

SPO: una ayuda en JAVA para la oralización de hipoacúsicos

J. M. Garin, P. D. Agüero, J. C. Tulli, E. L. Gonzalez y A. J. Uriz

Facultad de Ingeniería, Universidad Nacional de Mar del Plata, jmgarin@fi.mdp.edu.ar, Tel: +54 223 481 66 00 (251) , Juan B. Justo 4302(7600)

Resumen—Este proyecto en desarrollo aborda la problemática actual en Argentina 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 diseñó un programa que emplea los últimos avances tecnológicos en el área informática, utilizando el lenguaje de programación JAVA, el cual permite una amplia portabilidad 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 pronuncie, capturar su emisión 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 lingüísticos a fin de que pueda corregirlos con el continuo entrenamiento.

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, a través de un registro de las actividades y progresos.

Palabras clave— hipoacúsicos, entrenamiento del habla, distribución gratuita.

I. INTRODUCCIÓN

En la Argentina existe gran cantidad de personas que presentan una disminución en la capacidad auditiva, la cual les genera grandes dificultades al momento de comunicarse a través del habla. Esta disminución se la puede clasificar en dos grandes 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 malformaciones; 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 de niños [11], no cuentan con la realimentación auditiva necesaria, lo que les impide percibir con claridad los sonidos del ambiente y los que ellos mismos emiten.

Esta dificultad les produce grandes inconvenientes en el aprendizaje del habla [10], 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 [12]: dislalias, diglosias y disartrias.

En la actualidad existen diversos softwares para la asistencia de personas con este tipo de problemas, tales como: SpeechViewer de IBM [18], Laureate Learning Systems [19], Vocaliza [20, 21], etc. Sin embargo, ninguno de ellos se corresponde con el idioma de nuestro país ni con nuestros dialectos [13], y los que mas se ajustarían a las necesidades no son de distribución gratuita. Es por esto que se decidió comenzar a desarrollar una plataforma de entrenamiento de personas hipoacúsicas que tuviese como premisas principales: personalización para el español local, de distribución gratuita, y compatible con la mayoría de los sistemas operativos.

Actualmente este desarrollo se encuentra en una etapa intermedia, en la cual está diseñada la interfaz administrativa y el motor gráfico para las aplicaciones, pero la sección encargada del análisis de la señal se encuentra en una etapa inicial, que unicamente detecta frecuencia fundamental y energía. Los parámetros para el correcto funcionamiento del reconocedor fonético todavía no se han ajustado debido a que es necesario grabar una base de datos en español local.

En la sección siguiente se aborda el funcionamiento básico de la aplicación. Luego, en la sección 3 se detallan las opciones del menú de configuraciones del profesor, y en la sección 4 las aplicaciones disponibles y su funcionamiento. Seguido a esto, en la sección 5 se explican los métodos de análisis de las señales, y por último en la sección 6 las conclusiones y trabajos futuros.

II. SISTEMA DE PRÁCTICA DE LA ORALIZACIÓN

El desarrollo de esta plataforma surge de la necesidad de tener en las escuelas especiales de la Argentina un software que ayude al entrenamiento del habla de los niños y adultos hipoacúsicos [7]. Para poder satisfacer esta necesidad de manera gratuita se decidió utilizar JAVA© [16] como lenguaje de programación y Netbeans© [17] como plataforma de desarrollo, la cual nos permite crear una interfaz gráfica amigable, que pueda ser utilizada tanto por niños como por adultos.

El funcionamiento básico de las aplicaciones que administra esta plataforma se puede resumir en los siguientes pasos:

1. Indicarle en pantalla al usuario lo que se pretende que pronuncie.
2. Habilitar la línea del micrófono de modo que pueda grabarse la fonación.
3. Almacenar digitalmente la fonación.
4. Analizar los datos según el parámetro que se desea medir.
5. Mostrar en pantalla los resultados en un formato amigable y entendible por el usuario.

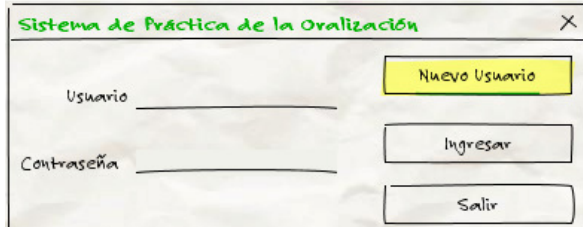


Fig. 1: Ventana Inicial

Algunos de estos pasos son levemente diferentes dependiendo de la aplicación, como se verá más adelante en la Sección 4.

