Perfiles de parámetros acústicos de la Voz, su uso e incidencia como método práctico para la implantación y rehabilitación de la Voz y el Habla.

Sergio Suárez Guerra
Centro de investigación en Computación (CIC) –IPN, CP 07738, México D.F.
Instituto de Cibernética, Matemática y Física (ICIMAF) de Cuba, Calle 15 e/ C y D Vedado, Ciudad de La Habana.
Número de teléfono: (552) 55- 5729 6000 ext. 56588. Fax: (552) 55- 5586 2936
E-Mail: ssuarez@cic.ipn.mx ; sergio@icmf.inf.cu

Resumen

La posibilidad de "ver lo que se dice", ha resultado muy novedosa como método para la implantación y rehabilitación de la Voz y el Habla. Visualizar de forma inmediata, mediante una gráfica, los perfiles acústicos de los principales parámetros de la señal de voz y asociarlos con imágenes que representan lo dicho, ha resultado una alternativa adicional muy estimulante en el campo de la Foniatría y en Escuelas Especiales. En protocolos de investigación realizados durante un año en Escuelas Especiales, se ha notado un adelanto sustancial en el aprendizaje de la correcta dicción, en aquellos alumnos que adicionalmente al método tradicional, utilizaron un sistema de extracción y visualización de perfiles acústicos, representación de imágenes asociada al sonido y realimentación auditiva del sonido patrón y del producido por el usuario durante la sesión de trabajo. En el presente año escolar, el sistema Exparam V.2.0, continua siendo una opción para su introducción a Nivel Nacional en la Escuelas Especiales de la República de Cuba. Estadísticas del uso de la aplicación de la versión 1.2 se obtuvieron en el curso 2000 - 2001 y se esperan resultados del uso de la versión 2.0 a finales del curso 2002 – 2003. En el CIC – IPN del DF, México, se termina un sistema que a modo de evaluación se prueba en el Instituto de la Comunicación Humana (InCH) de México D.F, que busca apoyar la gestión en consultas de foniatría mediante el análisis de la voz y que involucra de igual manera la representación de perfiles acústicos, así como la inclusión de una base de datos clínicos de los pacientes. La versión Exparam 2.0 es utilizada en el InCH, para la creación de carpetas de archivos de voces clasificados por problemas, a partir de las cuales se desarrollaran nuevas aplicaciones con el empleo del análisis de las características de perfiles acústicos de la voz.

Palabras clave: Speech, Voice proccesing, Voice parameters, Acoustic voice analysis.

1 Antecedentes

Desde mediados de la década de los 80 se inició el desarrollo y comercialización de sistemas de análisis de voz con graficación de perfiles paramétricos, entre los parámetros más comunes tenemos la intensidad de la señal y su cruce por ceros. Estos perfiles paramétricos se realizaban no sólo para la señal pura de voz, también se realizaban para determinadas bandas de frecuencias bien estudiadas en las cuales está el mayor contenido de la información hablada: formantes o frecuencias de resonancias del tracto vocal, así como la parte del espectro que caracteriza a los sonidos fricativos y el tono fundamental.
En la década de los 90 aparecieron sistemas, que sin mostrar los perfiles de parámetros acústicos, presentaban imágenes capaces de ser movidas o alteradas por la presencia de determinado nivel o duración de un parámetro en específico. A principios de esa década se construyó un equipo a la medida llamado VIDEOVOZ, cuyo objetivo es la extracción y representación de perfiles paramétricos acústicos de la señal de voz en tiempo real, con posibilidades de realizar comparaciones cualitativas de dos perfiles acústicos, con fines de entrenamiento. La posibilidad de representación de los perfiles paramétricos en tiempo real, incorpora el elemento de realimentación visual de "ver lo que se dice". De este equipo se construyeron una cantidad superior a 20 y fueron instalados en las escuelas provinciales de Educación Especial de la República de Cuba.
A finales de los 90's se concluyó la primera versión del software: "Sistema para la extracción y análisis de parámetros de la voz" EXPARAM V.1.2; el mismo contiene la versión de presentación que utiliza el VIDEOVOZ y adiciona la posibilidad de disponer de una biblioteca de archivos de voces para el auto entrenamiento, así como la realimentación auditiva del sonido representado en las gráficas de perfiles, elemento éste muy útil para percibir la dicción de lo que el usuario dice, o sea, doble realimentación: visual y auditiva. De manera

