

Adaptación a Jclic para alumnos con deficiencia motriz, mediante comandos por voz.

M. Lucrecia Moralejo

III LIDI, Facultad de Informática. Universidad de La Plata, La Plata, Buenos Aires, Argentina.

lmoralejo@lidi.info.unlp.edu.ar

Stefania Ostermann

Facultad de Informática. Universidad de La Plata, La Plata, Buenos Aires, Argentina.
stefaniaostermann@gmail.com

Cecilia V. Sanz

III LIDI, Facultad de Informática. Universidad de La Plata, La Plata, Buenos Aires, Argentina.

csanz@lidi.info.unlp.edu.ar

Resumen

En el presente trabajo se propone la adaptación de un software educativo mediante comandos por voz, para alumnos con deficiencia motriz y que no se han visto afectados en el habla. Forma parte de un trabajo de fin de carrera realizado en la Facultad de Informática de la UNLP.

Durante este proceso se analizaron algunos programas educativos y software adaptativos. Además se realizaron pruebas de integración de varios software adaptativos estudiados con Jclic, para analizar la ayuda que pueden ofrecer a los alumnos con algún tipo de deficiencia motriz en la resolución de actividades.

Además, se estudiaron diferentes motores de reconocimiento de voz (RV), y sus fundamentos teóricos. Se profundizó el análisis del motor de RV Sphinx-4.

Antes de realizar el prototipo de adaptación, se hizo un estudio de la arquitectura de diseño y desarrollo de la herramienta educativa elegida, Jclic. Finalmente, se obtuvo el desarrollo de un prototipo de adaptación del programa Jclic, con la integración de Sphinx-4 para proveer RV, en particular, para las actividades de asociación simple. Se presenta aquí parte de este trabajo realizado y los resultados y conclusiones obtenidas, luego de la evaluación del prototipo.

Palabras claves: TIC, software educativo, deficiencia motriz, reconocimiento por voz, ayudas técnicas.

Introducción

Actualmente, existe una gran cantidad de software orientado a la educación en sus distintos niveles. Muchos de ellos han sido adaptados o creados teniendo en cuenta la diversidad de alumnos, pero otros son sólo herramientas estándares que no brindan adaptación alguna, por lo que están destinados a un conjunto restringido de alumnos. Las personas que están afectadas de algún tipo de disminución motriz, suelen influenciar alguna de las habilidades básicas, relacionadas con la percepción (visual, auditiva y táctil), la comunicación, el desplazamiento y/o la manipulación. En consecuencia, se encuentran con numerosos obstáculos y barreras que les impiden el desarrollo de habilidades, la ejecución de actividades, la relación con las personas y el entorno, etc. Para las personas con necesidades especiales, la mera utilización de las TIC puede representar la consecución de un elevado grado de autonomía en su vida personal.

Algunos de los aspectos que se pueden abordar utilizando programas de computación son: el enfoque de la

atención, el estímulo visual, la disminución de la sensación de aislamiento, el entretenimiento, el control del ambiente, entre otras [1].

Una de las razones de la escasa implantación de las TIC en la educación especial es la diversidad y la especificidad de las necesidades. Su utilización como herramientas, en este campo, requiere desarrollos muy complejos y variados, algunos personalizados, que además van a ser utilizados por colectivos poco numerosos.

La situación actual presenta grandes retos a superar para que una persona con disminución pueda estar en una posición de igualdad de condiciones, respecto de las demás personas. Por lo tanto, se debe adaptar adecuadamente el entorno y utilizar ayudas técnicas, que permitan eliminar al máximo de barreras, que dificultan que una persona con discapacidad pueda interactuar en él. Y sin embargo, éste ha sido el punto principal de conflicto: las personas con disminución motriz no disponen con frecuencia de las ayudas técnicas y adaptaciones necesarias para interactuar en un entorno hostil. Las personas con gran compromiso motriz tienen poca autonomía para la realización de tareas, necesitando de una forma constante, la llamada ayuda de terceros, para cambiar de tareas, acceder a información, desplazarse, etc. En la computadora este problema

también se manifiesta, pues suele ser el profesional quien ayuda constantemente a la persona a realizar las tareas. Un mayor nivel de autonomía los haría sentir independientes, podría aumentar su autoestima y su autosuficiencia, y en consecuencia, hacerlos sentir útiles y capaces. También influiría considerablemente en su comportamiento, facilitando que puedan ser emprendedores y que puedan realizar actividades nuevas por iniciativa propia.

Por lo tanto, estas razones motivan el desarrollo de adaptaciones a un software muy utilizado en el ámbito educativo como es JClic, para facilitar su uso por parte de alumnos con dificultades motoras, y así incentivar su desarrollo intelectual.

Si bien se conoce la relación entre la discapacidad motora y la dificultad en el desarrollo del lenguaje, esto no se da en todos los casos. Este trabajo está destinado a las personas con problemas motores, pero sin consecuencias o con consecuencias leves en el desarrollo del lenguaje. Se pensó en este subconjunto de personas, ya que existen más variedad de ayudas técnicas para personas con discapacidad motriz mediante la utilización de diferentes partes del cuerpo y se considera que sería una buena alternativa, el uso de la voz, si la persona afectada se expresa oralmente

sin dificultades. Además, requeriría un menor esfuerzo si la persona pudiera usar la voz para manipular el ordenador y se evitarían las lesiones producidas por "esfuerzo repetitivo".