Como innovación a los sistemas existentes, se utilizará un sistema de asociación remoto vía internet, de manera que el usuario solo tenga que tener descargado el software en la pc, ya que todos sus datos personales estarán almacenados en un servidor, los cuales se descargarán temporalmente cuando el usuario ingrese sus datos.

Para poder utilizar el programa es necesario crear un usuario de tipo profesor y uno de tipo alumno, el cual, como se verá más adelante, debe ser asignado al profesor. Para la creación de nuevos usuarios, se debe ingresar en "Nuevo Usuario" en la ventana inicial (Fig. 1), llenar los campos correspondientes a los datos personales del usuario, y luego hacer click en "Guardar".

El funcionamiento del software se lo puede dividir en tres secciones: PROFESORES (Sec. 3), ALUMNOS (Sec. 4) y ANÁLISIS DE LA SEÑAL (Sec. 5).

III. PROFESORES

Si el usuario es un profesor, luego de iniciar la sesión, el programa visualiza en pantalla la sección mostrada en la Fig. 2, en donde podrá optar por las siguientes tareas a realizar:

- Seleccionar alumnos a su cargo (Sec. 3.1).
- Cargar datos de los alumnos (Sec. 3.2).
- Ver la evolución de sus alumnos (Sec. 3.3).
- Configurar sesiones (Sec. 3.4).
- Configurar el micrófono (Sec. 3.5).
- Probar las aplicaciones (Sec. 3.6).

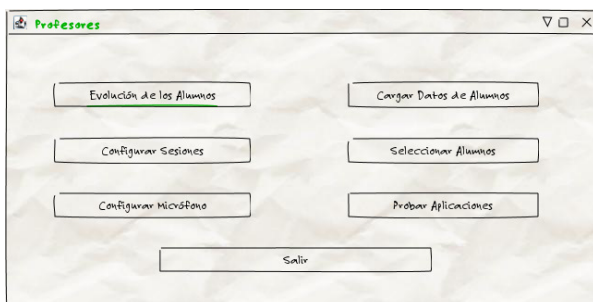


Fig. 2: Ventana de Profesores

III.I SELECCIONAR ALUMNOS A CARGO

Esta etapa de configuración le permite al profesional determinar que alumnos están su cargo, los cuales luego le aparecerán en los otros menús para poder configurarlos y también para ver su evolución. El proceso consiste en tomar los nombres de los alumnos de una lista y desplazarlos hacia la lista de alumnos seleccionados. Luego aceptando estos cambios el sistema actualiza la base de datos y ya no es necesario realizar esta tarea cada vez que se utiliza el software.

Este modo de operación está 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 sólo con su listado de usuarios.

III.II CARGAR DATOS DE LOS ALUMNOS

Esta ventana permite, luego de haber previamente seleccionado al alumno en una ventana intermedia, configurar los datos pertinentes a su condición, tales como las patologías (por ejemplo: dislalias), audiogramas, etc. Estos datos proporcionan al programa información adicional muy valiosa, ya que en el momento del análisis podrá saber previamente alguna dislalia, ahorrando tiempo de cálculo y analizando de forma más eficaz lo que dijo 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 buscar directamente una "a" en el principio de la grabación, sino que va a contemplar la opción de que la persona haya dicho "rbol".

Fecha	Hora	Alumno	Aplicación	Nº de Sesión
01/04/2010	15:30:25	Pocestelmm		
01/04/2010	15:39:54	Pocestelmm		
01/04/2010	15:40:04	Pocestelmm		

Fig. 4: Listado del Historial

III.III EVOLUCIÓN DE LOS ALUMNOS

Aquí los profesionales pueden seguir toda la evolución de sus alumnos asignados durante la utilización del software. Estos datos de las sesiones aparecen en forma de lista (Fig. 4), ordenado por fecha y hora, según el alumno que se elija de los que tenga previamente asignados.

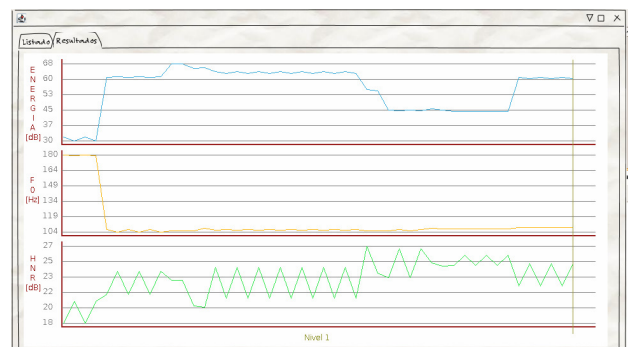


Fig. 5: Gráfico de Resultados

Luego, seleccionando una fila correspondiente a una sesión, en la lengüeta de resultados aparecerán gráficas de diversos parámetros acústicos analizados, según el tipo de aplicación que se haya utilizado. A modo de ejemplo en la Figura 5 se visualizan los valores de energía, pitch y HNR (Relación armónicos a Ruido), correspondientes a cada nivel jugado en la aplicación del pececito (Sec. 4).

