Revista RUEDES, Año 3- Nº 5- 2014, ISSN: 1853-5658, p. 85-101

Paradigmas de interacción hombre-máquina. Un análisis enfocado al ámbito de la educación especial

Lucrecia Moralejo[1]

Cecilia Sanz[2]

Patricia Pesado[3]

Universidad Nacional de La Plata

Resumen

Este trabajo se enmarca en el área de interacción hombre-máquina y los diferentes paradigmas que existe actualmente. Se revisan antecedentes y posibilidades vinculadas a la educación especial. Como caso de estudio, se presenta una propuesta de adaptación al software educativo JClic, mediante la utilización de comandos por voz, con el objetivo de ser utilizado por usuarios/alumnos con deficiencia motriz sin consecuencias o con consecuencias leves en el desarrollo del lenguaje. Como parte de esta propuesta de adaptación, se estudiaron diferentes motores de reconocimiento de voz (RV), y se profundizó el análisis del motor de RV Sphinx-4. Se presenta aquí parte de este trabajo realizado y los resultados y conclusiones obtenidas, luego de la evaluación del prototipo.

Palabras clave

Educación especial - Interacción hombre-máquina - JClic - Sphinx

[1] Lucrecia Moralejo[1], Licenciada en Sistemas, III LIDI, UNLP, lmoralejo@lidi.info.unlp.edu.ar

[2] Cecilia Sanz[2], Dra. En Computación, III LIDI, UNLP, csanz@lidi.info.unlp.edu.ar

[3] Patricia Pesado[3], Licenciada en Informática, III LIDI, UNLP, ppesado@lidi.info.unlp.edu.ar

Abstract

This work is part of the area of human-machine interaction and the different paradigms available today. Records are reviewed and possibilities related to special education. As a case study, we present a proposal to adjust the JClic educational software, using voice commands, in order to be used by users/pupils with motor impairment without consequences or mild implications in the development of language. As part of the proposed adaptation, we studied different speech recognition engines (RV), and deepened the analysis of the RV engine Sphinx-4. Presented here is part of this work, results and conclusions reached after evaluation of the prototype.

Keywords

Special education - Human Computation Interaction – jClic - Sphinx

Recibido: 30 de junio de 2012

Aceptado: 6 de agosto de 2013

Introducción

Actualmente, existe una gran cantidad de software orientado al ámbito de 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 discapacidad, 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 (Sancho, 1998).

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

Las TIC son herramientas que pueden utilizarse de forma creativa para mejorar el desarrollo de capacidades y destrezas, de las personas con necesidades especiales bajo una concepción interaccionista, que desplaza su enfoque desde las características individuales de los alumnos, a un modelo de apoyo curricular que actualmente se encuentra en proceso de expansión (Sánchez Montoya, 2002).

Interacción Hombre Ordenador

El área de Interacción Hombre Computador (o HCI: Human Computer Interaction), es la disciplina que se enfoca en el estudio de la interacción entre las personas y los sistemas computacionales. Su objetivo principal es mejorar esta interacción, haciendo que los sistemas computacionales sean más usables, de manera que aumente la productividad de las personas al trabajar con ellos. La Asociación de Maquinaria Computarizada (ACM) (ACM, 2012) define HCI como “La disciplina encargada del diseño, evaluación e implementación de sistemas computacionales interactivos para uso humano y del estudio de lo que los rodea”.

Pone énfasis así en HCI como una ciencia que analiza ambos aspectos: humano y computador, en conjunto. Esta es una de las razones primordiales por lo cual HCI es estudiada con un enfoque distinto, dependiendo de la ciencia. Desde el contexto humano, HCI es complementada por otras ciencias tales como: psicología, ciencias cognitivas, de la comunicación, de diseño gráfico e industrial, entre otras. En el contexto de computadores y maquinaria comprende: gráficos por computador, sistemas operativos, lenguajes de programación, y desarrollo de ambientes.