Elección de Jclic como herramienta a adaptar

Existen en el mercado, una gran cantidad de softwares con fines educativos.

Con el objetivo de tener un panorama variado de los programas disponibles, se analizaron algunos de ellos, los cuales ofrecen diferentes funcionalidades y además se podrían utilizar en diversos niveles educativos. Entre ellos, se encuentran JClic, Text toys, Hot Potatoes, Markin, Lim, y Wink, etc.

De estos programas analizados, se optó por adaptar la herramienta JClic, que es un entorno para la creación, realización y evaluación de actividades educativas multimedia.

Algunas de las características que llevaron a la elección de este software, fueron que está desarrollado bajo licencia GPL, lo que provee la oportunidad de contar con el código fuente del programa para poder estudiarlo y analizarlo. De esta manera, se puede llevar a cabo la integración propuesta. También es uno de los software más utilizados para realizar

actividades educativas (ya lleva años de uso en el ámbito), de manera que se creyó que sería interesante comenzar el desarrollo de un prototipo para ampliar la diversidad de usuarios de dichas actividades.

Otro punto fuerte en esta elección, es que es posible utilizar JClic en distintos sistemas operativos, tales como Windows, Linux, Solaris y Mac OS X. Esta cualidad viene dada porque JClic se encuentra enteramente desarrollado con tecnología Java, la cual es multiplataforma.

Características de JClic

Como se mencionó anteriormente JClic es un entorno para la creación, realización y evaluación de actividades educativas multimedia. El cual está compuesto por tres componentes fundamentales, *JClicAuthor* que permite crear, editar y publicar las actividades de una manera sencilla, visual e intuitiva, *JClicPlayer* para la resolución de las mismas (el cual puede presentarse como applet o como una aplicación JClic) y *JClicReports* que es el módulo encargado de recopilar los datos (tiempo empleado en cada actividad, intentos, aciertos, etc.), y presentarlos después en informes estadísticos de diversos tipos. Esta aplicación sirve para realizar diversos tipos de actividades educativas: rompecabezas,

asociaciones, ejercicios de texto, palabras cruzadas, etc. [2]. Las actividades, por lo general, no se presentan solas, sino empaquetadas en proyectos. Un proyecto está formado por un conjunto de actividades y una o más secuencias, que indican el orden en que se han de mostrar.

Pruebas de integración de JClic con diferentes ayudas técnicas

Con el objetivo de obtener un mayor conocimiento de las posibilidades presentadas a un alumno con deficiencia motriz para resolver una actividad de JClic, se realizaron algunas pruebas del uso de diferentes softwares de autoayuda en conjunto con JClic.

Los programas de autoayudas se definen como un grupo de aplicaciones o utilidades que se cargan en la memoria de la computadora antes que el programa principal. Su propósito es facilitar el uso de aplicaciones informáticas de tipo general a los usuarios con discapacidad que, de otra forma, se verían obligados a usar sólo programas específicamente diseñados para ellos [3].

Algunos ejemplos de este tipo de autoayudas son: Mouse Por Barrido, Emuclic, Kanghooru y Ratón Facial.

Para estas pruebas se utilizaron sólo Emuclíc, Mouse por barrido y ratón facial.

Emuclíc convierte la acción de un switch (conmutador) estándar, o el sonido captado por un micrófono, en una o varias combinaciones de teclas o clics del mouse [4].

El objetivo del programa es emular diversas funciones del teclado o del mouse, de manera que cada vez que se accione el switch o se genere sonido, la computadora interprete que se ha presionado una determinada tecla, realizado un clic con el mouse, etc. De esta forma, cualquier programa que espere la presión de esa tecla reaccionará a la acción del conmutador o el sonido.

Mouse Por Barrido (MPB) es un software de distribución gratuita, que corre bajo Windows. La versión 3.0 de este programa fue liberada por Antonio Sacco¹. Está pensado especialmente para personas que por problemas motrices no pueden utilizar un mouse convencional con precisión. El objetivo del programa es permitir al usuario el manejo total del puntero del mouse mediante la sola presión de un switch

(o pulsador) estándar. Trabaja en modo residente, sin interferir con la ejecución de otros programas, y está destinado a funcionar en conjunto con otras aplicaciones. Una vez que se ejecuta queda funcionando en modo residente.

Ratón Facial es un programa que, unido a una webcam USB estándar, permite controlar totalmente el ordenador sin las manos, mediante ligeros y suaves movimientos de alguna zona del cuerpo previamente configurada [5]. Se puede utilizar como sustituto o complemento del ratón convencional. Está especialmente diseñado para personas con discapacidades motrices en manos o brazos. Este software le proporciona al usuario una solución efectiva y fiable para la interacción con el ordenador.