III.IV 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, configurarle un plan de trabajo. Este consiste en un conjunto de sesiones para que trabaje con las distintas aplicaciones del programa.

Esta tarea consiste en seleccionar la aplicación de un listado para luego elegir las opciones que se presenten en la pantalla, asignando así las configuraciones según el tratamiento deseado. Una vez guardada la sesión, se pueden seguir agregando más actividades, de modo de planear un tratamiento semanal o mensual del alumno.

En el caso de las aplicaciones que trabajan con fonética, 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 específica que contiene los parámetros que luego van a ser utilizados durante el análisis en las diversas aplicaciones.

III.V CONFIGURAR EL MICRÓFONO

Antes de utilizar las aplicaciones es necesario establecer en el sistema los umbrales de sonido detectados por el micrófono, los cuales dependen de cada computadora y el ambiente. Para esto se creó una instancia de configuración, en la que se establecen estos parámetros y se almacenan, para que no sea necesario calibrarlos cada vez que se ejecute el programa.

III.VI PROBAR LAS APLICACIONES

Esta sección le ofrece al profesional la opción de probar las aplicaciones antes de asignárselas en los tratamientos a los alumnos, permitiéndole de este modo saber que parámetros debe ajustar al momento de asignársela a un alumno en una sesión.

IV. ALUMNOS

Al iniciar la sesión, si el usuario es un alumno, el sistema corrobora que este tenga sesiones por realizar. Si es así, se inicializa un menú con las aplicaciones disponibles. Seleccionando una de estas se abre la pantalla inicial de la aplicación (Fig. 6), donde se puede leer las instrucciones del juego y luego proceder a su uso.

Si la última sesión quedó sin completarse, se sigue con esta a partir del último punto realizado.



Fig. 6: Pantalla inicial de la aplicación

Las aplicaciones con que cuenta actualmente la plataforma trabajan con dos visualizaciones gráficas distintas. La primera, correspondiente a un tipo de análisis fonético, se implementó con una aplicación en la cual un mago (Fig. 7) le pide al usuario que pronuncie una determinada palabra, la cual es grabada a través del micrófono y almacenada digitalmente. Seguido a esto, los datos son analizados por el reconocedor fonético, el cual almacena los resultados. Por último, la aplicación toma estos resultados y se los retorna al usuario en función de la correcta duración (ancho) y energía (altura) de los fonemas, modificando el tamaño de las letras que componen la palabra [4].



Fig. 7: Mago

El segundo tipo de visualización trabaja en tiempo real, en donde el usuario pronuncia un sonido indicado por el profesional y un objeto en la pantalla se desplaza según los resultados obtenidos de energía o pitch [5].

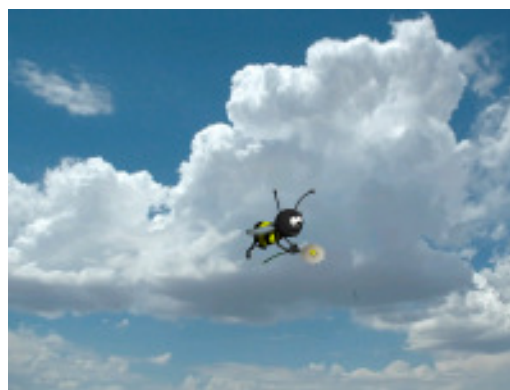


Fig. 8: Abeja



Fig. 9: Pececito

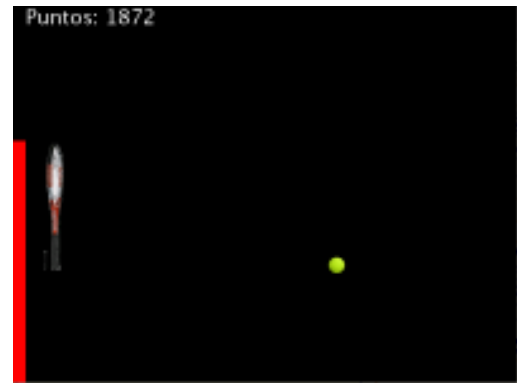


Fig. 11: Frontón

Para este tipo de visualización se crearon cuatro aplicaciones:

