

# Sistema de Práctica de la Oralización

J. M. Garin<sup>†</sup>    P. D. Agüero<sup>†</sup>    J. C. Tulli<sup>†</sup>    E. L. Gonzalez<sup>†</sup>    G. Aranda<sup>†</sup>

<sup>†</sup>Laboratorio de Comunicaciones, Universidad Nacional de Mar del Plata, Mar del Plata, Argentina  
jmgarin@fi.mdp.edu.ar  
pdaguero@fi.mdp.edu.ar    jctulli@fi.mdp.edu.ar

**Abstract**— This on-going project focuses on the need of software for the assistance of deaf people in Argentina, with a special emphasis in the development of freeware with total support of Argentinian dialects.

Our software employs the latest technological advances in the software area by using JAVA programming language. The main advantage of this approach is the high portability and the avoidance of multiple code development to support different platforms.

The user interacts with the software through the computer screen. The system asks the user to utter some specific sound, word or sentence, which is recorded using a microphone. Then, the speech signal is analyzed and compared with existing reference patterns in a speech database. The results are shown in the screen. They include both the user's mistakes and possible actions to correct them.

All the data generated during the interaction between user and software are sent via Internet to the servers at the university where the development of the project takes place. This information is important for future improvements of the analysis tools and to check the correct performance of the implemented algorithms under different recording conditions.

**Keywords**— hipoacúsicos, programación java, entrenamiento del habla, distribución gratuita,

## 1. RESUMEN

Este proyecto en desarrollo analiza la problemática argentina actual sobre la escasez de programas para computadora destinados a la asistencia de personas hipoacúsicas, de distribución gratuita y en idioma español local.

Debido a esto, se decidió diseñar un programa que emplee los últimos avances tecnológicos en el área informática, utilizando el lenguaje de programación JAVA, que nos permite una amplia portabilidad del programa y elimina las limitaciones de incompatibilidad de programación entre sistemas operativos.

Su funcionamiento consiste en indicarle en pantalla al usuario lo que se pretende que este pronuncie, capturar su habla a través de un micrófono, analizarla y comparar los resultados con patrones preestablecidos. Luego de esto se

le informa nuevamente a través de la pantalla los posibles errores de fonación a fin de que pueda ir corrigiéndolos con el continuo entrenamiento [1].

Los datos de este proceso son enviados vía Internet a los servidores del grupo de desarrollo permitiendo realizar análisis posteriores tanto por el profesional a cargo como por los desarrolladores, lo que facilita mantener un registro de las actividades y progresos [2].

## 2. INTRODUCCION

En la actualidad existen gran cantidad de personas que presentan una disminución en la capacidad auditiva, las cuales se pueden clasificar en dos grupos, los hipoacúsicos y los sordos profundos (cofóticos).

La hipoacusia es la disminución de la percepción auditiva. La misma puede provenir de distintas causas congénitas tanto de origen genético, neurosensorial, como por mal formaciones; o bien adquirirse a través de enfermedades tales como otitis, tímpanosclerosis, colesteatomas, etc.

Las personas que nacieron con alguna de estas anomalías o que las adquirieron durante la infancia [3], no cuentan con la realimentación auditiva necesaria, lo que les impide percibir con claridad los sonidos del ambiente y los que ellos mismos pronuncian.

Esta dificultad les produce grandes inconvenientes en el aprendizaje del habla [4], lentificando el desarrollo de la dimensión del lenguaje, del pensamiento abstracto, y obstaculizando la categorización de los objetos, la estructura y sistematización de la realidad.

Los sordos profundos, son personas que no tienen percepción auditiva. Debido a esto, desde niños aprenden a comunicarse a través del lenguaje de señas, lo que lleva a que tengan aún más dificultades que los hipoacúsicos en la comunicación a través del habla.

A su vez en ambos grupos se originan distintas alteraciones en la articulación de fonemas por ausencia o alteración de estos por otros, dependiendo del grado de sordera que tengan. Estas alteraciones se pueden clasificar en tres grupos [5]: dislalias, diglosias y disartrias.