Pruebas realizadas

Para llevar a cabo estas pruebas de integración, se generó una actividad JClic de prueba, la cual consta de una asociación simple donde se presentan dos grillas: una con sumas y otra con los resultados. El usuario deberá unir adiciones a su respectivo resultado. La forma de resolver esta actividad es la siguiente: el usuario hace clic en una celda, mueve el mouse hacia una celda de la otra grilla y hace clic, si la unión que eligió es correcta (la suma que se muestra en la celda de una de las

¹ Antonio Sacco - Graduado en Ingeniería en Sistemas de Información, en la Universidad Tecnológica Nacional. Especialista en Informática Aplicada en la Educación. Ha realizado varios aportes en el área de tecnología aplicada a la educación nivel internacional. Algunos de ellos son: Emuclíc, MPB (Mouse Por Barrido), entre otros. En <http://www.antoniosacco.com.ar/> puede encontrarse más información.

grillas da como resultado el número que se muestra en una celda de la otra grilla), JClic marca esas celdas para que no se puedan volver a elegir y otorga un punto de acierto al usuario. Se debe repetir esta serie de pasos hasta que todas las celdas se unan quedando resuelta la actividad.

Se describen aquí sólo dos de las pruebas de integración realizadas:

Integración de Mouse por barrido

Fue posible realizar la actividad utilizando MBP, ya que este software provee el manejo completo de la funcionalidad de un mouse a través de un switch. Permite seleccionar las celdas necesarias, a través de las pulsaciones del switch, para resolver la actividad. Cuando el programa residente muestra la imagen de la funcionalidad que se necesita (por ejemplo, clic con el botón izquierdo) se debe pulsar el switch.

La desventaja o dificultad que se experimentó mientras se resolvía la actividad de prueba, fue que resultó tediosa la espera de la imagen que se quería ejecutar. Además, si por ejemplo se detenía el mouse a muy poca distancia de la celda que se quería presionar, avanzar esa mínima distancia resultaba complicado, y era más sencillo alejarse de ella y volver a

intentar posicionar el cursor con otro movimiento.

El tiempo que se necesita para resolver la actividad es bastante mayor al que llevaría resolverla sin ninguna adaptación.

Integración con ratón facial

Fue posible realizar actividades con JClic a través del uso de ratón facial, ya que este software provee el manejo del mouse, a través de una cámara web que lo reemplaza, pudiéndose realizar dichas actividades sin problemas. Las pruebas se realizaron configurando el uso de clic por espera, es decir, que se debe mantener el mouse sin moverse por cierta cantidad de tiempo para que el ratón ejecute el clic.

En principio se notó que la precisión del ratón depende de la cámara web utilizada y de la calidad del video, por lo que a veces puede dificultar su utilización.

En particular, para MBP y Ratón Facial, de ser necesaria la utilización del teclado, para realizar actividades como las de texto es necesario utilizar un teclado virtual para complementar la funcionalidad.

Si bien es posible utilizar algunos de estos softwares junto con JClic, se vio al reconocimiento de voz como una alternativa viable, la cual apunta al uso del habla cuando sea posible para el

usuario, siempre y cuando éste se desenvuelva oralmente sin dificultades o con dificultades leves. Se considera importante destacar, que es necesario introducirse en el área de reconocimiento de voz para facilitar el uso de la computadora para las personas con problemas motrices, ya que con el uso de la voz se evita el esfuerzo físico que se genera al intentar usar ayudas técnicas como el switch.

Particularidades de las personas con dificultades motrices

La discapacidad motriz podría definirse como la pérdida o restricción de la capacidad de movimiento, desplazamiento y equilibrio de todo o de una parte del cuerpo. Se entiende como la alteración o deficiencia orgánica del aparato motor o de su funcionamiento, que afecta al sistema óseo, articulario, nervioso y muscular. Según la causa de la discapacidad, y el grado de ésta, pueden existir además restricciones para el desarrollo del lenguaje, o de la manipulación de objetos, siendo oportuno señalar que este tipo de restricciones no tiene por qué afectar a los procesos cognitivos o de aprendizaje, aunque en ocasiones, puede presentarse un enlentecimiento en algunas actividades y habilidades, como la comunicación, la

lectoescritura, o el desarrollo de actividades de autocuidado [6].

A nivel educativo, se plantea el problema en cuanto a la intervención centrada en sus posibilidades de aprendizaje y potencialidades educativas, y no en el déficit que presenta el alumno.

Desde el punto de vista de la discapacidad física, el ordenador, a modo de prótesis, abre posibilidades hasta ahora poco imaginables. Personas con un pequeño resto de control motriz pueden llegar a comunicarse con los demás, ocupar su tiempo de ocio de forma gratificante, realizar cursos de formación e integrarse laboralmente. El ordenador, ha ayudado a cambiar la idea de que muchas personas con discapacidades físicas graves deberían estar recluidas [3].