experimental se instaló en una escuela de niños sordos y con trastornos del lenguaje, realizándose una evaluación del mismo en las actividades de enseñanza para niños de 4to y 5to grado, los resultados fueron alentadores, los profesores se familiarizaron con el uso de computadoras personales estándares y el software aplicado, los niños aceptaron el producto como un elemento de aprendizaje que les posibilitaba además el acceso a las computadoras.
A inicios del siglo XXI se continúa el desarrollo de aplicaciones para la educación y se inicia el diseño y programación de sistemas para el análisis de voz en el área médica de consultas de foniatría.

2 Nuevos desarrollos y presentaciones

Parámetros de trabajo en el proyecto

Actuales	Futuro inmediato
• Amplitud - AO • Cruce de ceros - NZ • Formante 1 - F1 • Formante 2 - F2 • Alta frecuencia - RO • Tono Fundamental - FO • Espectro de frecuencia - FFT	• Jitter • Shimmer. • Clasificación de perturbaciones. • Representaciones gráficas del jitter, contorno de variabilidad del jitter y clasificación del tipo de fonación.

El objetivo central para los nuevos desarrollos se presenta en dos líneas: los sistemas educacionales y las aplicaciones médicas.

En la Figura No.1, podemos ver el esquema de trabajo del proyecto actual.

La figura No.1 representa el conjunto de bloques, procesos y su ínter relación, de las tareas a diseñar y poner en funcionamiento, en un proyecto de investigación que se lleva acabo de forma conjunta entre el Centro de Investigación en Computación (CIC – IPN de México) y el Instituto de Cibernética, Matemática y Física (ICIMAF – Cuba).

En el esquema, las tesis: 1, 2 y 3, son trabajos que en la actualidad se realizan para titulaciones de Maestros en Ciencias de Computación, las cuales se apoyan mutuamente, así como en módulos que son el resultado del trabajo de especialistas de análisis de señales. Las aplicaciones educacionales, en Escuelas Especiales, se corresponden con el uso y ensamble de los diferentes módulos obtenidos.

Sistema de procesamiento de voz para aplicaciones foniátricas

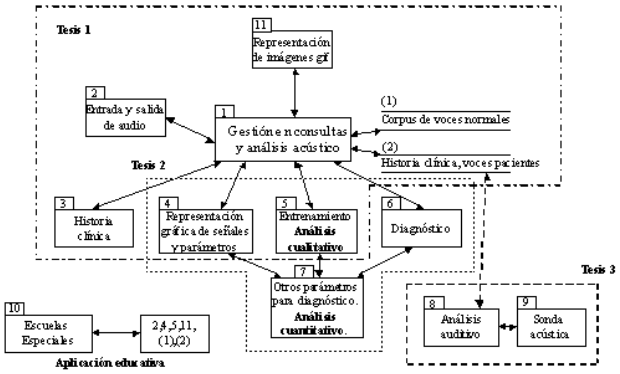


Figura No. 1. Nuevos desarrollos

3 Extracción y medición de parámetros

El intervalo de tiempo utilizado para el procesamiento de la señal de voz, durante el cual se extraen, grafican y calculan los valores de la mayoría de los parámetros de la señal voz es de 20 mseg. Para el caso de las mediciones relacionadas con el tono fundamental el intervalo es mayor. Los parámetros con mayor información utilizados para el análisis de la señal voz, se detallan en la tabla No.1.

