

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

M. Lucrecia Moralejo^{1,2}, Cecilia V. Sanz², Patricia Pesado^{2,3}
{lmoralejo, csanz, ppesado}@lidi.info.unlp.edu.ar

¹Becaria UNLP (Universidad Nacional de La Plata), Buenos Aires, Argentina.

²III LIDI, Facultad de Informática, UNLP, La Plata, Buenos Aires, Argentina.

³CIC (Comisión de Investigaciones Científicas de la Pcia. de Buenos Aires), Argentina.

Abstract. Este trabajo se enmarca en el área de interacción hombre-máquina (HCI), en donde se analizan algunos de sus diferentes paradigmas aplicados al escenario de educación especial. Así, se presenta aquí una revisión de antecedentes. Luego, se describe la utilización del paradigma de HCI basado en la utilización de comandos por voz para el desarrollo de una adaptación al software educativo JClic, con el objetivo de ser utilizado por usuarios/alumnos con deficiencia motriz sin consecuencias o con consecuencias leves en el desarrollo del lenguaje. Para llevar a cabo esta aplicación, 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 estudio y los detalles propios de la implementación de un prototipo (JClicVoice), que lleva adelante la adaptación a JClic. Finalmente, se describen los resultados y conclusiones obtenidas, luego de la evaluación del prototipo.

Keywords: Educación especial, Interacción hombre-máquina, JClic, Sphinx, comandos por voz.

1 Introducción

El desarrollo tecnológico, posibilita la realización de las actividades de la vida diaria de cualquier persona, esto es aún más evidente en las personas con discapacidad al contribuir, con sus tecnologías y herramientas técnicas, a hacer una vida autónoma más independiente. Pero también es cierto que deben considerarse los elementos que hagan útil esa tecnología para el mayor número de personas, con una atención y búsqueda constantes en sus técnicas y recursos, evitando que puedan convertirse en barreras de comunicación, de información o de accesibilidad. Esto es, debe conducirse a un desarrollo tecnológico con patrones de diseño universal, favoreciendo un verdadero marco de inclusión en la sociedad. Esta reflexión debe hacerse ya que, como afirma Grau [1], tradicionalmente, las tecnologías han sido concebidas, proyectadas, producidas y aplicadas con arreglo al patrón de la persona media, sin

tener en cuenta, o en menor medida, la atención a las diferencias que se derivan de una discapacidad, lo que obliga, a veces, a la realización de algunas adaptaciones de técnicas o recursos necesarios a personas o casos particulares [2].

2 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) [3] 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 disciplinas 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 el Centro para el Estudio de Librerías Digitales [4] 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).

3 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 más paradigmas de interacción. Por otra parte, 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 [5], revisados en esta investigación

Visión como medio de interacción (Visual-based 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 las siguientes:

- Análisis de Expresión Facial.
- Seguimiento del movimiento del cuerpo.
- Reconocimiento de gestos.
- Detección de la Mirada (Seguimiento del movimiento de los ojos).

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).

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.

- 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.

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.

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).

Computación Ubicua (Ubicomp).

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.

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.

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

Como parte de este trabajo, se viene realizando una revisión de antecedentes en el área de aplicación de paradigmas de HCI y educación especial. A modo de ejemplo se mencionarán algunos.

El proyecto Pictogram Room [6] plantea que con la ayuda de la realidad aumentada, la posibilidad de usar pictogramas superpuestos sobre objetos reales puede ayudar a personas autistas a ver la conexión entre imagen real y pictograma en tiempo real. En este caso, el paradigma de interacción que prevalece es visual-based.

Otro ejemplo, es el del proyecto NAVI [7]. Se trata de una aplicación orientada a personas con dificultades en la visión. Recolecta datos visuales del entorno, tales como formas, colores, velocidad relativa de los objetos, etc. Los procesa y brinda indicaciones verbales al usuario que presenta dificultades en el sentido de la vista. También, utiliza un cinturón que vibra para indicar la proximidad de obstáculos.