En (CSDL, 2008) se plantea el modelo conceptual de HCI que contempla cuatro elementos: (a) las personas; usuarios del sistema, (b) la tarea; diferentes pasos a realizar para llevar a cabo una o más actividades, (c) el ambiente; aspectos físicos, organizacionales y sociales del ambiente y, (d) la tecnología: cualquier artefacto con el cual se interactúa.

A diferencia de los aspectos del ambiente, y de la tarea; la interacción entre personas y tecnología se realiza por medio de un componente un tanto implícito: la Interfaz. Esta se conforma de varios componentes, entre estos se puede nombrar, Interfaces de hardware: teclado, mouse, touchpads, lápices, etc. e interfaces de software como la Interfaz Gráfica de Usuario (GUI).

Paradigmas de HCI

Los paradigmas de HCI se proponen responder a la necesidad de contar con interfaces lo más naturales posibles para el ser humano. Con el pasar del tiempo y con el avance de la tecnología, han aparecido cada vez mas paradigmas de interacción, ciertas formas de interacción que antes se consideraban paradigmas por separado se han unificado bajo un solo paradigma. Se presenta aquí un breve resumen de algunos paradigmas de interacción (Karray, Milad, Abou, Arab, 2008).

Visión como medio de interacción (Visual-based HCI): La visión como medio de interacción, es probablemente el área más extendida en la investigación de HCI. Teniendo en cuenta el alcance de las aplicaciones y la variedad de problemas abiertos y los enfoques, los investigadores trataron de hacer frente a diferentes aspectos de las respuestas humanas que pueden ser reconocidos como una señal visual. Algunas de las principales áreas de investigación en esta sección son los siguientes:

Análisis de Expresión Facial.

Seguimiento del movimiento del cuerpo.

Reconocimiento de gestos.

Detección de la Mirada (Seguimiento del movimiento de los ojos).

Dentro de esta categoría, se encuentra lo que se conoce como realidad mediada por computador. Este paradigma consiste en el uso del computador para modificar la percepción de la realidad. Este se divide en dos campos:

La realidad virtual (RV) que es una simulación por ordenador en la que se emplea el grafismo para crear un mundo que parece realista. Además este mundo no es estático, sino dinámico y responde a las órdenes del usuario (a través de gestos, voces, entre otros). Su clave es la interactividad en tiempo real y el sentimiento de inmersión al participar de lo que se desarrolla en la pantalla.

La realidad aumentada (RA) agrega información sintética a la realidad. Algunos la definen como un caso especial de realidad virtual (RV), otros como algo más general y ven a RV como un caso especial de RA.

Siguiendo la definición del autor Ronald Azuma un sistema de RA tiene 3 requerimientos (Azuma, 1997):

Combina la realidad y lo virtual. Al mundo real se le agregan objetos sintéticos que pueden ser visuales (como texto u objetos 3D), de sonidos, Haptics (sensibles al tacto) y /u olores.

Es interactivo en tiempo real. El usuario ve en tiempo real un mundo real con objetos sintéticos agregados, que le que ayudarán a interactuar con la realidad (Azuma, 2001).

Existen tres alternativas para la visualización en la RA: a. las imágenes sintéticas pueden proyectarse sobre los objetos reales y el usuario visualiza la escena normalmente, sin ningún dispositivo especial; b. la combinación imágenes sintéticas y reales puede proyectarse sobre una pantalla o gafas especiales “see-through” o “video-see-through” que le permitan al usuario la visión del mundo real aumentado; c. una última alternativa, es verla a través de una pantalla de un dispositivo móvil como un PDA o un teléfono móvil.

Imágenes registradas en espacios 3D. La información virtual tiene que estar vinculada espacialmente al mundo real de manera coherente. Se necesita saber en todo momento la posición del usuario con respecto al mundo real, y de esta manera la mezcla de información real y sintética podrá registrarse.

Audio como medio de Interacción (audio-based): este paradigma toma como base el uso de sonidos como medio para dar o recibir instrucciones hacia y desde los sistemas computacionales.