Tabla No.1. Parámetros con mayor información en la voz.
Descripción de estos parámetros y su utilidad para un segmento bajo análisis, donde fm = frecuencia de muestreo; donde S_i son las muestras i en el intervalo:

AO: Caracteriza la energía media de la señal de voz en el intervalo.

$$AO_m = 1/N \sum_{i=0}^{N} S_i^2$$

NZ: Da una medida de la frecuencia en el intervalo.

$$NZ = \sum_{i=0}^{N} \text{Signo}(S_{i+1}) - \text{Signo}(S_i) > 0$$

F1: Valor medio de la frecuencia en la banda de 250 a 1000 Hz; SF1 = muestra en la banda F1:

$$NZ1 = \sum_{i=0}^{N} \text{Signo}(SF1_{i+1}) - \text{Signo}(SF1_i) > 0$$

$$F1_m = fm * (NZ1\text{-}1) / \sum_{i=0}^{NZ1-1} n_i$$

Donde n_i es la cantidad de muestras para un período de la señal en el segmento de análisis y $\sum n_i$ es la cantidad de muestras que contiene el intervalo que ocupan los períodos detectados de la señal en el segmento. Con este cálculo la F1 que se obtiene, es la F1 promedio en el segmento de análisis. Al estar limitada en banda el análisis, pues el resultado es aceptable. Para una medición más precisa de los componentes armónicos en cada segmento, se puede utilizar el análisis espectral con ayuda de la Transformada Rápida de Fourier (FFT) que es una de las opciones del sistema.

F2: Valor medio de la frecuencia la banda de 1000 a 3500 Hz: SF2 = muestra en la banda F2.

$$NZ2 = \sum_{i=0}^{N} \text{Signo}(SF2_{i+1}) - \text{Signo}(SF2_i) > 0$$

$$F2_m = fm * (NZ2\text{-}1) / \sum_{i=0}^{NZ2-1} n_i$$

RO: Valor medio de la frecuencia en la banda de 3500 a 6400Hz: SRO = Muestra en la banda RO.

$$NZRO = \sum_{i=0}^{N} \text{Signo}(SRO_{i+1}) - \text{Signo}(SRO_i) > 0$$

$$RO_m = fm * (NZRO\text{-}1) / \sum_{i=0}^{NZRO-1} n_i$$

FO: Valor medio de la frecuencia del tono fundamental en el intervalo.

$$F0_m = 1/N \sum^{N} F0_i$$

i=0

El tono fundamental (pitch) de la voz, es el parámetro más importante a tener en cuenta en el análisis de Voz y Habla, pues a partir de este es que se producen los sonidos que caracterizan los segmentos sonoros en la fonación. Cualquier perturbación en el tono fundamental, se refleja inmediatamente en la salida de información y altera la correcta dicción.

Aislar y extraer la señal del tono fundamental es el primer paso para aplicar cálculos de análisis de comportamiento estadísticos y de medición de la estabilidad en esta señal. Por otra parte, durante la articulación de palabras, la producción del tono fundamental se ve interrumpida, de ahí que los análisis del comportamiento del tono fundamental se realicen para segmentos sonoros con contenido invariable, ej. una vocal sostenida: 'a','e'.

A la señal del tono fundamental aislada, se le realiza la medición del valor de la frecuencia correspondiente a cada ciclo durante el segmento de análisis seleccionado, obteniendo los valores $F0_i$ correspondientes. También es posible medir la variación de la amplitud del tono fundamental AF0i, lo cual es útil para determinar otras características de comportamiento de estabilidad y calidad en la producción de Voz y Habla.

FFT: Espectro de frecuencia.

El cálculo del espectro de frecuencia se realiza para toda la señal bajo análisis, utilizándose para ello el algoritmo de la Transformada rápida de Fourier (FFT).