Las dislalias son trastornos en la articulación de los fonemas, generándolos de manera equívoca o directamente excluyéndolos. Por ejemplo dicen “mai” en vez de “maíz” o “tre” en vez de “tres”.

Las diglosias son trastornos de origen orgánico provocado por lesiones físicas o malformaciones de los órganos articulatorios y periféricos. Estas últimas, según la parte

del aparato fonatorio que afectan, se clasifican en: labiales, mandibulares, dentales, linguales y palatales, para las cuales se pueden citar, como ejemplo, las correspondientes afecciones: labio leporino, prognatismo, agenesias dentarias, glossectomía y paladar hendido.

Por último, las disartrias son anomalías en la articulación de naturaleza motriz, generadas por lesiones en el sistema nervioso central, como así también por enfermedades en los nervios o músculos de la lengua, faringe y laringe. A éstas se las puede clasificar a su vez en cinco grupos: flácida, espástica, atáxica, extrapiramidal y mixta. En las dos primeras la afección está localizada en las neuronas motrices inferior y superior. La atáxica es ocasionada por una afección en el cerebelo generada por lesiones cerebelosas bilaterales o generalizadas. La extrapiramidal es generada por lesiones en el sistema que lleva el mismo nombre, el cual forma parte del sistema paleocéfalo. Y por último las mixtas, que son una conjunción de las mencionadas anteriormente.

Las personas que tienen estas alteraciones son tratadas por los especialistas del habla a través de la utilización de métodos auditivo-visuales, llevando a cabo una reeducación ortofónica y logopédica, siendo ellos los que interpretan el sonido y le informan al paciente lo que dijo a través de señas o dibujos. Esto implica que el profesional no cuenta con más información sobre lo que dijo el paciente que lo que pudo captar e interpretar a través de la experiencia.

Debido al avance tecnológico en el área informática, se dispone de diversos tratamientos asistidos por computadora que complementan el trabajo del profesional y permiten un conocimiento más amplio y preciso de los sonidos del paciente y de sus anomalías [6].

En la actualidad existen diversos programas para computadoras orientados a la asistencia de personas hipoacúsicas, tales como el SpeechViewer de IBM [7], Laureate Learning Systems [8], Vocaliza [9], etc.

El SpeechViewer de IBM es un visualizador fonético para los profesionales del habla, que cuenta con dos partes principales: terapia clínica, que incluye ejercicios de habla directa, y gestión de terapia, que suministra información sobre el rendimiento del paciente y sus progresos.

Por su parte la compañía Laureate Learning Systems desarrolló diversos paquetes de programas para la asistencia de personas con capacidades diferentes, de los cuales uno está dirigido a niños con problemas en el habla. Dentro de este paquete, se pueden encontrar programas tanto de ejercitación general como específicos, como ser de pronunciación de sustantivos, verbos, pronombres personales, etc. Uno de ellos es el "TalkTime with Tucker" que está basado en juegos, en los que el niño va interactuando y avanzando niveles a través del habla. Esto es ampliamente recomendable ya que estimula el aprendizaje.

El programa Vocaliza [10] pertenece al proyecto Comunica de la Universidad de Zaragoza, el cual trabaja en tres niveles de lenguaje: fonológico, semántico y sintáctico. El mismo está desarrollado para el tratamiento de niños, ya que se basa principalmente en juegos e interacción con dibujos. También cuenta con un sintetizador de voz, que les permite a las personas con alguna percepción auditiva saber

como se debe pronunciar correctamente una palabra. Este último, a diferencia de los mencionados anteriormente, es para el idioma español y su distribución es gratuita lo que permite el acceso a todas las personas.

En la actualidad no existe ningún programa de asistencia de personas con hipoacusia para el idioma español argentino. Si bien Vocaliza está desarrollado para el idioma español, la transcripción fonética no se corresponde con la de nuestro país, por lo que no es apto para ejercitar el habla de nuestros niños. Por esta razón hemos decidido desarrollar un programa diseñado en lenguaje de programación  $JAVA^{TM}$  (por su portabilidad) y de distribución gratuita, para que sea accesible a todas las personas que padezcan estas afecciones en nuestro país.

