

Desarrollo y prueba de un Software de Apoyo para la Mejora de la Pronunciación en Alumnos con Déficit del Habla por Dislalia

José Manuel Soto-Corzo¹, José Juan Hernández-Mora², José Federico Ramírez-Cruz³, Blanca Estela Pedroza-Méndez⁴

^{1,2,3,4}Instituto Tecnológico de Apizaco,
Apizaco, Tlaxcala, México

¹jmsc_1984@live.com.mx, ²jhhmora@itapizaco.edu.mx,
³framirez@itapizaco.edu.mx, ⁴thelismedina@hotmail.com

Paper received on 24/07/12, Accepted on 12/09/12.

Resumen. Vivimos en una etapa de evolución informática orientada en gran parte hacia una evolución constante en el desarrollo de interfaces y métodos de interacción humano computadora. Dentro de esta evolución se observa una clara mejoría en la forma como las aplicaciones se adaptan a las necesidades de personas que presentan alguna deficiencia en sus capacidades. Desde nuestra perspectiva esta misma evolución debe converger no solo en la interacción, sino en el propósito incluyente del software mismo. De esta forma se planteó la necesidad de desarrollar software enfocado a la rehabilitación de personas déficit del habla por Dislalia como una herramienta de apoyo en su proceso de rehabilitación que les permita mejorar su capacidad de pronunciación por medio de actividades.

Palabras Clave: Dislalia, Rehabilitación, Reconocimiento Automático del habla, Ingeniería de Software.

1 Introducción

El término “trastornos del aprendizaje” se aplica a los problemas que plantean obstáculos al rendimiento académico o escolar. Estos problemas afectan a uno de cada diez niños, y se pueden detectar desde los cinco años de edad. Dependiendo de qué tan graves sean estos trastornos pueden afectar el aprendizaje del niño, y también sus relaciones con otras personas. Las dificultades escolares se pueden clasificar por su etiología en dificultades de origen primario, dificultades específicas y dificultades derivadas del entorno socio-familiar y cultural [1]. Dentro de las dificultades específicas, que se refieren a los trastornos del aprendizaje, se encuentra a la Dislalia, la cual se define como la dificultad para pronunciar un fonema determinado. Muchos de los niños que presentan este déficit del habla son

diagnosticados de manera tardía y su desarrollo cognitivo se ve limitado dependiendo del nivel del trastorno que presentan y del tiempo que tarde su rehabilitación. Algunos niños con dislalia pertenecen a sistemas educativos de atención especial debido a la presencia de este trastorno de aprendizaje. Existen productos comerciales enfocados a la rehabilitación de personas con déficit del habla, pero muchos no están habilitados para el idioma español y aquellos que si lo están, representan una inversión que muy pocas instituciones públicas de educación especial pueden adquirir. El reto consiste en proveer de aplicaciones de software libre diseñados específicamente para apoyar los procesos de enseñanza y/o rehabilitación de personas con déficit del habla. El sistema propuesto esta desarrollado para ser empleado en escuelas de educación especial de nivel primaria en el estado de Tlaxcala y cuenta con un repositorio de actividades a manera de juegos para que los alumnos puedan mejorar su capacidad de pronunciación. El sistema cuenta con un modulo de reconocimiento fonético que permite determinar si el alumno ha pronunciado correctamente los fonemas, silabas o palabras solicitadas a través del desarrollo de las actividades contenidas en el mismo sistema. Asi también cuenta con una base de datos embebida en el sistema y un visualizador de graficas de progreso.

2 Entornos de la investigación

En esta sección se presenta de manera breve una descripción de los aspectos relacionados con la investigación requerida para el desarrollo del software.