Los límites y precisión en el cálculo de la FFT permiten ajustar sus resultados en función de la relación del ancho de la ventana M que es una potencia de 2 y el número de muestras N, N menor o igual a M. Si N < M las muestras N < n < M se hacen igual a CERO. Así es posible realizar análisis de tiempo corto o largo. Para el análisis de tiempo corto se escoge el segmento o intervalo de tiempo N de la señal que contiene por lo general un sólo ciclo del tono fundamental y así se eliminan ruidos y dispersiones de los otros ciclos. La frecuencia de muestreo, fm, y su relación con el ancho de la ventana, determinan la precisión de respuesta, delta f, en los resultados de la FFT.

$$H(k) = \sum_{n=0}^{M-1} S(n)\, e^{-j(2pi/M)\, n\, k} \qquad \text{para } k = 0, 1, \ldots, M \qquad M = 2^{x}$$

$$H(w) = \sqrt{Re(w)^2 + Im(w)^2} \text{ para } w = 0, 1, \ldots, M$$

$$\text{Delta } f = fm / M$$

También, para eliminar ruidos por truncamiento de la señal en las fronteras de la ventana, se aplica alguna función de ventana. La más utilizada es la de Hamming.

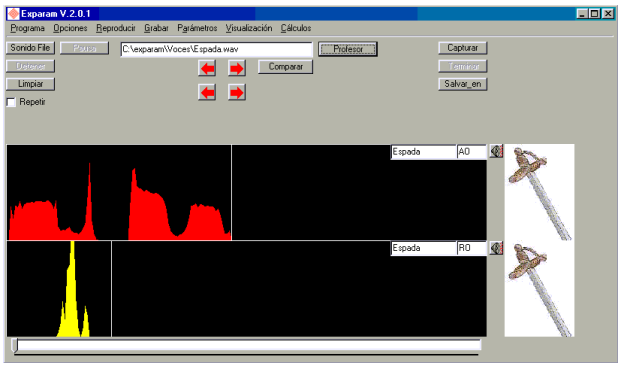
$$Whm(k) = 0.54 + 046 \cos(2pik/M-1) \quad k = 0,1, \ldots, M-1$$

Jitter: Medida de la inestabilidad de la frecuencia del tono fundamental F0.

El Jitter es una medida de la inestabilidad a corto tiempo de la F0 durante la producción del tono fundamental. El Jitter medio está definido como:

$$Jitter_m(Hz) = 1 / N - 1 \sum_{i=0}^{N-1} \left| F0_i - F0_{i+1} \right|$$

Shimmer: Medida de la inestabilidad de la amplitud del tono fundamental AF0.



El Shimmer es una medida de la inestabilidad a corto tiempo de la AF0 durante la producción del tono fundamental. El Shimmer medio está definido como:

$$\text{Shimmer}_m(\text{dB}) = 1 / N - 1 \sum_{i=0}^{N-1} \left| 20 \log (AF0_i / AF0_{i+1}) \right|$$

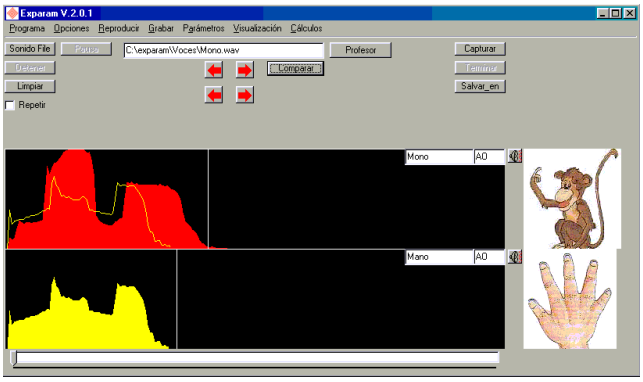
Otros valores estadísticos a considerar durante el análisis del tono fundamental son calculados a partir del comportamiento de los valores F0i, AF0i.
Clasificación de perturbaciones:
Desviación estándar y coeficiente de variación, son calculados para conocer la variación de F0i y la Intensidad del tono fundamental.
Porcentaje de variación y promedio de perturbación, son valores estadísticos que por su importancia, se calculan a modo de ofrecer una visión más particular de la perturbación del tono fundamental de la señal bajo análisis, tanto para la frecuencia como la amplitud.

4. Sistemas y resultados alcanzados.

4.1 Area educacional.

- Paquete de programas para la extracción y análisis de parámetros de la voz. EXPARAM 2.0.