En este artículo se describe el programa y sus dos modos de trabajo dependiendo del usuario: un profesional (Sec. 3.1) o un alumno (Sec. 3.2). Seguido a esto, se encuentran las conclusiones (Sec. 4) y la bibliografía.

### 3. DESCRIPCION DEL PROGRAMA

El programa está dividido en dos partes principales, una manejada por los profesionales (Sec. 3.1) y otra por los alumnos (Sec. 3.2).

La primer ventana del programa (Fig. 1) le permite al usuario ingresar en la sección de trabajo que le corresponde: profesional o alumno.

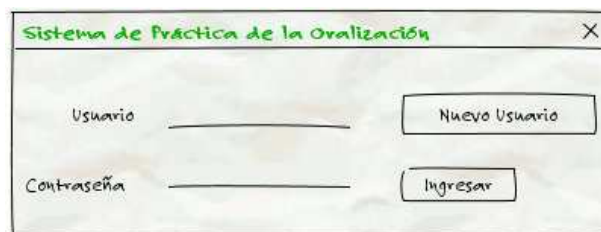


Figura 1: Ventana Inicial

También permite crear un nuevo usuario, el cual debe configurarse como profesor o como alumno y debe cargar todos sus datos personales y aquellos de pertinentes al trabajo con la aplicación (nombre de usuario, contraseña, etc).

#### 3.1. Profesionales

Si el usuario es un profesor, se abre una pantalla en donde se puede optar por 4 diferentes tareas (Fig. 2), las cuales se detallan en las siguientes secciones: cargar datos de los alumnos (Sec. 3.1.1), evolución de los alumnos (Sec. 3.1.2), seleccionar alumnos a cargo (Sec. 3.1.3) y configurar sesiones (Sec. 3.1.4).

##### 3.1.1. Cargar datos de los alumnos

Esta ventana permite, luego de haber previamente seleccionado al alumno en una ventana intermedia (Fig. 3), configurar los datos pertinentes a su condición, como ser las patologías (por ejemplo: dislalias), audiogramas, etc, proporcionándole al programa información adicional muy valiosa ya que en el momento del análisis este podrá saber previamente alguna dislalia, ahorrando tiempo y principalmente pudiendo analizar de forma más eficaz lo que dijo

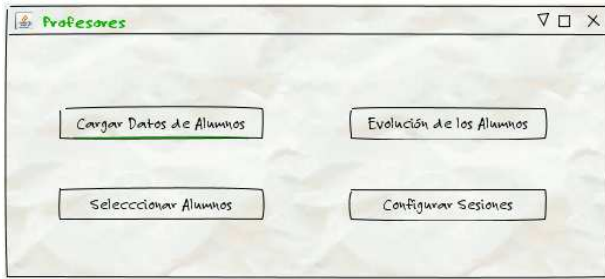


Figura 2: Ventana de Profesores

la persona. Por ejemplo si se conoce que la persona omite vocales de las palabras, cuando se le pida que pronuncie la palabra “árbol”, el programa no va a forzar la detección de una “a” en el principio de la grabación, sino que va a contemplar la opción de que la persona haya dicho “rbol”.

### 3.1.2. Evolución de los alumnos

Aquí, los profesionales pueden ver en forma gráfica la evolución que tuvieron sus alumnos con el uso del programa.

A través de gráficos temporales se indica el porcentaje de efectividad al pronunciar las palabras y se enumeran los errores que cometen con mas frecuencia, permitiéndole de este modo saber hacia que debe apuntar su tratamiento y que debe corregir en el alumno. Para el caso en que ya se hayan aplicado medidas correctivas, se podría saber la efectividad de estas respecto a otras etapas del tratamiento.

Luego, el profesional puede optar por exportar los datos del alumno para realizar otros tipos de análisis de estos y así obtener, resultados en otros formatos, según sus necesidades.

### 3.1.3. Seleccionar alumnos a cargo

Esta etapa de la configuración permite al profesional asignarse alumnos a su cargo, los cuales luego le aparecerán en los otros menús para poder configurarlos y también para ver su evolución (Fig. 3).