2.1 Dislalia

La dislalia se define como la dificultad para pronunciar un fonema determinado. Las posibles causas en este trastorno del habla son:

- a) **Retraso fonológico:** Relacionados a un retraso en el desarrollo del habla. El niño simplifica los sonidos pues no aprendió a producir los más complejos de su lengua. Su capacidad articulatoria no está afectada.
- b) **Trastorno fonético:** Cuando el niño no adquirió correctamente los patrones de movimiento necesarios para la producción de algunos sonidos del habla, por lo que no mueve los músculos que se encargan del habla, como debería, y por eso comete omisiones, sustituciones y distorsiones de algunos sonidos de la lengua.
- c) **Alteraciones físicas:** presencia de malformaciones físicas en los órganos que intervienen en el habla y que impiden pronunciar muchos sonidos. Las lesiones en el sistema nervioso pueden ocasionar alteraciones en el movimiento y en la coordinación de los músculos implicados en el habla. (ejemplo, labio leporino o hendido), pueden ocasionar una dislalia.

Un programa de rehabilitación para la dislalia consiste en los siguientes pasos:

1. Estimulación de la capacidad del niño para producir sonidos.

2. Estimulación de la coordinación de los movimientos necesarios para la pronunciación de sonidos, ejercicios labiales y linguales.
3. Realización de ejercicios para producir sonidos dentro de sílabas hasta automatizar el patrón muscular necesario para articular el sonido.
4. Empezar con palabras completas a través de juegos.
5. Una vez que el niño es capaz de pronunciar los sonidos difíciles, tratar que lo realice en conversaciones espontáneas (fuera de terapia).

2.2 Fonología y Fonética

La fonología es un sub campo de la lingüística. Mientras que la fonética estudia la naturaleza acústica y fisiológica de los sonidos, la fonología describe la manera en que los sonidos funcionan (en un lenguaje o en una lengua en general) en un nivel abstracto o mental [2]. Un sonido o fono se caracteriza por una serie de rasgos fonéticos y articulatorios, el número de dichos rasgos y la identificación de los mismos es tarea de la fonética. Un fono es cualquiera de las posibles realizaciones acústicas de un fonema. Los fonemas son unidades teóricas, postuladas para estudiar el nivel fonético-fonológico de una lengua humana. Entre los criterios para decidir, qué constituye o no un fonema se requiere que exista una función distintiva. Un fonema es una unidad fonológica diferenciadora, indivisible y abstracta.

2.3 Reconocimiento Automático del Habla

Dado que el habla se transmite a través del sonido y que se pueden extraer patrones de una señal de sonido, el reconocimiento automático del habla recae en el área del reconocimiento de patrones. Para nosotros los humanos es una tarea intuitiva pues forma parte del núcleo de nuestra inteligencia, sin embargo en el área de la computación representa uno de problemas con más retos [3]. La forma de extraer estos patrones es a través de métodos estadísticos y probabilísticos. Dependiendo del propósito para el cual se requiera del reconocimiento automático del habla se han de emplear diferentes técnicas de extracción y clasificación de datos. Dado que la explicación detallada de estas técnicas está más allá del propósito de este documento se sugiere consultar la bibliografía citada para más información [3], [4], [5].

3. Trabajos y herramientas similares

En esta sección se hace mención de algunos trabajos y herramientas similares para poder proporcionar un panorama muy general del estado del arte en cuanto a software de apoyo para personas con déficit del habla se refiere. En cuanto a sistemas comerciales los más representativos son los siguientes.

Speech Viewer.- Se trata del software comercial más antiguo diseñado para tal propósito. Fue desarrollado por la IBM. El programa utiliza animaciones a partir de entradas de voz, la versión actual es el Speech Viewer III. Está dirigido a problemas de lenguaje, audición y habla. Contiene varios ejercicios que tienen una respuesta

inmediata por medio de la voz y es utilizado en personas de todas las edades que cuenten con una discapacidad que afecte el desarrollo de su lenguaje [6].

Sistema Visha.- Su diseño gráfico es poco atractivo, sin embargo el sistema Visha consta de un conjunto de componentes de hardware (memoria, tarjeta de sonido, etc.) que deben instalarse en el ordenador, y un conjunto de programas de software que permiten el análisis de los parámetros del habla, la síntesis y codificación visual de la señal acústica y el reconocimiento de los sonidos [7].

Sistema AVEL.- El sistema AVEL incluye una tarjeta analógica para la computadora. La señal es captada por un micrófono unido a la computadora y almacenada de esta manera. En este sistema se encuentran aplicaciones de desmutización, articulación vocálica correcta, ejercicios de ritmo, control de tono e intensidad [8].