Hoy en día, existen muchos fundamentos teóricos sobre la capacidad de las ayudas técnicas e informáticas para adaptarse a los formatos de la actividad escolar e incidir positivamente en el proceso de enseñanza y aprendizaje de los niños y jóvenes con necesidades educativas especiales, sin embargo, al descender a la práctica, se ve que no es sencillo encontrar las TIC adecuadas a cada necesidad [7]. Para disminuir la brecha existente entre las personas con necesidades especiales y la tecnología,

se utilizan lo que se conoce como ayudas técnicas. Estas ayudas, facilitan el acceso al ordenador, mediante adaptaciones que se usan como complemento para utilizar el teclado o el mouse, además se existen dispositivos que se pueden utilizar para sustituir a algunos de los comúnmente usados. Es decir, existen ayudas técnicas para acceder a las interfaces estándar (varilla bucal, sujeta teclas, licornio, etc.), y otras para sustituirlas (Joystick, Switch, teclado virtual, etc). Además de estas, existen otras tecnologías consideradas “ayudas técnicas”, tal como es el caso del reconocimiento de voz.

Con el uso de esta tecnología, se podría utilizar cualquier programa, dándole las órdenes a la computadora verbalmente mediante letras, palabras, o frases. La interacción resulta más eficaz que si utiliza, por ejemplo, el lento proceso de comunicarse a través de un conmutador siguiendo la técnica de barrido.

Si bien se ha mencionado la relación entre la discapacidad motora y la dificultad en el desarrollo del lenguaje, esto no se da en todos los casos. Este trabajo, como ya se mencionó, está destinado a las personas con problemas motores, pero sin consecuencias o con consecuencias leves en el desarrollo el lenguaje.

Ayuda técnica seleccionada: Reconocimiento de voz

El avance tecnológico ha aportado al ser humano nuevas y mayores posibilidades de desarrollar un modo de vida más completo, pero al mismo tiempo exige continuamente nuevos y específicos conocimientos y habilidades en el individuo para poder hacer uso de las posibilidades que le ofrecen. En las personas con algún tipo de discapacidad, la progresiva complejidad del medio social puede tener, sin embargo el efecto contrario al buscado por el progreso social [8].

Así se encuentra en el reconocimiento de voz una alternativa para la comunicación con la computadora, permitiendo que las personas con discapacidades motoras que no pueden acceder al teclado estándar y al mouse puedan, con el habla, realizar acciones que sin esta tecnología no le serían posibles, en otras palabras, el objetivo es convertir el habla humana en acciones interpretables por la computadora.

Esta tecnología, es una parte de la Inteligencia Artificial, que tiene como objetivo permitir la comunicación hablada entre seres humanos y computadoras electrónicas, es decir, es el proceso de conversión de un mensaje hablado a texto, que permite al usuario una comunicación con la

computadora. El problema que se plantea en un sistema de RV es el de hacer cooperar un conjunto de informaciones que provienen de diversas fuentes de conocimiento (acústica, fonética, fonológica, léxica, sintáctica, semántica y pragmática), en presencia de ambigüedades, incertidumbres y errores inevitables para llegar a obtener una interpretación aceptable del mensaje acústico recibido [9]. Un sistema de reconocimiento de voz es una herramienta computacional capaz de procesar la señal de voz emitida por el ser humano, y reconocer la información contenida en ésta, convirtiéndola en texto o emitiendo órdenes que actúan sobre un proceso. En su desarrollo intervienen diversas disciplinas, tales como: la fisiología, la acústica, el procesamiento de señales, la inteligencia artificial y la ciencia de la computación.

Existen algunos componentes de gran importancia para los sistemas de RV, que son: el diccionario, la gramática, el modelo acústico y el modelo de lenguaje. Donde el diccionario representa el conjunto de palabras o sonidos a reconocer. A diferencia de un diccionario normal, cada entrada no tiene por qué ser una única palabra. Estas pueden ser tan largas como una oración o dos. Los vocabularios pequeños pueden tener una o dos sonidos a reconocer, mientras que los

vocabularios muy grandes pueden tener cientos de miles o más. La gramática se define a partir de las palabras que debe aceptar la aplicación, y puede estar dada a través de un estilo similar a BNF.

El modelo de lenguaje puede ser abordado a través de modelos estadísticos (Statistical Model Language - SLM) o utilizando gramáticas de estado finito (Finite Grammar State - FGS) [10]. Un modelo estadístico captura la probabilidad de las palabras y de las secuencias de palabras. Es utilizado en el decodificador para limitar la búsqueda, y generalmente, hace una contribución significativa a la exactitud del reconocimiento. Un buen modelo es aquel que modela con precisión la entrada esperada. Se caracteriza por su orden, en términos de "n-gram", donde "n" indica el tamaño de la ventana sobre la cual se computan las estadísticas. En general cuanto más grande sea "n" más preciso será el modelo. También, a mayor "n" se necesitan más datos para asegurar que las estadísticas se estiman sólidamente. Una gramática de estados finitos define las posibles palabras, así como también el posible orden de dichas palabras.

Un modelo acústico se crea a partir de grabaciones, sus respectivas transcripciones, y el uso de software para crear representaciones

estadísticas de los sonidos que componen cada palabra [11]. La performance del reconocimiento producida por el modelo acústico puede mejorarse aún más, mediante un modelo de lenguaje, el cual contribuye a evitar ambigüedad entre varias palabras similares producidas por el modelo acústico.