Esto esta pensado para el caso en que una misma computadora sea utilizada por distintos turnos de alumnos y profesores, de modo que cada profesor pueda trabajar solo con su listado de usuarios.

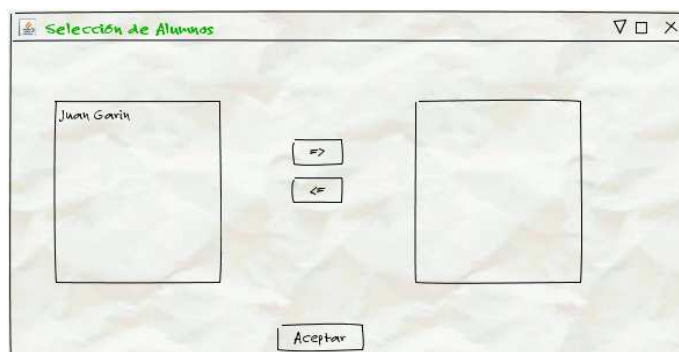


Figura 3: Selección de Alumnos

### 3.1.4. Configurar sesiones

La configuración de las sesiones va a ser la tarea principal de los profesores.

Esta instancia permite, luego de seleccionar al alumno en una ventana intermedia (Fig. 3), configurarle un plan de trabajo. Es decir, un conjunto de sesiones para que trabaje con las distintas aplicaciones del programa. Este proceso consiste primeramente en seleccionar la aplicación, y luego el listado de palabras que el alumno va a tener que pronunciar durante la sesión de trabajo. Una vez guardada esa sesión, se pueden seguir agregando otras de modo de organizar un plan de trabajo semanal o mensual del alumno. Las palabras de ejercitación están incluidas en el programa y no son configurables por el profesor, ya que cada una pertenece a una base de datos especifica que contiene los parámetros que luego van a ser utilizados durante el análisis en las diversas aplicaciones.



Figura 4: Primera instancia del juego de Merlin

## 3.2. Alumnos

En el caso de ingresar como alumno, el programa verifica que este tenga sesiones configuradas, y si es así, continúa con la última sin realizar. De este modo se abrirá la aplicación que el profesor asignó en la sesión y le pedirá que pronuncie la primer palabra del listado que tiene asignado.

La aplicación que actualmente utiliza el programa como preestablecida es la del mago Merlin. En ella se solicita que pronuncie una determinada palabra del listado, como muestra la Figura 4. Luego se habilita el micrófono y comienza la grabación. Una vez que el programa detecta que se pronunció una palabra, detiene la grabación y comienza con la etapa de análisis, la cual esta comprendida por 3 pasos:

1. Recorte la señal, eliminando los silencios de comienzo y fin de la grabación.
2. Análisis con Dynamic Time Warping para obtener las fronteras de los fonemas.

3. Extracción de la energía y de frecuencia fundamental en cada fonema [11]. En una etapa siguiente se agregará la evaluación de los fonemas).



Figura 5: Segunda instancia del juego de Merlin

Luego de terminado el análisis de la señal, se procede a comparar los valores obtenidos con los de la base de datos de referencia, comparando la duración, la energía o el pitch (dependiendo de la terapia elegida). Finalmente, se estima un parámetro de escala para poder graficar las letras permitiendo que el usuario pueda detectar sus posibles errores como lo muestra la Figura 5.

En dicha figura se puede apreciar las variaciones de amplitud que simbolizan, para el caso de una letra estirada, un pitch o energía mas altos que la referencia. Para el caso en que la duración del fonema haya sido mayor o menor que la referencia, se indica a través del ancho de la letra. Por último, si el alumno se siente conforme con los resultados obtenidos, pasa a la palabra siguiente, y si no, puede volver a repetir la palabra cuantas veces desee.

#### 4. CONCLUSIONES Y TRABAJO FUTURO

El estado actual de desarrollo de la aplicación puede apreciarse en la Figura 6.

Las versiones siguientes de este software incluirán los siguientes avances. En lo que respecta a las aplicaciones, se implementarán otras variantes para lograr abarcar los diversos niveles del lenguaje.