Como trabajo de investigación y diseño de prototipo uno de los más citados para el idioma español es desarrollado por Liseth García Robles y Sergio Suárez Guerra [9]. En su propuesta se observa la utilización de diversas técnicas para la extracción de características de la señal acústica, como son, el análisis, oscilográfico, el análisis espectral y la codificación por predicción lineal. El sistema realiza dos funciones: la primera consiste en permitir el seguimiento y rehabilitación de pacientes con trastornos del habla por medio de la visualización de su señal de voz, previamente analizada con las técnicas antes descritas. La segunda, consiste en permite grabar la voz de personas y por medio de sus análisis diagnosticar si el paciente tiene algún trastorno. El trabajo está orientado al reconocimiento del habla continua y no de fonemas o palabras aisladas. De estas herramientas de apoyo y software específico se observa que aquellos que son comerciales son una buena opción para terapeutas enfocados a la rehabilitación de personas con problemas del habla en general, sin embargo por cuestiones de precio o distribución no son accesibles para la mayoría de las instituciones de educación especial del estado de Tlaxcala. Los proyectos de investigación y prototipos no comerciales analizados, son buenos trabajos y aportan grandes ideas al desarrollo del sistema, pero su aplicación es muy general. La intención del sistema mostrado en este trabajo, se debe a la necesidad de un desarrollo orientado a aspectos específicos como la edad de los usuarios finales, su nivel de atención, etc.

4. Descripción del Sistema

El trabajo consiste en el desarrollo de una aplicación de computadora que sirva como herramienta de apoyo en los procesos de rehabilitación de alumnos de educación especial con déficit del habla por dislalia originada por retraso fonológico y/o trastorno fonético. El sistema está diseñado para ser administrado por un profesor que pueda llevar el control de los registros almacenados de los alumnos dados de alta en el sistema, así como observar y extraer información estadística del progreso de los alumnos. Los alumnos registrados en el sistema podrán llevar a cabo sesiones con actividades a manera de juegos que les permitan mejorar su capacidad de pronunciación de fonemas específicos. Cada alumno tendrá asignado un nivel de acuerdo a su capacidad de pronunciación. Esta escala de tres niveles es como se describe a continuación: el nivel A para alumnos con dificultad para pronunciar

fonemas de manera aislada; el nivel B para aquellos que no pueden pronunciar correctamente los fonemas dentro de sílabas; el nivel C es para quienes presentan dificultad para pronunciar los fonemas dentro de palabras completas. Dichos niveles plantean una escala de complejidad, por lo que el nivel deberá ser asignado al alumno por el profesor encargado de apoyarlo en su proceso de rehabilitación.

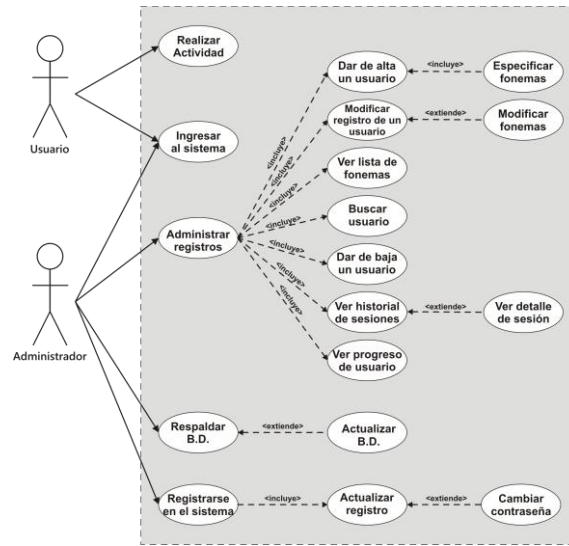


Figura 1. Diagrama de casos de uso del sistema.

En la Fig. 1 se pueden observar el diagrama de casos de uso. La mayoría de las funciones están contempladas para el administrador del sistema, sin embargo la más importante, corresponde a la realización de actividades por parte de los alumnos. La arquitectura general del sistema denominado *Speech View* se muestra en la Fig. 2.