Las áreas de investigación en esta sección, se pueden dividir en las siguientes partes:

Reconocimiento de voz

Reconocimiento del hablante.

Análisis auditivo de emociones.

Detección de ruidos/señales emitidas por el hombre.

Interacción con la música

Sensores como medio de Interacción (Sensor-based): Esta sección es una combinación de diversas áreas con una amplia gama de aplicaciones. El carácter común de estas diferentes áreas es, que al menos, se utiliza un sensor entre el usuario y la máquina para proporcionar la interacción. Como por ejemplo:

Interacción basada en lápiz (como es el caso de los dispositivos móviles).

Ratón y Teclado.

Joysticks.

Sensores de movimiento de seguimiento y Digitalizadores.

Sensores hápticos.

Sensores de presión.

Sensores de Sabor / Olor.

Dentro de esta categoría se encuentra, la interacción háptica y teleháptica. En el campo de la tecnología, Háptica se refiere al sistema que tiene como medio de interacción el tacto; esto incluye la aplicación de fuerzas, vibraciones y/o movimientos por parte del usuario. Por otro lado, teleháptica se refiere a la generación de impulsos sensibles al tacto por medio de un computador. Por lo general, este tipo de impulsos generados por computador se utilizan como un medio de retroalimentación de un sistema computacional hacia el usuario. Un ejemplo común, dentro de la industria de los videojuegos, puede ser la tecnología Dual Shock de Sony empleada en el Play Station. En esta plataforma la retroalimentación es causada por medio de pequeños motores colocados dentro del control de juego. Mientras el jugador interactúa, el juego está programado para que dependiendo de ciertos eventos: colisiones, explosiones, etc. el motor gire haciendo vibrar el control en las manos del jugador. De esta forma hace que la experiencia de juego sea más inmersiva.

También se puede incluir dentro de los sensor-based la interfaz cerebro computador. Este paradigma plantea la implementación de una vía de comunicación directa entre la actividad cerebral o nerviosa de un ser humano, animal o una red neuronal viva con un dispositivo externo. Los sistemas que se rigen por este tipo de paradigma se ven divididos por la dirección que sigue el flujo de información. De esta forma, existen sistemas de una vía como de dos vías.

Interacción Multimodal: Se pueden dar, además de las mencionadas, interacciones multimodales, también denominadas MMHCI. El término multimodal, se refiere a la combinación de múltiples modalidades. En los sistemas MMHCI, estas modalidades en su mayoría se refieren a las formas en que el sistema responde a las entradas, es decir, los canales de comunicación. La definición de estos canales se hereda de los tipos humanos de la comunicación que son, básicamente, sus sentidos: vista, oído, tacto, olfato y gusto. Las posibilidades para la interacción con una máquina los incluyen, pero no están limitados a estos.

Por lo tanto, la interfaz multimodal actúa como un facilitador de la interacción humano-computadora a través de dos o más modos de entrada que van más allá del teclado y el ratón tradicionales. El número exacto de los modos de entrada soportados, sus tipos y la forma en que trabajan juntos, pueden variar ampliamente de un sistema multimodal a otro.

Un aspecto interesante de la multimodalidad es la colaboración de las diferentes modalidades para ayudar a los reconocimientos. Por ejemplo, el seguimiento de movimiento de los labios (visual-based) puede ayudar a los métodos de reconocimiento de voz (audio-based) y los métodos de reconocimiento de voz (audio-based) pueden ayudar a la adquisición de comandos en el reconocimiento de gestos (visual-based) (Karray, Milad, Abou, Arab, 2008).

Computación Ubicua (Ubicomp):La inteligencia ambiental o la computación ubicua , la cual es llamada la Tercera Ola, está tratando de integrar la tecnología en el entorno, de modo que sea más natural e invisible al mismo tiempo (Karray, Milad, Abou, Arab, 2008).