El proyecto ABI (Adaptive Brain Interface) es un ejemplo de sustitución sensorial muy útil para personas con discapacidad [8]. Hace posible que una persona transmita órdenes a la computadora mediante impulsos eléctricos emitidos por su cerebro cuando piensa en realizar un determinado movimiento. 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.

Como parte de esta investigación, se ha planteado el desarrollo de un prototipo de adaptación de un software educativo utilizando el paradigma de HCI basado en audio, según la clasificación antes presentada. En la siguiente sección se describe los pasos realizados para alcanzar este objetivo.

4 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 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 [9]. 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. Estos sistemas no cuentan con una fiabilidad del 100%, por lo que es un área en la que se necesita una profunda investigación, y que puede colaborar en mejorar la autonomía y calidad de vida de las personas, entre muchas otras aplicaciones posibles.

5 Jclivoice : Adaptación al software JCliv mediante comandos por voz.

Como una aplicación de los estudios que se vienen desarrollando en el área, se presenta uno de los aportes realizados al escenario de educación especial: jClivVoice, que consiste en una adaptación al software educativo JCliv [10], 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 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”.

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

JCliv está formado por tres aplicaciones, una de las cuales, se utiliza para la resolución de las actividades [10]:

- JCliv Player : Esta componente para resolver las actividades, pueden presentarse como applet o como aplicación jCliv.
- JCliv autor: La herramienta de autor que permite crear, editar y publicar las actividades.
- JCliv 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 JclivVoice.

5.1 JClickVoice como prototipo adaptado utilizando comandos por voz

5.1.1 Antecedentes del trabajo.

La adaptación propuesta ha abordado la modificación de las actividades de JClick, de manera que se puedan resolver a través de la utilización de comandos por voz.

En una primera instancia se tomaron en cuenta, sólo, las actividades de asociación simple. En trabajos previos se presentó una explicación detallada de cada una de las etapas transitadas para diseñar el prototipo de adaptación. En este caso se detallarán las nuevas actividades implementadas como parte de JClickVoice [11]. Sin embargo, se mencionan algunas decisiones previas para dar contexto a lo aquí expuesto. Uno de los primeros pasos ha sido analizar la utilización del motor de reconocimiento de voz. Se estudiaron diferentes y se realizó una comparación. Finalmente, se seleccionó un software de reconocimiento de voz llamado Sphinx [12], más precisamente, la versión 4 desarrollada en Java. Además, se debió analizar de qué forma los usuarios harían referencia a los elementos de interacción que aparecían en las actividades. Se decidió utilizar etiquetas numeradas para poder mencionarlas en los comandos por voz. Luego de realizado el primer prototipo, se planteó un proceso de evaluación a través de juicio de expertos. Se diseñaron una serie de encuestas, que fueron respondidas por expertos del área de educación especial, informática, y tecnología educativa.

Como conclusión de este proceso de evaluación, se pudo conocer que se consideró una buena elección el software educativo a adaptar, como así también, la utilización de comandos por voz como ayuda técnica para el grupo destinatario. Además, se propusieron algunas modificaciones al prototipo que se llevaron a cabo. Por ejemplo, dar la posibilidad al docente de seleccionar para cada alumno si utilizará la opción de comandos por voz o no. Otras modificaciones se vincularon con el diseño de la interfaz y la visualización de las etiquetas. El resultado más importante ha sido el demostrado interés por los especialistas y profesionales del área en el uso de este prototipo. El trabajo fue presentado durante un evento del área de educación especial, con un numeroso grupo de terapeutas y docentes de este escenario educativo, y se evidenció el interés de poder tener disponible el JClickVoice para ser descargado.

Teniendo en cuenta, los resultados obtenidos, se ha extendido el prototipo a otro tipo de actividades que JClick permite crear, que son las actividades de asociación compleja, puzzle de intercambio y puzzle de agujero.