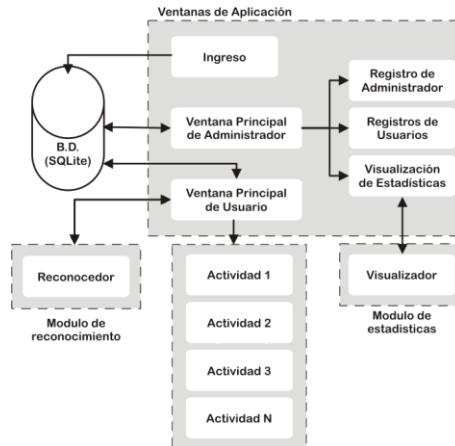


Figura 2. Arquitectura general del sistema.

Finalmente se presenta el diagrama de componentes del sistema en la Fig. 3.

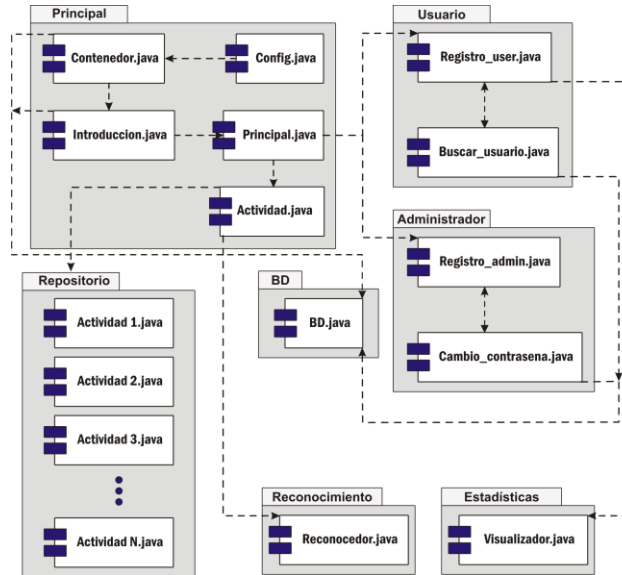


Figura 3. Diagrama de componentes del sistema.

A continuación se muestran imágenes de algunas pantallas del sistema.



Figura 4. a) Pantalla inicial; b) Ventana principal del sistema; c) Ventana de registro de administrador; d) Ventana para el registro de alumnos

Desde las pantallas mostradas en la Fig. 4 se puede ingresar al sistema (a), se puede llevar a cabo los escenarios de uso del administrador (b), se puede llevar el control de registro del administrador (c) y de los alumnos (d) respectivamente.

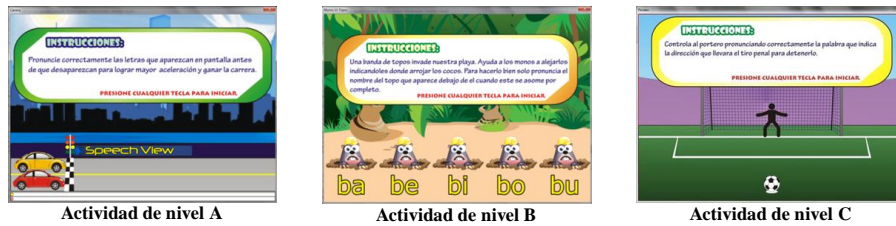


Figura 5. Ventanas iniciales de las actividades de pronunciación.

Las imágenes mostradas en la Fig. 5 corresponden a las actividades de pronunciación diseñadas e implementadas en el sistema.

5. Desarrollo del sistema

El sistema fue desarrollado empleando una metodología basada en prototipos, por medio de un patrón de desarrollo conocido como MVC (Modelo- Vista – Controlador) y empleando las siguientes herramientas de desarrollo:

Tabla 1. Herramientas de desarrollo y requerimientos asociados.

Herramienta	Requerimiento de desarrollo
Java SE Development Kit 1.7	Aplicación multiplataforma de software libre
Java Speech API 1.0	Reconocimiento automático del habla
Paquete Swing de Java	Desarrollo agíl de la interfaz de usuario
SDK de JavaFX 2.0	Elaboración de actividades de pronunciación animadas
SQLite 3.7	Resguardo de datos de los usuarios
JFreeChart 1.0.13	Generación de gráficas de progreso de los usuarios

A continuación se describe brevemente cada herramienta de desarrollo.