En 1991, Mark Weiser (Xerox PARC) explica que esta interacción trataría de extender la capacidad computacional al entorno del usuario, permitiendo que la capacidad de información esté presente en todas partes en forma de pequeños dispositivos muy diversos que permiten interacciones de poca dificultad, conectados en red a servidores de información.

El diseño y localización de estos dispositivos deben ser ideados especialmente para la tarea objeto de interacción. La computación, por tanto, deja de estar localizada en un único punto para pasar a diluirse en el entorno.

El problema es crear un entorno de software/hardware que soporte la multi–persona, el multi–dispositivo, en diferentes lugares de una manera integrada, hecho a partir de la tecnología existente, pero aportando coherencia sobre el entorno como un todo. Esto implica, al mismo tiempo, integración del sistema de tal manera que todos los dispositivos puedan ínteroperar.

Antecedentes de diferentes paradigmas de HCI enfocados al ámbito de la educación especial

En este aparatado se revisan algunos antecedentes de aplicación de diferentes paradigmas de interacción para personas con necesidades especiales.

Una aplicación de realidad aumentada aplicada al área de educación especial, es la utilizada en el proyecto Pictogram Room (PictogramRoom, 2012). En este proyecto, se ha creado una habitación de realidad aumentada para enseñar a comprender los pictogramas a personas con trastornos del espectro del autismo. Los pictogramas conforman sistemas de comunicación alternativa, pero las personas con autismo, pueden presentar dificultades en su reconocimiento, cuando se presentan mínimos cambios en el pictograma, ya sea en el grosor, color, entre otros. El proyecto, plantea que con la ayuda de la realidad aumentada, la posibilidad de usar pictogramas superpuestos sobre objetos reales, ayude a estas personas a ver la conexión entre imagen real y pictograma en tiempo real. En este caso, el paradigma de interacción que prevalece es visual-based.

El proyecto **NAVI** (NAVI, 2012) pone en juego diferentes paradigmas de interacción como el basado en realidad mediada por computador, teleháptica y el basado en audio. Está orientado a personas con dificultades en la visión. Utiliza Kinect para recolectar datos visuales del entorno, tales como formas, colores, velocidad relativa de los objetos y varios parámetros más, los cuales son procesados por una notebook que la persona asistida lleva en su espalda. La información es traducida en indicaciones verbales y en vibraciones de advertencia en un cinturón especialmente diseñado, que le informan a la persona sobre la proximidad de obstáculos y las características espaciales del lugar donde se encuentra.

A medida que las cámaras de Kinect van generando la imagen tridimensional, la notebook identifica los posibles obstáculos que pueden presentarse en la trayectoria de la persona, e inmediatamente transmite esta información a un controlador que le envía una señal a 3 motores que producen vibraciones, indicándole a la persona en qué dirección se encuentra el obstáculo. Asimismo, se emite una advertencia sonora mediante un par de auriculares Bluetooth. El sistema también utiliza ARToolKit (un

framework para desarrollar aplicaciones de realidad aumentada) para brindarle información adicional al no-vidente, señalando puertas, salidas y otros objetos de interés.

El proyecto ABI (Adaptive Brain Interface) es un ejemplo de sustitución sensorial muy útil para personas con discapacidad (Millar, Hausen, Renkens, 2002). Hace posible que una persona transmita órdenes a la computadora mediante impulsos eléctricos, emitidos por su cerebro cuando piensa en realizar un determinado movimiento. En una prueba realizada con 15 individuos, y tras sólo unas pocas horas de aprendizaje, el sistema reconoció tres estados distintos, con el 70% de aciertos y sólo el 5% de errores (el resto de las veces el equipo no actuó para evitar daños). La actividad cerebral asociada a la imaginación de un movimiento es distinta a su ejecución y lo que resulta más útil aún es que el cerebro visualiza el movimiento muscular milisegundos antes de que se ejecute realmente. El proyecto ABI y otros, actúan localizando la zona del cerebro que se activa un instante antes de que la persona alargue la mano para mover el cursor o el teclado del ordenador. Se colocan unos cascos con electrodos sensibles y un software apropiado. Se propone que una persona pueda escribir un texto mediante un simulador de teclado o manipular una silla robotizada, por ejemplo, y se capturan los impulsos eléctricos generados previos al movimiento para que se realice la acción (Sánchez Montoya, 2012). En este caso, se presenta una aplicación de paradigma basado en interfaz Cerebro Computador (Sensor-based).