Este sistema fue culminado en julio del 2002 y se encuentra en evaluación en Escuelas Especiales de la Ciudad de La Habana, a finales del curso escolar 2002 – 2003 se recibirán los resultados.
Como diferencia fundamental con la versión 1.2, se tiene el soporte de programación sobre Delphi 5 (Pascal Orientado a Objetos y presentación visual).



Otras adecuaciones son:

- ❖ Se incorporan representaciones de imágenes para cada sonido del corpus de voces que se recibe con el sistema, de forma tal que además de poder ver la señal acústica de la voz y los perfiles paramétricos que se extraen de cada sonido, el usuario puede ver el significado del sonido en una figura.

- El usuario puede incorporar nuevos archivos de voces. Si desea ver la representación en forma de imagen, tiene que incorporar la misma en la carpeta correspondiente, en el formato JPG.
- Se añadieron representaciones gráficas de parámetros de la voz como son: tono fundamental (pitch) y espectrogramas de frecuencia.

En las figuras No.2 y 3, se muestran representaciones de los perfiles paramétricas de varias palabras, así como de las imágenes para el significado de cada sonido.

Figura 2 a)

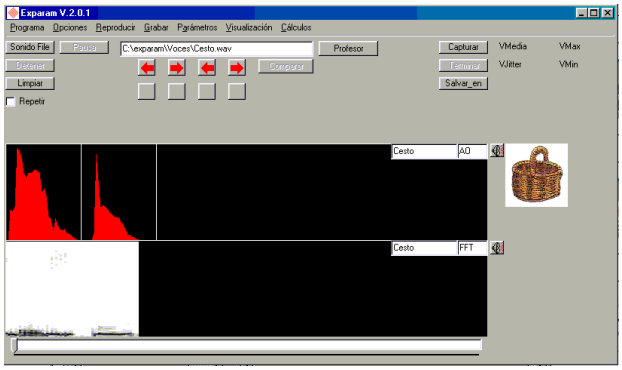


Figura 2 b)

Figuras No. 2: a) Palabra “espada”, representación de los perfiles acústicos de intensidad, parte superior (vocales con mayor nivel) y de frecuencias alta, parte inferior (fricativo ‘S’); b) Palabra “cesto”, representación del perfil acústico de intensidad, parte superior y del espectrograma, parte inferior

Figura 3 a)

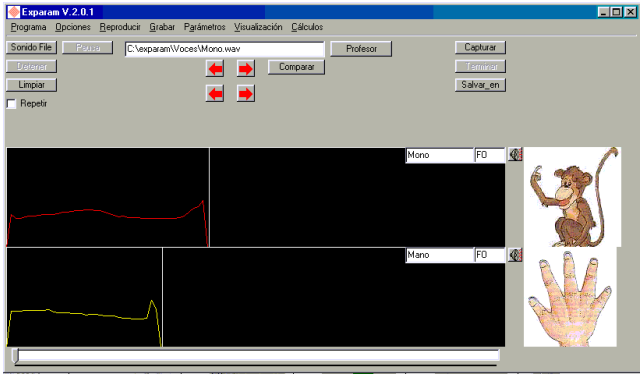


Figura 3 b)

Figura No. 3, palabra “mano y mono”: a) Representación de los perfiles de intensidades de ambas palabras; b) Representación de los perfiles del Tono Fundamental para cada palabra.

Como se puede apreciar de las figuras 2 y 3, las representaciones de los perfiles paramétricos acústicos, es muy versátil. En la figura No.3 a), se presenta en la ventana superior el efecto de comparación de perfiles, con un contorno idéntico al del perfil inferior, sobre la gráfica del perfil superior, este efecto es posible verse para cualquier representación de perfiles, excepto cuando una de las gráficas es un espectrograma (figura No.2 b)).

Los iconos de bocinas, a la derecha de las ventanas: superior e inferior; se utilizan para reproducir el sonido desde el inicio de la pantalla, hasta la posición que ocupan los cursores respectivos.