Java SE Development Kit.- La edición estándar del kit de desarrollo de Java en su versión 1.7 es un entorno de desarrollo para construir aplicaciones, applets y componentes usando el lenguaje de programación Java [10].

Java Speech API 1.0.- Java Speech Recognition define un estándar, fácil de usar, así como una plataforma que combina el software para el desarrollo de interfaces con el estado del arte de la tecnología en reconocimiento del habla [11].

Paquete Swing de Java.- El paquete Swing es parte de la JFC (*Java Foundation Classes*) en la plataforma Java. La JFC provee facilidades para ayudar a la gente a construir GUIs [12].

SDK de JavaFX 2.0.- El SDK de JavaFX permite crear y desplegar RIA's (*Rich Internet Applications*) que se comporten de forma consistente en múltiples

plataformas. Construida sobre la tecnología Java, la plataforma JavaFX proporciona un rico conjunto de gráficos y API's multimedia para hardware de alto rendimiento, aceleración de gráficos y los motores multimedia [13].

SQLite 3.7.- SQLite es una biblioteca que implementa un sistema autónomo, sin servidor, sin configuración, así como un motor de base de datos SQL. SQLite lee y escribe directamente en archivos de disco normal, por tanto, una base de datos SQL completa con varias tablas, índices, triggers y vistas, estará contenida en un archivo de disco único que puede ser autocontenido por la aplicación [14].

JFreeChart 1.0.13.- JFreeChart es una biblioteca gráfica de Java 100% libre, que hace que sea fácil para los desarrolladores poder mostrar gráficos de calidad profesional en sus aplicaciones [15].

6. Pruebas y resultados

Además de las pruebas unitarias para cada módulo de código durante la fase de desarrollo, fueron realizadas pruebas para evaluar el desempeño del módulo de reconocimiento del habla y el desempeño del sistema como herramienta de apoyo a procesos de rehabilitación de personas con déficit del habla por dislalia.

6.1 Pruebas al módulo de Reconocimiento Automático del Habla

Se llevaron a cabo 29 sesiones de pronunciación con tres niños en un rango de seis a ocho años de edad, todos ellos sin problemas de pronunciación, puesto que lo que se buscó fue validar el porcentaje de reconocimiento correcto del módulo. Cada uno de ellos realizó actividades de un nivel distinto. Fueron 6 sesiones de 5 actividades para el nivel A, 20 sesiones de 3 actividades para el nivel B y 3 sesiones de 3 actividades para el nivel C. A continuación se presentan los resultados y observaciones obtenidos. Los datos en la Tabla 2 muestran la tasa de reconocimiento general obtenido de las pruebas realizadas para cada tipo de elemento. Dentro de esta, las vocales y consonantes forman parte de las pruebas realizadas por medio de la actividad de nivel A, las sílabas por medio de la actividad de nivel B y finalmente las palabras corresponden a pruebas por medio de la actividad de nivel C.

Tabla 2. Tasas de reconocimiento generales por tipo de elemento

Tipo de elemento	Tasa de reconocimiento
Vocales	90.76 %
Consonantes	96.20 %
Sílabas	88.15 %
Palabras	92.33 %

El porcentaje de reconocimiento general del sistema, con base en los datos de la Tabla 2, es de 91.86 %. Se observa que los porcentajes de reconocimiento varían de acuerdo a diversos factores como lo son el ruido ambiental, la calidad del dispositivo de entrada de audio empleado, etc.

Otra observación importante, es que debido a que el motor de reconocimiento discrimina la salida que genera con base en los elementos de la gramática especificada, tiende a presentar errores de reconocimiento cuando dicha gramática contiene elementos cuyo sonido es demasiado similar. El sonido que representa un fonema esta clasificado con base en los factores que intervienen en su formación [16], por tanto entre mayor sea la similitud en la forma como son generados, mayor será la dificultad para diferenciarlos por parte del módulo de reconocimiento. Un ejemplo de este hecho, se observa rápidamente en las pruebas realizadas con las vocales, donde el sonido de la /e/ y la /i/ tienden a generar un error de reconocimiento por parte del sistema.