Para la elección de la herramienta a utilizar, se analizaron distintos softwares de reconocimiento de voz, entre ellos Loquendo, Xvoice, NicoTollkit, Sphinx y Dragon Naturally Speaking. Se estudiaron sus principales características, funcionalidades y requisitos.

De los analizados, se optó por Sphinx, particularmente la versión 4. Es un sistema desarrollado en la Universidad de Carnegie Mellon (CMU). Este framework es un sistema basado en los Modelos Ocultos de Markov (HMM, Hidden Markov Model), por lo que, para su funcionamiento primero debe aprender las características (o parámetros) de un conjunto de unidades de sonido, y luego utilizar lo que ha aprendido de estas unidades para encontrar la secuencia de unidades de sonido más probable para una señal de voz dada. Se optó por esta herramienta, ya que es ampliamente utilizada por investigadores y desarrolladores que se dedican al área del reconocimiento de

voz, y en consecuencia, se encuentra en constante desarrollo y actualización. Por las características de su licencia, es posible utilizarlo libremente en cualquier desarrollo e investigación. Además, también se puede obtener su código fuente, en caso de que sea necesaria alguna modificación o estudiar su funcionamiento a bajo nivel. Está completamente desarrollado con tecnología Java, al igual que JClic. De manera que servía al propósito de integrar ambos componentes sin dificultades provocadas por incompatibilidad de lenguajes. Además, ha sido diseñado con un alto grado de flexibilidad y modularidad, donde cada elemento del sistema puede ser fácilmente reemplazado o modificado. A través del Configuration Manager, es que el framework brinda la posibilidad de cargar y configurar los distintos módulos dinámicamente, en tiempo de ejecución. Determinando así, que componentes van a ser usados y la configuración particular de cada uno de ellos. En particular, es posible indicar el diccionario y la gramática que se utilizaran durante el reconocimiento. A continuación, se presenta la propuesta particular para este trabajo.

Propuesta de adaptación

La adaptación propuesta ha abordado la modificación de las actividades de JClic de manera que se puedan

resolver a través de la utilización de comandos por voz. Para ello se tomó, inicialmente, la actividad del tipo asociación simple.

En este caso de actividad, que JClic permite crear, el usuario tiene que descubrir las relaciones existentes entre dos conjuntos de información. Es decir, se presentan dos grupos de datos que tienen el mismo número de elementos, donde a cada elemento del origen le corresponde un elemento del destino. Es por ello que se la denomina simple, a diferencia de la asociación compleja, donde a cada elemento del origen puede corresponderle 0, 1, o más elementos del destino.

Como primera medida para llevar a cabo esta integración se debieron tomar algunas decisiones, las cuales se detallan a continuación.

Etapa 1: Análisis

Una de las decisiones que se consideró fue, cómo tomar conocimiento de que se desea realizar la actividad utilizando comandos por voz.

Se consideró que en esta situación, el usuario deba contar con la asistencia del docente, ya que es éste quien toma la decisión para cada alumno en particular, si es adecuado o no utilizar RV en la resolución. El programa, para ello, muestra un mensaje en pantalla al momento de comenzar la actividad.

Este pide que se indique si se desea utilizar reconocimiento de voz.

Otra cuestión de suma importancia, ha sido decidir qué mecanismo proveer para identificar cada elemento de la pantalla que presente interactividad, con el fin de resolver la actividad. Para esto se analizaron diferentes posibilidades. Esta identificación que utiliza el usuario para nombrar un elemento se la denominará etiqueta, de aquí en más.

En primer instancia, se pensó en utilizar las letras del alfabeto como etiquetas, pero al momento de llevarlo a la práctica, se encontró la dificultad de que ciertas letras, tales como la “b” y la “d”, eran muy similares en su pronunciación, por lo que la tasa de aciertos del reconocedor disminuía considerablemente.

Por otro lado, si se ampliaba el número de casilleros a utilizar, resultaba más natural usar combinaciones de dígitos (por ejemplo 10) que utilizar letras (por ejemplo ab). También, se debían utilizar letras alternadas, quitando del diccionario del RV, aquellas que causaban conflictos como los ya mencionados o aquellas que resultaban muy complejas en cuanto a su pronunciación (por ejemplo, el caso de la letra ‘r’). Considerado esto, se decidió la posibilidad de utilizar números para la creación de las etiquetas. Esta solución presenta

ciertas ventajas, respecto a la planteada anteriormente.

Además, se decidió realizar las adaptaciones necesarias para evitar dificultades de pronunciación de ciertos números. Para esto, se tuvieron en cuenta otras palabras alternativas a la correcta, por ejemplo, se admite que el usuario diga “tes” en lugar de “tres”, “tinco” en lugar de “cinco”, “acetar” en lugar de “aceptar”, entre otras.