La descripción del funcionamiento del sistema está disponible en el Manual de Usuario.

4.2 Area médica.

- Sistema para la gestión y análisis acústico en consultas de foniatría. FONAVOZ 1.0.

Como resultado de la culminación de un trabajo de tesis de Maestría, se dispone de una versión prototipo del sistema FONAVOZ V.1.0 el cual integra el llenado de una base de datos de la información personal de los pacientes atendidos, con la situación diagnóstica que determina el especialista médico y adiciona la recopilación de archivos de voces in sito, para realizar análisis acústicos de la voz del paciente.

Este sistema, en su base de datos, contiene las características que el personal del servicio de foniatría del Instituto de la Comunicación Humana (InCH) de la Secretaria de Salud del D.F. de México. La grabación de los archivos de voces y la visualización de los perfiles paramétricos acústicos, se corresponde con la experiencia que se ha tenido en las representaciones de los sistemas educacionales desarrollados.

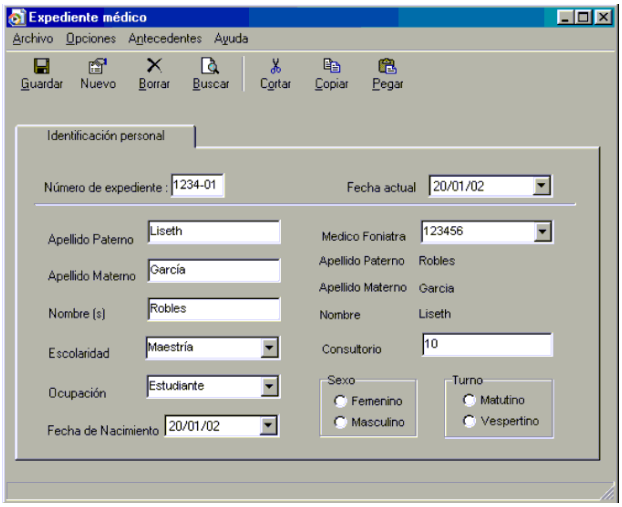
Dentro de las características de representaciones, al igual que EXPARAM 2.0, se tiene la posibilidad de observar: señal real, parámetro de frecuencia e imagen con el significado de la palabra dicha. Con estas representaciones el especialista puede comparar los gráficos correspondientes a una voz normal vs. la del paciente bajo estudio y además de diagnosticar, proponer sesiones de entrenamiento, donde el paciente mejore su dicción, si es el caso.

El sistema está realizado en Delphi 5.0 y la base de datos es del tipo ADO.

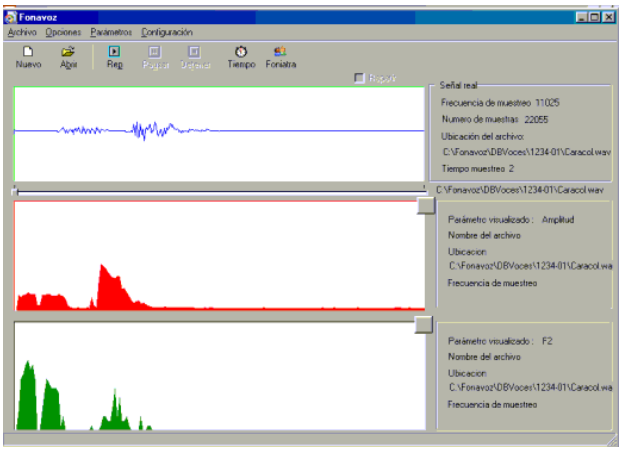
Se trabaja en la recopilación de un corpus de voces de personas con problemas de voz y habla, con el fin de realizar una investigación de clasificación diagnóstica, utilizando la técnica de análisis de señales en la señal de voz.

Los resultados del uso de este sistema y recomendaciones de extensión de funcionalidades se esperan para la primera mitad del año 2003.

En la figura No.4, se pueden apreciar dos de las pantallas que el sistema presenta.