El reconocimiento de voz como paradigma de interacción para personas con dificultades motoras

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 estas posibilidades. 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 (Madrid Vivar —2002). 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. Es importante decir que el reconocimiento de voz es aún un problema sin resolver. No existe aplicación que haya demostrado poder realizar el reconocimiento de la voz con la misma habilidad que los humanos han desarrollado. Tal es así, que aunque se hayan presentado muchas maneras de resolver este problema, no se ha alcanzado el auge en esta tecnología. Tal vez, a causa de esto, sea que el reconocimiento de voz automatizado se ha transformado en un tema de estudio atractivo para muchos investigadores, desarrolladores y estudiantes. Esto se traduce en que estos sistemas no cuentan con una fiabilidad del 100%, por lo que es un área en la que se necesita una profunda investigación, en la cual se puede incurrir para mejorar la autonomía y calidad de vida de las personas, entre muchas otras aplicaciones posibles.

jclicVoice: Adaptación al software jClic mediante comandos por voz.

Como caso de estudio del presente artículo, se presenta el aporte realizado en el marco de esta investigación: jClicVoice, que consiste en una adaptación al software educativo jClic (jClic, 2012), con el fin de incorporar un nuevo modo de interacción en esta aplicación, mediante comandos por voz.

Este trabajo está destinado a las personas con problemas motores, pero sin consecuencias o con consecuencias leves en el desarrollo el 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”.

En las siguientes secciones se explicará la metodología utilizada para el desarrollo de JClicVoice, previamente se explicarán los componentes del software JClic, necesarios para comprender la adaptación aquí propuesta.

JClic está formado por tres aplicaciones, una de las cuales, se utilizan para la resolución de las actividades (jClic, 2012):

JClic Player.Esta componente puede ser presentada de dos maneras diferentes: como applet o como aplicación JClic. El "applet", permite incrustar actividades JClic en una página web para ejecutarlas en cualquier navegador. La aplicación JClic es un programa independiente que una vez instalado permite realizar las actividades desde el disco del ordenador.

JClic autor. La herramienta de autor que permite crear, editar y publicar las actividades.

JClic reports. Es el módulo encargado de recopilar los datos (tiempo empleado en cada actividad, intentos, aciertos, etc.), y presentarlos, luego, en informes estadísticos de diversos tipos.

En la siguiente subsección se presenta la metodología empleada para llevar a cabo la adaptación y conformar el software JClicVoice.

Metodología

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 único 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.

Este trabajo se llevó adelante en diferentes fases o etapas que se detallan a continuación.

Etapa 1: Análisis

En la etapa de análisis se abordaron varios aspectos a estudiar. Una de las decisiones que se consideró fue cómo tomar conocimiento de que se desea realizar la actividad utilizando comandos por voz.

Se decidió que en esta situación, el usuario deba contar con la asistencia del docente o facilitador, ya que es éste quien toma la decisión para cada alumno en particular, si es adecuado o no utilizar reconocimiento de voz 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 comandos por voz.

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

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 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 abordaron 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.

Otro aspecto a resolver fue el hecho 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.

Finalmente, se analizó la forma de 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 reconocimiento de voz.

Etapa 2: Configuración de Sphinx-4

Sphinx es un framework de reconocimiento de voz, flexible, modular y fácil de incorporar a otras aplicaciones. En particular, la versión 4 fue desarrollada en Java.

Para utilizar Sphinx-4, en primer lugar, se debe descargar la aplicación desde el sitio oficial (CMU Sphinx, 2011). Allí está disponible el código fuente de la herramienta, aunque si no se desea modificar código (como en este 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 reconocimiento de voz, que realizan este tipo de solución (Magnus, 2010).