Si bien esta decisión implica un diccionario de mayor tamaño, presenta consecuencias positivas en cuanto al aumento de usuarios que podrían utilizar el prototipo. Así, se intentó lograr un equilibrio entre performance de la aplicación y usabilidad del producto.

El segundo tema a resolver fue el de conocer cuándo el usuario termina de nombrar los dos elementos a unir. Para ello, se pensó en utilizar palabras “nexo”. Por ejemplo, “uno con tres aceptar”; lo que se interpreta de esta sentencia es lo siguiente: el primer número representa un casillero del primer conjunto de información, la palabra “con” (nexo) indica que se va a nombrar el casillero del segundo conjunto, representado por el segundo número de la frase. La palabra “aceptar” indica que el usuario quiere realizar la unión de los casilleros nombrados.

También, con respecto a las etiquetas, se debió pensar en qué momento

agregar el código necesario para que se inserte la etiqueta en el componente que representa al casillero con la información. Cabe mencionar, que se generan cuando se ejecuta JClicPlayer, sólo si se indica que se desea realizar la actividad con comandos por voz. Esto implicó una decisión, ya que había que mantener la presentación de la información de los dos conjuntos en forma aleatoria, de manera que no apareciera la actividad resuelta, a causa de las etiquetas.

Finalmente, se agregó el código necesario de manera que la aplicación muestre un mensaje pidiendo confirmación de lo dicho por el usuario. Así, cuando éste nombra los casilleros que desea unir, el programa presenta un mensaje mostrando las palabras reconocidas. Para dar confirmación positiva al mensaje, se debe decir “aceptar”, y en caso contrario, “cancelar”. A continuación, se presenta la segunda etapa de trabajo, que ha sido decidir (y llevar a la práctica) cuestiones vinculadas al motor de RV.

Etapas 2: Configuración de Sphinx-4

En primer lugar, para utilizar Sphinx, se debe descargar la aplicación desde el sitio oficial [12]. Allí está disponible el código fuente de la herramienta, aunque si no se desea modificar código (como en nuestro caso), alcanza con

incluir el archivo .jar en la aplicación donde se va a integrar.

Actualmente, Sphinx-4 dispone de modelos que han sido creados utilizando SphinxTrain (herramienta que provee para el entrenamiento), y puede descargarse desde el sitio de cmusphinx.org.

En un principio, se pensó como una alternativa válida crear el diccionario utilizando el modelo WSJ_8gau_13dCep_16k_40mel_130Hz_6800Hz que viene incluido con la distribución de Sphinx-4 y, si bien está entrenado para el idioma en inglés, reemplazando los fonemas puede reconocer español. Existen trabajos revisados del área de RV, que realizan este tipo de solución².

Si bien los fonemas pertenecen al idioma inglés, en un primer momento, fueron utilizados para generar el diccionario para la integración con JClic.

Esta solución fue parcialmente válida, ya que el reconocedor funcionaba con un alto porcentaje de acierto. Pero, a pesar de esto, se encontraron dos falencias. Por un lado, había errores en la precisión del reconocedor en ambientes ruidosos. Esto sería un problema en los casos en que la adaptación fuera utilizada en escuelas,

donde las aulas se comparten entre varios alumnos. Por otro lado, si se deseaba extender el diccionario y utilizar palabras con la letra “ñ”, no existían fonemas en el idioma inglés que lo represente.

A partir de estas conclusiones, se decidió utilizar un modelo basado en el idioma español. Luego de investigar sobre el tema, surgieron dos alternativas viables. Por un lado, se podía entrenar el reconocedor, utilizando la herramienta SphinxTrain, y por otro, utilizar modelos ya entrenados y testeados. En el presente desarrollo se optó por utilizar un modelo ya entrenado, pero se hicieron además algunas pruebas con el entrenador, de manera tal, de entender y estudiar su funcionamiento.

Para esto se utilizó un modelo ya entrenado, disponible en la web para su libre utilización. El proyecto se llama Diálogos Inteligentes Multimodales en Español (DIME), dentro del cual hay más de un modelo acústico. El modelo elegido para este trabajo recibe el nombre de DIMEx30-T22 [13].

A partir de esta lista de unidades fonéticas se creó el diccionario a utilizar en la integración con JClic. Podría haberse utilizado el diccionario, tal cual lo presenta DIMEx30, pero había palabras que no se encontraban en él, por lo que se optó por redefinirlo, respetando las unidades fonéticas presentadas. Respecto del modelo de

²Entre ellos, se consultó el proyecto Mouse Advanced GNU Speech (Magnus): <http://magnusproject.wordpress.com/>

lenguaje, la definición del modelo acústico y su arquitectura, se respetó el proporcionado por DIMEx30.

Para incorporar estos archivos a la aplicación JClic, se debió crear en primer lugar un archivo .jar que, por convención, debía respetar la estructura de directorios de los modelos provistos por Sphinx-4.

Luego de armado el archivo .jar, se incluyó en el classpath de la aplicación. También, se debió configurar Sphinx-4 para incorporar los nuevos archivos del modelo acústico, el diccionario, la gramática y el modelo de lenguaje. Esto se realizó a través del archivo de configuración (Configuration Manager). En la siguiente sección, se detallan cuestiones referidas al desarrollo del prototipo.