Expediente Clínico 4 a)



Gráficas de señal y parámetros 4 b)

Figura No. 4 Presentaciones de FONAVOZ a) Expediente clínico; b) gráficas de señal y parámetros.

5. Aplicaciones en rehabilitación y diagnóstico. Parámetros.

Una de las “habilidades” de los sistemas en aplicación y desarrollo es la posibilidad de comparar, mediante la superposición de gráficas, los dos perfiles acústicos que aparecen en las ventanas de presentación. Con esta opción es que se puede realizar el análisis cualitativo de dos sonidos provenientes de fuentes diferentes, dos locutores o usuarios.

La superposición de las gráficas de los perfiles paramétricos acústicos facilita a los usuarios del sistema observar que tanto se asemejan en su composición y pronunciación los dos sonidos de una forma objetiva. Si además se cuenta con la posibilidad de oír los sonidos correspondientes a cada ventana de presentación, pues se puede establecer una asociación de la apreciación objetiva, lo que se ve, vs. lo que se escucha.

Para trabajos de rehabilitación fonética el poseer ambas realimentaciones, sonora y visual, ha resultado ser un factor muy importante, pues el usuario del sistema puede detectar con mayor precisión en que parte de la fonación está incurriendo en una falta, omisión o producción inadecuada.

Esta posibilidad se puede observar en la figura No. 5, donde la misma palabra ha sido pronunciada por dos personas diferentes, los perfiles acústicos de la energía son similares, con la diferencia que el perfil superior tiene menos definida la vocal ‘a’ terminal.

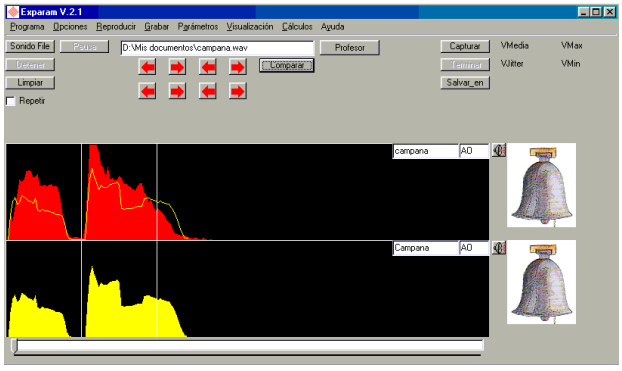


Figura No. 5. Comparación cualitativa de dos perfiles paramétricos acústicos.

Como parte de los trabajos enfocados al diagnóstico de problemas de Voz y Habla, se introduce el análisis cuantitativo, que se realiza con la incorporación de dos cursores en las ventanas de presentación. De esta forma es posible delimitar el espacio del sonido que es necesario analizar y realizar las mediciones de los valores reales que alcanza el parámetro en ese intervalo.

El análisis cuantitativo está limitado en su aplicación al tono fundamental - F0 (Pitch), ya que el mismo es el responsable de la producción del resto de los parámetros acústicos durante la fonación y cualquier alteración en su producción es reflejada en los que de él dependen: formantes F1, F2, ..., Fn.

Con las mediciones objetivas del comportamiento del tono fundamental en cuanto a su frecuencia y amplitud, es posible detectar problemas que estarían vinculados bien con acciones motoras (músculos), de comunicación nerviosa o ambas. Entre los parámetros a medir en este tipo de análisis tenemos: valor medio de la frecuencia fundamental, variabilidad (jitter), amplitud del tono fundamental, variabilidad (Shimmer); así como distribución estadística de la variación de la frecuencia y amplitud.

Para trabajos de diagnóstico de problemas de Voz y Habla, la medición de estos parámetros no es por si solo suficiente, pero si son una parte muy importante para orientar el uso de otras pruebas y mediciones del tracto faríngeo y las cuerdas vocales, con el fin de establecer las posibles causas del problema.

En la figura No. 6, parte superior derecha, se muestran los resultados de la medición de las características del tono fundamental para un segmento de voz enmarcado entre los dos cursores, para el mismo sonido, producido por dos personas diferentes.