6.2 Pruebas al sistema como herramienta de apoyo

Para las pruebas de evaluación al sistema como herramienta de apoyo al proceso de rehabilitación de personas con dislalia, se llevaron a cabo sesiones con cada uno de los tipos de actividades que presenta el sistema. Se realizaron 9 sesiones de pronunciación con tres niños en un rango de edad entre cuatro a seis años, específicamente una niña de 4 años, y dos niños de 5 y 6 años respectivamente. Cada uno de ellos llevo a cabo actividades de un nivel distinto. Al igual que en las otras pruebas, un niño efectuó actividades de nivel A, otro niño llevo a cabo actividades de nivel B y el otro niño ejecutó actividades de nivel C. Cada uno de los niños realizo 3 sesiones con 5 actividades para cada sesión.

Es apropiado mencionar que las actividades de nivel A y B son prácticamente de refuerzo para poder llevar a cabo las actividades de nivel C, puesto que es en el nivel C donde realmente se realizan las actividades que permiten practicar la correcta práctica de los fonemas o sílabas que presentan dificultad al usuario dentro de palabras muy específicas. Por ello se buscó que los niños empleados en estas pruebas al sistema presentaran las características necesarias para poder determinar si el sistema cumple con su objetivo principal a través de la ejecución de las actividades de pronunciación.

La niña asignada a actividades de nivel A, tiene un conocimiento intermedio del alfabeto, razón por la cual se buscó observar si por medio del uso del sistema puede reforzar la asociación mental entre un símbolo del abecedario y su correspondiente fonema, los elementos seleccionados para esta prueba fueron: b, v, i, y, c. El niño asignado al nivel B, posee un correcto conocimiento del alfabeto y los grupos silábicos que se pueden formar con las consonantes, en este caso se buscó observar la efectividad del sistema para practicar y mejorar el aprendizaje de la correcta pronunciación del grupo silábico conformado por la consonante /g/. Se eligió este grupo silábico por tener la peculiaridad de que en las sílabas: ga, go, y gu, el sonido de la consonante es suave, mientras que en las sílabas: ge y gi, la consonante presenta un sonido fuerte. El niño asignado al nivel C, presenta dislalia evolutiva por sustitución de sonidos entre /r/ y /r/. La frecuencia de sustitución en los sonidos es moderada, sin embargo se trata de un caso que nos permite observar el uso del sistema.

Se observa que en todos los casos se presenta una tendencia positiva en cuanto a las evaluaciones obtenidas entre cada sesión. En las pruebas de nivel A se observó

una mejora promedio del 13% entre cada sesión, para las pruebas de nivel B se observó una mejora promedio del 6% entre cada sesión, finalmente en las pruebas de nivel C se observó una mejora promedio del 6% entre sesiones. La descripción tabular de estas observaciones, se muestra en la Tabla 3.

Tabla 3: Tabla de evaluación entre sesiones de prueba.

Nivel	Evaluación entre sesiones.
A	+ 13%
B	+ 6%
C	+ 6%

A pesar de que los resultados obtenidos son positivos, será necesario observar durante un tiempo mayor los resultados que produce en la capacidad de pronunciación, la aplicación del sistema en las aulas de educación especial para poder realizar una evaluación más completa del sistema. También es necesario recordar que el sistema por si mismo nunca podrá sustituir el apoyo de expertos en el tratamiento de la dislalia, y que el propósito principal del sistema es servir únicamente como herramienta de apoyo.

7 Conclusiones

De acuerdo a las Tablas 2 y 3, se concluye que el software diseñado e implementado puede ayudar progresivamente a niños con déficit del habla por Dislalia en sus procesos de enseñanza y/o rehabilitación. Sin embargo su nivel de eficiencia es directamente proporcional al buen desempeño del módulo de reconocimiento automático del habla que el sistema use y a la capacidad de las actividades desarrolladas que están tengan de captar y mantener la atención del usuario.