- **Abeja.** En esta aplicación una abeja vuela por la pantalla, en donde la altura de vuelo esta determinada por la energía sonora emitida por el jugador (Fig. 8).
- **Pececito.** Al igual que el juego de la Abeja el pececito sube y baja dependiendo de la energía, pero para poder superar los niveles debe sortear los obstáculos que aparecen en la pantalla. De este modo, se puede avanzar por cinco niveles de distinta dificultad (Fig. 9).
- **Globo.** El funcionamiento es similar al del juego de la abeja, con la diferencia de que el protagonista es un globo aerostático, y lo que determina el movimiento vertical no es la energía sino la frecuencia fundamental (Fig. 10).
- **Frontón.** El juego se basa en una pelota de tenis que rebota entre una pared y una raqueta, la cual se maneja con la energía sonora emitida por el jugador, desplazándose en un eje vertical y deteniéndose en la altura indicada por la barra de desplazamiento roja ubicada a la izquierda de la misma (Fig. 11).



Fig. 10: Globo

V. ANÁLISIS DE LA SEÑAL

El principal objetivo es conseguir un conjunto de parámetros acústicos representativos de la señal, que permitan una comparación con voces de personas oyentes, y de esta manera ofrecer una evaluación. Para esto se realiza un análisis de la señal de voz del usuario que consiste en dos etapas.

En la primer etapa se realiza un filtrado y una segmentación. El filtrado elimina componentes de baja frecuencia y de muy alta frecuencia, que no corresponden a voz. Y la segmentación divide a la señal en un conjunto de ventanas de 20ms de duración y con un solapado del 50%.

En la segunda etapa se procede a realizar el cálculo de diversos parámetros acústicos a cada uno de los segmentos, tales como:

- **Frecuencia fundamental:** frecuencia de vibración de la glotis, que permite calcular diversos parámetros acústicos útiles para el análisis de la voz, tales como jitter y shimmer [2].
- **Energía:** energía de la señal, calculada en base al período de la señal para sonidos sonoros, y para todo el segmento en el caso de sonidos sordos [9].
- **Sonoridad:** indicación de la probabilidad de pertenencia del sonido del segmento a los fonemas sonoros [6].
- **Relación de armónicos a ruido:** parámetro que indica el nivel de armónicos en el sonido, con el objeto de distinguir sonidos vocálicos y no vocálicos [3].
- **Coefficientes cepstrales de frecuencia Mel (MFCC):** parámetros acústicos utilizados por el sistema de segmentación de fonemas [9].

Los primeros cuatro parámetros son utilizados para realizar las diversas acciones descritas en las aplicaciones de las secciones anteriores. En cambio, los parámetros MFCC son utilizados por el sistema de reconocimiento de voz SPHINX [22] para la segmentación del audio en fonemas [14, 1]. De esta manera es posible un análisis individualizado de los fonemas, como es el caso de la aplicación del mago.

VI. CONCLUSIONES Y TRABAJO FUTURO

Actualmente el software se encuentra en una etapa intermedia de desarrollo, en la cual ya se programó en mayor parte toda la sección administrativa y algunas aplicaciones, tal como puede observarse en la Figura 12 con

mas detalle. También es posible descargar una versión de demostración que se encuentra alojada en nuestro servidor web[15].

Las versiones siguientes de este software incluirán diversos aspectos. En lo que respecta a las aplicaciones actuales, se las mejorará en lo relativo a la visualización gráfica, haciendolas aún mas completas y atractivas para los usuarios.

La aplicación del mago se implementará para ser utilizada con un análisis fonológico, para lo cual se debe ajustar el reconocedor fonético y realizar una base de datos de voces del dialecto local para poder entrenar los modelos acústicos.

La implementación de aplicaciones como las que actualmente cuenta el software, les permite a los alumnos ejercitar de manera indirecta su habla, ya que su principal motivación es ganar el juego. A su vez, el profesional puede realizar un análisis de la evolución del alumno consultando y comparando sus historiales de trabajo.

El avance mas importante que va a presentar esta aplicación respecto a las existentes es la forma de trabajo remota. Al descargar el programa el usuario crea una cuenta en el servidor, la cual se va a utilizar en 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 la flexibilidad de ingresar en diversas computadoras durante sus prácticas sin tener que transportar su información en medios extraíbles (por ejemplo: pendrives). De esta manera podrá ingresar al programa, descargar todas sus configuraciones y sus sesiones de trabajo, trabajar el tiempo deseado, cerrarlo y continuar luego en otra computadora que puede estar en la escuela junto al profesor, en un cybercafé o en su domicilio.