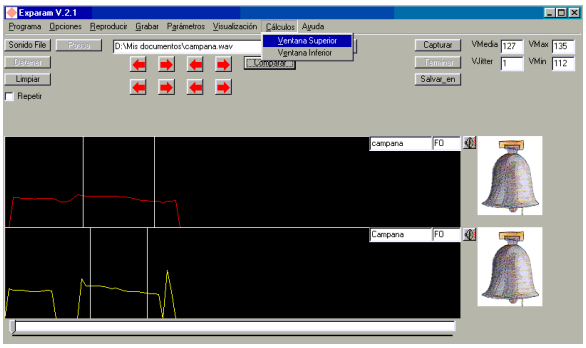


Figura 6 a) Cálculo parte superior

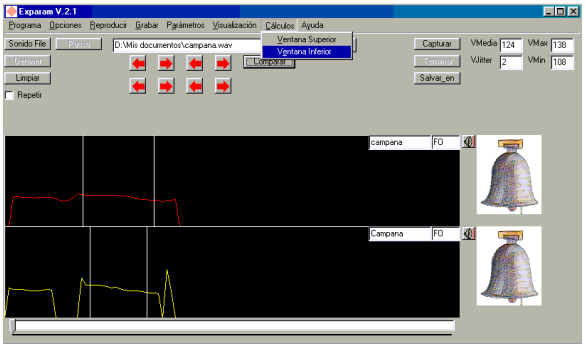


Figura 6 b) Cálculo parte inferior

Figura No. 6, medición de características del tono fundamental para un segmento de voz.

Cabe señalar, que tanto para las aplicaciones de rehabilitación como para el diagnóstico de problemas de Voz y Habla, los sistemas aquí presentados constituyen herramientas auxiliares, cuya efectividad puede ser altamente

positiva o no en dependencia del problema bajo tratamiento o estudio. Hay problemas de rehabilitación que requieren de tratamiento quirúrgico y luego realizar ejercitación vocal, para lo cual estos sistemas pueden ser muy útiles. Para el caso del diagnóstico se utilizan otros tipos de análisis y pruebas que son fundamentales, no es posible solamente con la medición de las características del comportamiento del tono fundamental decidirlo todo.

6. Conclusiones.

Los sistemas presentados son el resultado de varios años de trabajo en la línea de procesamiento de voz para aplicaciones de educación y medicina. Ambos están instalados en centros de atención y con ellos se llevan a cabo trabajos de análisis, entrenamiento y rehabilitación de problemas de Voz y Habla.

Los parámetros que se calculan y representan tienen el mayor significado para los trabajos de entrenamiento, rehabilitación y en especial para diagnóstico de problemas de Voz y Habla. Tal es el caso del tono fundamental F0.

La comparación de perfiles paramétricos acústicos es rápida de entender y factible de ser utilizada eficientemente por los usuarios de los sistemas.

La aceptación de los especialistas en el área de educación y medicina ha sido muy favorable, por ser sistemas fáciles de manipular y ofrecerse a bajo costo. Ya se encuentran en proceso de redacción manuales orientados a la Educación Especial, para diferentes niveles de escolaridad.

Bibliografía

1. Sistema para la Extracción y Análisis de parámetros de la voz EXPARAM V.2.0. 2000-2002. Manual de Usuario. CENDA, ICIMAF, Ciudad de la Habana, Cuba. ISBN 959-7056-17-8.
2. Sistema para la gestión y análisis acústico en consultas de Foniatría (FONAVOZ), Tesis de Maestría en Ciencias de la Computación, CIC – IPN, México D.F. Diciembre de 2002. Lic. Liseth García Robles.
3. Workshop on Acoustic Voice Analysis. Ingo R. Titze, Ph.D. 17 – 18 of february, 1994, Denver Colorado, National Center for Voice and Speech. Summary Statement.

Referencias

1. http://www.sqlab.com/scParamAcoustFR.htm
2. http://www.sqlab.com/scEggFR.htm