Etapas 3: Desarrollo del prototipo

En esta sección se describirán aspectos del prototipo que incluyen a ambos componentes utilizados para la integración. Una de ellas, es cómo se realizó la incorporación del framework de reconocimiento de voz a JClic. Para ello, se creó una clase en JClic que representa al reconocedor, llamada VoiceRecognizer, donde se encuentran sus principales métodos, tales como el método que se utilizó para crearlo, así como también, el método que se encarga de realizar el reconocimiento. Se generó un paquete llamado

“reconocimiento” dentro del paquete “src” de JClic. Luego, esta clase, es utilizada en el método constructor de la clase Player, si el usuario eligió trabajar con reconocimiento de voz. Allí se crea el reconocedor y se puede comenzar a utilizarlo.

También, se hizo que la clase que representa al reconocedor heredara de SwingWorker aunque no utilice Swing, de manera que JClic y el reconocedor se ejecuten en hilos separados, interactuando entre ellos, para paralelizar tareas. De esta manera, ambos componentes, pueden ejecutarse sin problemas.

Para llevar a cabo la tarea de resolver una actividad de tipo Asociación Simple, lo que se implementó fue, que al crearse, el reconocedor ejecute un método llamado getCommand() en la clase que representa dicha actividad. Este método es el encargado de procesar la entrada de voz del usuario y tomar las decisiones correspondientes.

Al recibir la entrada de voz “aceptar”, el sistema muestra un cartel con los valores que se van a procesar, el usuario deberá confirmar estos valores para que la acción se lleve a cabo.

Para la confirmación es necesario pronunciar nuevamente la palabra “aceptar”. Luego de confirmado, se invoca a un método que se encarga de ejecutar la acción que el usuario desea realizar. En este método se buscan los

casilleros nombrados, si existen y no fueron elegidos antes. Luego, se verifica dentro de la estructura interna del elemento, si forman una correspondencia correcta, es decir, si las celdas seleccionadas son parte de la solución. Si es así, se eliminan de los posibles elementos a elegir y se continúa con la próxima correspondencia, hasta llegar a la última. Cuando se llega a ésta, se da por terminada la actividad.

JClic provee un módulo capaz de contabilizar el tiempo empleado en cada actividad, intentos, aciertos, etc. Si bien el tiempo puede variar si se utiliza reconocimiento de voz, se pensó en mantener igualdad en el contador de intentos y aciertos para que el docente pueda evaluar al alumno que está resolviendo la actividad. Es por esta razón, que se decidió agregar un cartel donde el usuario vea y confirme que es lo que desea unir, ya que existe, en la mayoría de los reconocedores, cierta tasa de error, con lo cual, podría darse la situación en que se procese una entrada errónea y JClic lo contara como intento fallido, perjudicando la evaluación del alumno. Con los agregados mencionados, el docente que creó la actividad podrá usar el contador de errores que provee, por defecto, JClic.

El prototipo desarrollado al momento abarca, como se dijo, la resolución de las actividades de asociación simple.

Sin embargo, como parte de este trabajo, se ha hecho una propuesta de cuál sería la estrategia para extender el prototipo al resto de las actividades y se abordará esto en trabajos futuros. A continuación se presenta la evaluación realizada de las estrategias de integración planteadas al momento.

Evaluación

Se decidió someter el prototipo presentado a prueba de expertos (vinculados a las distintas áreas que se involucran en este trabajo) para que ellos expresaran sus opiniones respecto de este trabajo.

Se consideró más apropiado realizar primero este tipo de prueba, y analizar los resultados para tomarlos como líneas futuras de trabajo e investigación. Después de esta etapa será posible testear el prototipo con los usuarios finales, los cuales serían, docentes y alumnos. Esto se creyó importante, para no someter a los alumnos a situaciones de posibles fracasos propios del testeo del software y de la estrategia en sí misma. Además, esta metodología tiene como ventaja la calidad de la respuesta y el nivel de profundización por parte del experto. Este puede ofrecer sus reflexiones acerca del objeto a evaluar. Mediante el juicio de expertos, se pretende tener estimaciones razonablemente buenas, las mejores

conjeturas, en situaciones donde no se pueden o no es conveniente obtener cuantificaciones exactas [14]. Sin embargo, estas estimaciones pueden y deben ser confirmadas o modificadas a lo largo del tiempo, según se vaya recopilando información sobre el objeto de estudio.

Como instrumento de evaluación se eligió una encuesta con preguntas abiertas y cerradas, de manera tal de poder recoger la información que se cree necesaria para someter a juicio el prototipo.

Como conclusión de las encuestas realizadas a expertos, se considera que se ha realizado una buena elección del software educativo a adaptar, como así también, la utilización de comandos por voz como ayuda técnica. Como mencionó uno de los expertos, esta opción puede usarse de forma complementaria con otras herramientas y no necesariamente es mejor o peor que otra adaptación, sino que es una alternativa diferente, la cual abre un camino de nuevas posibilidades. Si bien unos pocos expertos se manifestaron acerca de la elección del motor de reconocimiento de voz, coincidieron en que la misma es acertada. El aspecto fundamental a resaltar es su disponibilidad y sus posibilidades en cuanto a funcionalidad. En el marco de este trabajo, se considera que, la utilización de sphinx-4 resultó conveniente,

rescatando también lo manifestado por los encuestados.