En un primer momento, se utilizó esta opción 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 JClicVoice 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 adoptar un modelo ya entrenado, pero se hicieron además algunas pruebas con el entrenador, de manera tal, de entender y estudiar su funcionamiento.

Se eligió un modelo ya entrenado, disponible en la web y de libre uso. 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 (DIME, 2011). Es importante aclarar que para el entrenamiento de este modelo, se utilizaron 30 hablantes diferentes, y de cada hablante se recolectaron: 50 oraciones diferentes entre sí y 10 oraciones comunes entre los hablantes. Es decir, en total: 1500 oraciones diferentes y 300 oraciones comunes.

A partir de esta lista de unidades fonéticas se creó el diccionario a utilizar en la integración con JClic. Podría haberse incorporado 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 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.

Etapa 3: Desarrollo del prototipo

En esta sección se describen 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 usó 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, donde se crea el reconocedor para empezar a trabajar.

También, se hicieron las modificaciones necesarias 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 activad 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”, luego de nombrar los dos casilleros de la manera ya explicada, el sistema muestra un cartel con los valores que se van a procesar, el usuario debe 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. Cuándo 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 y, además, las de asociación compleja, puzzle de intercambio y puzzle de agujero.

Se puede descargar a la aplicación en la dirección https://projectes.lafarga.cat/projects/jclicvoice/downloads.

En la Figura 1, se ejemplifica una actividad de asociación simple, utilizando la adaptación de JClic.

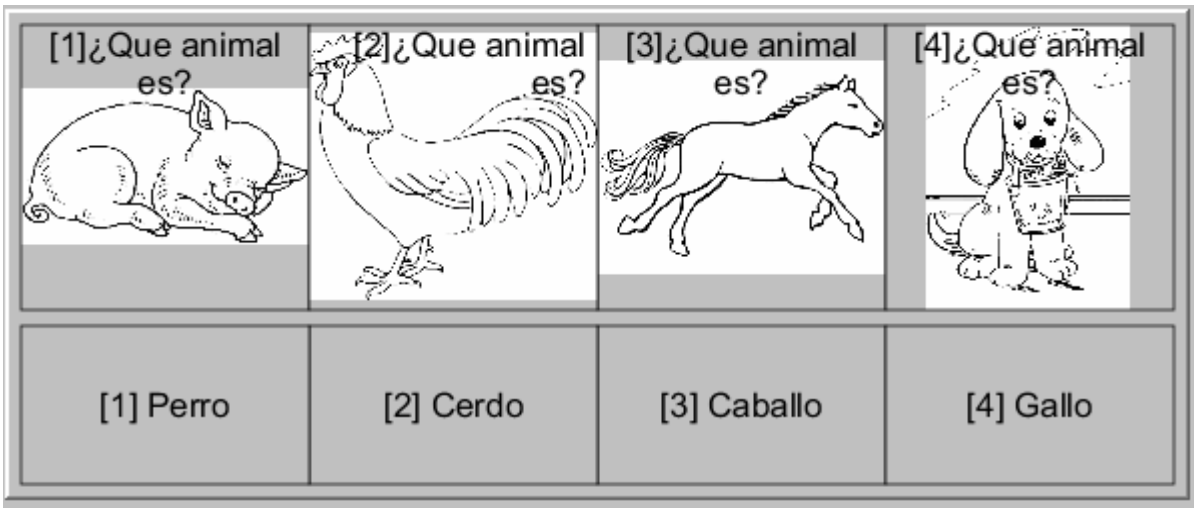

Figura 1: Actividad de asociación simple adaptada.

A continuación se presenta la evaluación realizada, para esta primera etapa, de las estrategias de integración planteadas.

Resultados y discusión

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.

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 (Arquer, 2010). 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 para el grupo destinatario. 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. Actualmente, el prototipo ha sido presentado a docentes del área de educación especial, con buena aceptación de la adaptación.