Una vez concluido el desarrollo del programa, este se podrá utilizar en otros países de habla hispana simplemente modificando su base de datos de referencia. Debido a que se prevee almacenar los archivos de audio de las sesiones en el servidor, se generará una base de datos sobre voces patológicas (bajo consentimiento de los usuarios), lo que nos va a permitir realizar mejores trabajos futuros [8].

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.

REFERENCIAS

[1] Antonio Jordi Adell and Antonio Bonafonte. Towards Phone Segmentation For Concatenative Speech Synthesis. 5th ISCA Speech Synthesis Workshop, 2004.

[2] Paul Christopher Bagshaw. Automatic prosodic analysis for computer aided pronunciation teaching. PhD thesis, 1994.

[3] Paul Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. IFA, 17:97-110, 1993.

[4] Mukund Devarajan. A computer-based articulation training aid for short words. PhD thesis, 2003.

[5] Elina Eriksson and Olle Bälter and Olov Engwall and Anne-Marie Öster and Hedvig Sidenbladh-Kjellström. Design Recommendations for a Computer-Based Speech Training System Based on End-User Interviews. 2005.

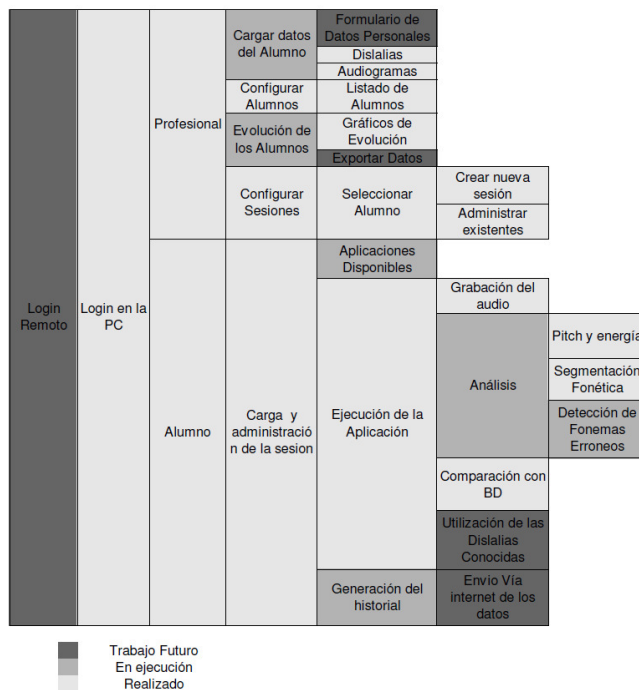


Fig. 12 Desarrollo de la Aplicación

[6] Antonio Galves and Jesus Garcia and Denise Duarte and Charlotte Galves. Sonority as a basis for rhythmic class discrimination. In Speech Prosody, :323-326, 2002.

[7] Juan Manuel Garin and Pablo Daniel Agüero and Juan Carlos Tulli. Nuevos Aportes al entrenamiento de Personas Hipoacúsicas. TISE, 2008.

[8] Juan Manuel Garin and Pablo Daniel Agüero and Juan Carlos Tulli and Esteban Lucio Gonzalez. Efectos de la compresión MP3 en la determinación del pitch. RPIC, 2009.

[9] Xuedong Huang and Alex Acero and Hsiao-Wuen Hon. Spoken language processing. Prentice Hall PTR, 2001.

[10] Rafael Lizandra La Plaza. Dificultades en el desarrollo del lenguaje oral e intervención. Technical report, E.O.E.P. N°1 de Zaragoza, 2006.

[11] A. M. Redondo Romero and J. Lorente Aledo. Trastornos del Lenguaje. 2006.

[12] Emagnet Consultoria SA. Dislalias. Technical report, Gabinete Psicopedagógico logopedia, 2007.

[13] Oscar Saz and William Rodríguez and Eduardo Lleida and Carlos Vaquero. A novel corpus of children's disordered speech. Technical report, Aragon Institute for Engineering Research, University of Zaragoza, 2008.

[14] D. T. Toledano and A. Hernández Gómez and Luis Villarrubia Grande. Automatic phone segmentation. IEEE Transactions on Speech and Audio Processing, 2(6):617-625, 2003.

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

[16] <http://java.com/es/>.

[17] <http://netbeans.org>.

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

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

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

[21] Carlos Vaquero and Oscar Saz and Eduardo Leida. Vocaliza: an application for computer-aided speech therapy in Spanish language. IV Jornadas en Tecnología del Habla, 2006.

[22] Willie Walker and Paul Lamere and Philip Kwok and Bhiksha Raj and Rita Singh and Evandro Gouvea and Peter Wolf and Joe Woelfel. Sphinx-4: A flexible open source framework for speech recognition. Technical report, CMU, 2004.