La aplicación del mago trabaja en un nivel fonológico, que en una primer instancia no comprueba la identidad fonética sino solo la energía y su pitch.

En una siguiente etapa a esta aplicación se le va a poder activar la verificación de los fonemas, para poder así indicar si se pronunció el fonema correcto o se reemplazó por otro.

Para trabajar un nivel mas bajo, en el cual solo se pretende ejercitar la potencia sonora del alumno, se van a implementar aplicaciones que permitan visualizar estos parámetros de forma adecuada, de modo de darle al usuario la posibilidad de entrenar otras características de su voz.

Cuando el alumno ya maneja el lenguaje básico puede optar por trabajar con un juego de aventura, en el que un personaje avanza a través de un camino con obstáculos que solo se pueden sortear si se pronuncia de manera correcta una palabra indicada en pantalla. Para esto se preconfigura un umbral de exactitud inicial, el que ira incrementando a medida que avanzan los niveles. Este tipo de aplicaciones es muy recomendable ya que genera que el alumno se divierta mientras aprende y que se motive a seguir jugando y a ejercitarse de manera indirecta.

El avance mas importante que va a presentar esta aplicación respecto a las existentes es la forma de trabajo remota, es decir que el usuario al descargar el programa crea una cuenta en el servidor, la cual va a utilizar el programa para almacenar todos los datos personales y los referentes a su trabajo diario. Esto permite que el usuario solo tenga que tener instalado el programa y no toda su información personal, ofreciéndole así una mayor confidencialidad con los otros usuarios de la computadora, y sobre todo ofrecer la posibilidad de que pueda ingresar en diversas computadoras durante su entrenamiento sin tener que transportar su información en medios extraíbles.

Una vez concluido el desarrollo del programa, se va a poder utilizar en otros países de habla hispana simplemente modificando su base de datos de referencia.

Debido a que se van a almacenar los archivos de audio de las sesiones en el servidor, se va a ir generando una base de datos sobre voces patológicas (bajo consentimiento de los usuarios), lo que nos va a permitir realizar mejores trabajos futuros.

La elección de la imagen de Merlin para la aplicación del mago se debió a que era necesario una imagen para realizar el motor de graficación del programa. Pero esta tiene “copyright”, por lo cual en las versiones que serán publicadas en la página web se van a utilizar diseños propios.

Este proyecto cuenta con una página web en donde se irán colocando los avances del sistema y el programa para su descarga [12].

#### Referencias

- [1] E. E. Crawford, “Acoustic signals as visual biofeedback in the speech training of hearing impaired children,” Ph.D. dissertation, 2007.
- [2] J. M. Garin, P. D. Agüero, J. C. Tulli, and E. L. Gonzalez, “Efectos de la compresión mp3 en la determinación del pitch,” *RPIC*, Septiembre 2009.
- [3] A. M. R. Romero and J. L. Aledo, “Trastornos del lenguaje,” 2006.
- [4] R. L. L. Plaza, “Dificultades en el desarrollo del lenguaje oral e intervención,” E.O.E.P. N°1 de Zaragoza, Tech. Rep., 2006.
- [5] E. C. SA, “Dislalias,” Gabinete Psicopedagógico logopedia, Tech. Rep., 2007.
- [6] E. Eriksson, O. Bälter, O. Engwall, A.-M. Öster, and H. Sidenbladh-Kjellström, “Design recommendations for



Figura 6: Desarrollo de la Aplicación

a computer-based speech training system based on end-user interviews,” 2005.

[7] <http://www.dif.gob.mx/cta/soluciones/speech.html>

[8] <http://www.laureatelearning.com>

[9] <http://www.vocaliza.es>

[10] C. Vaquero, O. Saz, and E. Leida, “Vocaliza: an application for computer-aided speech therapy in spanish language,” *IV Jornadas en Tecnología del Habla*, Noviembre 2006.

[11] P. C. Bagshaw, “Automatic prosodic analysis for computer aided pronunciation teaching,” Ph.D. dissertation, 1994.

[12] <http://elaf1.fi.mdp.edu.ar/pegasus/>