Dentro de las herramientas para el reconocimiento automático del habla consultadas, la selección del API de Java Speech como herramienta de desarrollo se justifica por su fácil integración con todas las demás herramientas de desarrollo y su buen desempeño como se observa en las pruebas hechas al módulo de reconocimiento. En cuanto al desarrollo de actividades de pronunciación atractivas a los usuarios se observó que la opción más recomendable para el desarrollo de actividades es por medio de animaciones. La opción más viable fue el SDK de JavaFX, cuya versión 2.0 se presenta como una herramienta completa y robusta para dicho fin.

El desarrollo de un software con fines de apoyo a procesos de rehabilitación de la dislalia, requiere de un amplio soporte de documentación y expertos en el área, pues no es posible plantear actividades de apoyo si no se cuenta con un fundamento teórico que respalde su propósito específico. Lo anterior se concluye de toda la información consultada y recabada sobre este trastorno del aprendizaje y los temas relacionados con el mismo, además de la asesoría proporcionada por los profesores de la escuela de educación especial No. 2 de Tlaxcala. Finalmente se concluye que la metodología de desarrollo basado en prototipos es una opción viable para la

generación de software de apoyo a proceso de rehabilitación, puesto que el entendimiento completo del sistema, tanto por parte de los usuarios finales, como de los desarrolladores fue más fácil de obtener mediante una constante validación de los procesos por medio de prototipos.

Referencias

1. Magaña, R., Lazaro, P.: Trastornos específicos del aprendizaje. *Sociedad Española Española de Pediatría Extrahospitalaria y Atención Primaria (SEPEAP)*. Mayo de (2010).
2. Gómez, P., Bernal, P., Bobadilla, J.: Reconocimiento de voz y fonética acústica. Alfaomega Ra-Ma, España, (2000).
3. Huang, X., Acero, A., Hsiao-Wuen J.: Spoken Language Processing, A Guide to Theory, Algorithm, and System Development. Prentice Hall International Inc, 133 -200. Upper Saddle River, New Jersey, U.S.A. (1994).
4. Lawrence, R., Biing-Hwang, J.: Fundamentals of Speech Recognition. Prentice Hall International Inc, 69 -139. Upper Saddle River, New Jersey, U.S.A. (1993).
5. Jurafsky, D., Martin, J.: Speech and Language Processing. Prentice Hall International Inc. 233- 281. Engewood Cliffs, New Jersey, U.S.A, (1999).
6. CECAPROIN.: SPEECHVIEWER III - Visualizador Fonético. (2010). <http://www.adaptat.com/productos/Comunicacion/speevi.htm>.
7. Consuelo, B.: VISHA (Visualización del habla) (2010). <http://www.uv.es/bellochc/pdf/visha.pdf>
8. Juaresti, J.: Tarjetas reconocedoras de voz para la intervención logopédica. II Congreso de CIIEE 2000, (2000).
9. García, L., Suárez, S.: Análisis de la voz, aplicaciones en sistemas de entrenamiento y rehabilitación. VIII Congreso Internacional de Informática en la Educación. (2002).
10. Oracle, Corp.: Java SE Overview - at a Glance. <http://www.oracle.com/technetwork/java/javase/overview/index.html>
11. Sun Microsystem, Inc.: Java Speech API Programmer's Guide. 35-36. (1998).
12. Calderón, M., Davis, E.: Swing, la solución actual de Java para crear GUIs. (2012). <http://www.dcc.uchile.cl/~lmateu/CC60H/Trabajos/edavis/swing.html>
13. Oracle, Corp.: What is JavaFX? (2012). <http://docs.oracle.com/javafx/2.0/overview/jfxpub-overview.htm>
14. SQLite, Consortium.: About SQLite. (2012). <http://www.sqlite.org/about.html>
15. JFree.org.: Welcome To JFreeChart! (2011). <http://www.jfree.org/jfreechart/>
16. García, L.: Transformada Wavelet Aplicada a la Transformación en Señales de Voz. PhD thesis, Departament de Teoria del Señal i Comunicacions, U. P. Catalunya, España (1998).