Finalmente, se analizó la estrategia planteada de solución, y los expertos manifestaron su acuerdo con la misma y presentaron algunas alternativas a tener en cuenta en trabajos futuros.

En la siguiente sección se detallan algunos de estos aspectos.

Conclusiones y líneas de trabajo futuras

En este artículo se ha realizado una presentación de los lineamientos generales de un prototipo de adaptación del software JClic, mediante el uso de reconocimiento por voz.

Como primer paso, se revisaron algunos programas educativos disponibles, junto con sus adaptaciones. Posteriormente al estudio de las herramientas, se seleccionó JClic para la implementación de la adaptación. También, se estudiaron diferentes ayudas técnicas, y se pusieron a prueba algunas de ellas junto con JClic, para analizar su funcionamiento.

Paralelamente, se abordó el tema de la discapacidad motriz, y lo que ella representa para las personas que la poseen, tanto a nivel cognitivo como social.

Se pensó en el reconocimiento de voz como una buena alternativa para implementar la adaptación deseada de

JClic. Se continuó entonces, con el estudio y análisis de diferentes sistemas reconocedores. Se realizaron consultas también a diferentes usuarios de estos sistemas para tomar contacto con sus bondades y dificultades. Luego, se profundizó en los aspectos referidos al reconocimiento de voz, de manera de poder contar con los conocimientos necesarios para lograr introducir esta tecnología, y llevar a cabo el desarrollo del prototipo.

Obtenida la base teórica, se pusieron en práctica estos conocimientos a través del uso y estudio de diferentes motores de RV. Se eligió y se profundizó el estudio de Sphinx 4. En particular, se tuvo que tomar conocimiento acerca de la estructura del motor, su forma de trabajo, y los modelos disponibles para el español. Así como también, aspectos vinculados a su configuración.

Como último paso, se implementó el prototipo y se realizaron encuestas a expertos para que expresen su opinión sobre la estrategia de solución del producto desarrollado.

Si bien se ha obtenido un amplio conocimiento de diferentes herramientas, tanto educativas como relacionadas con RV, quedan ciertas modificaciones, mejoras y extensiones a desarrollar en la adaptación presentada.

Como líneas de trabajo futuras, se plantean

- Realizar pruebas con alumnos y docentes del área.
- Permitir la configuración de las etiquetas (el docente podría elegir rotular cada casillero como lo crea más conveniente)
- Extender la implementación a todas las actividades disponibles en JClic, acorde a la estrategia planteada.

Referencias

- [1] Consultado en 2010.
<http://www.niee.ufrgs.br/eventos/CIIEE/2003/bloque2/comunicaciones/La%20utilizacion%20de%20Software%20de%20uso%20general%20y%20Aplicaciones%20esp.doc>.
- [2] Sitio de Jclic.
<http://clic.xtec.cat/es/jclic/index.htm>
- [3] Sánchez Montoya. "Ordenador y Discapacidad. Guía práctica a las personas con necesidades educativas especiales". Madrid, 2002
- [4] EmuClic –Consultado en 2010.
<http://www.antoniosacco.com.ar/emuclic.htm>
- [5] Ratón Facial – Consultado en 2010 -
<http://www.crea-si.com/esp/rfacial.php>
- [6] Guía de Apoyo a Personas con Discapacidad – Consultado en 2010 -
http://www.urjc.es/comunidad_universitaria/universidad_saludable/programa_discapacidad.html
- [7] Tecnologías Adaptadas – Consultado en 2010.
<http://tecnoadaptada.blogspot.com/>
- [8] Nuevas tecnologías y educación de personas con dificultades- Consultado en 2010
<http://www.tecnoneet.org/docs/2002/2-82002.pdf>

[9] Bernal Bermúdez, Bobadilla Sancho, Gómez Vilda – “Reconocimiento de voz y fonética acústica”. México, Alfaomega grupo Editor, 2000.

[10] Modelo del lenguaje. Sitio de Sphinx – Consultado en 2010 -

http://sphinx.subwiki.com/sphinx/index.php/Language_model

[11] Rocha Luis. “Sistemas de reconocimiento de voz”. Revista telegráfica electrónica. Agosto 1986. Pp. 1172-1180.

[12] Proyecto Sphinx – Consultado en 2010-<http://cmusphinx.sourceforge.net/sphinx4/>

[13] Modelo en español – Consultado en 2010-

<http://leibniz.iimas.unam.mx/~luis/DIME/recursos.html>

[14] Arquer “Fiabilidad humana: métodos de cuantificación, juicio de expertos”. Consultado en 2010.

http://www.insht.es/InshtWeb/Contenidos/Documentacion/FichasTecnicas/NTP/Ficheros/401a500/ntp_401.pdf