Conclusiones y trabajos futuros

Se considera que JClicVoice es un aporte significativo para el área de Educación Especial. Incorpora uno de los paradigmas de interacción (basado en audio), revisados en este trabajo con el fin de dar la oportunidad, a usuarios con dificultades motoras y sin compromisos en el habla, de abordar la realización de actividades educativas mediante comandos por voz.

Acorde a los trabajos revisados en referencia a paradigmas de HCI y educación especial, es de interés profundizar en este estudio, ya que cada paradigma abre las puertas a que más usuarios puedan hacer provecho de las posibilidades de las TIC.

Como trabajos futuros se está planificando el uso de JClicVoice en escenarios de educación especial, pera poder evaluar sus posibilidades con alumnos y docentes del área. Además se está avanzando en el desarrollo de nuevas aplicaciones basadas en otros paradigmas como el de Realidad Aumentada.

Referencias bibliográficas

Asociación de Maquinaria Computarizada (ACM).(2012) http://www.acm.org.

Arquer (2010). *Fiabilidad humana: métodos de cuantificación, juicio de expertos* http://www.insht.es/InshtWeb/Contenidos/Documentacion/FichasTecnicas/NTP/Ficheros/401a500/ntp_401.pdf, 2010.

Azuma, R. (1997). A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments 6*, 4 (August), 355-385.

Azuma, R. (2001). Recent Advances in Augmented Reality. *IEEE Computer Graphics and applications* (Nov-Dec 2001), 34-47.

Castellano, Sacco, Zurueta (2003). *La utilización de software de uso general y aplicaciones específicas en el área de las discapacidades motrices*. IV Congreso Iberoamericano de Informática en la Educación Especial. Disponible en http://www.niee.ufrgs.br/eventos/CIIEE/2003/. Recuperado en 2012

Centro para el Estudio de Libreras Digitales(CSDL) (2008) *Curso de HCI (CPSC 436).* http://www.csdl.tamu.edu/ leggett/courses/436/part1/sld015.htm. Consultado en 2008.

CMU Sphinx (2011). http://cmusphinx.sourceforge.net/sphinx4/.

Diálogo Inteligente Multimodales en Español (s/f) http://leibniz.iimas.unam.mx/~luis/DIME/recursos.html. Consultado en 2011.

jClic, (2012). http://clic.xtec.cat/es/jclic/index.htm.

Karray, M. y Abou, A. (2008). *Human-Computer Interaction: Overview on State of the Art*. http://www.s2is.org/Issues/v1/n1/papers/paper9.pdf. Recuperado en 2012.

Madrid Vivar (2002). http://www.tecnoneet.org/docs/2002/2-82002.pdf. Recuperado en 2012.

Mouse Advanced GNU Speech (Magnus): http://magnusproject.wordpress.com/.

Millar, Hausen, Renkens (2002). *Adaptative Brain Interface - ABI: ABI: Simple Features, Simple Neural Network, Complex Brain*. Disponible en www.cs.cmu.edu/~tanja/BCI/ABI2000.pdf. Recuperado en 2012.

NAVI (2012). www.webayunate.com/ojos-virtuales-para-personas-ciegas-gracias-a-kinect.

Pictogramroom (s/f) http://fundacionorange.es/areas/22_proyectos/proy_230.asp. Consultado en 2012.

Sánchez Montoya, R. (2012). *Computer human interaction adaptative brain*. Consultado enhttp://sce.uhcl.edu/boetticher/CSCI5931%20Computer%20Human%20Interaction/Adaptive%20Brain%20Computer%20Interface.pdf. Consultado en 2012.

Sánchez Montoya, R. (2002). *Ordenador y Discapacidad. Guía práctica a las personas con necesidades educativas especiales*. Madrid.

Sancho, (1998). *La tecnología: un modo de transformar el mundo cargado de ambivalencias*. Barcelona: HORSORI