5.1.2 Descripción de las nuevas actividades incluidas en JClickVoice.

Una de las nuevas actividades incorporadas a JClickVoice es el puzzle de agujero. Este consta de una sola ventana con las piezas desordenadas y una casilla vacía. El contenido de esta casilla, que es una de las piezas elegida por el programa aleatoriamente, aparece a la derecha. Una vez resuelto el puzzle, la pieza a la derecha (correspondiente al faltante o casilla vacía), se autocompleta.

Los desplazamientos de las piezas están restringidos, solo se pueden moverse las piezas contiguas a la casilla vacía y eso hace que esta modalidad resulte la más

complicada de todas, especialmente si el contenido no es gráfico o si tiene un elevado número de piezas.

En la Figura 1, se ejemplifica una actividad de puzzle de agujero, utilizando la adaptación de JClic.

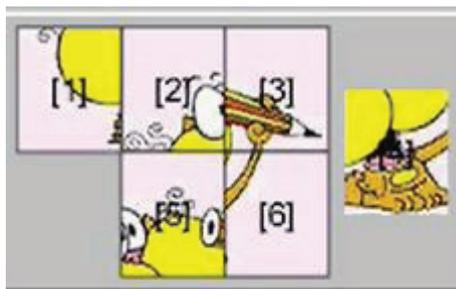


Figura 1: Actividad de asociación. Puzzle de agujero.

Para resolver este tipo de actividad en JClicVoice, el alumno deberá pronunciar la etiqueta asociada al elemento que se desea mover hacia el agujero o casilla vacía y la palabra "Aceptar". Así, sólo requerirá utilizar los comandos de voz indicados para resolver la actividad.

En el caso del puzzle de intercambio se presenta la información desordenada y el alumno cuenta con un sólo panel para reconstruir el contenido. De esta manera, se tienen que ir cambiando las piezas de lugar hasta que las casillas estén ordenadas.

Por ejemplo, teniendo en cuenta la partida que se muestra en la imagen de la figura 2, el alumno deberá pronunciar: "Uno con Tres Aceptar" para intercambiar el casillero con la etiqueta 1, con el que posee la etiqueta 3.



Figura 2. Puzzle de intercambio

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". Es entonces cuando se verifica 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.

Resumiendo, tanto en el caso de las actividades de asociación simple, como en las de asociación compleja y en el puzzle de intercambio, el alumno deberá pronunciar la etiqueta asociada a un elemento del primero conjunto de datos, la palabra "Con", y nuevamente la etiqueta asociada a un elemento del segundo conjunto de datos. Para confirmar su decisión, deberá decir la palabra "Aceptar".

Para llevar a cabo esta adaptación, además, fue necesario realizar modificaciones al componente JClic Author, de manera tal que sólo permita crear las actividades que han sido adaptadas en jClicVoice, obteniendo así, una aplicación cerrada y totalmente funcional.

La adaptación propuesta se encuentra disponible para poder ser descargada¹. Al mismo tiempo, se ha diseñado un instrumento de evaluación que es posible obtener desde el mismo servidor² y que se propone enviar a los autores para poder seguir evolucionando este prototipo con los comentarios de la comunidad.

6 Conclusiones y trabajos futuros

Este trabajo constituye un primer paso en un camino que se ha propuesto recorrer. Al momento se ha trabajado con un paradigma de HCI, enfocando su utilización al escenario educativo en cuestión. Actualmente, se están revisando las posibilidades de utilización de paradigmas basados en la visión para otros grupos destinatarios del ámbito de educación especial. En lo que respecta a jClicVoice, se está trabajando en un proceso de evaluación que permita llegar a los alumnos y docentes del área y poner en juego las actividades diseñadas.

¹ <https://projectes.lafarga.cat/projects/jclicvoice/downloads>

² <https://projectes.lafarga.cat/projects/jclicvoice/surveys>

7 Referencias bibliográficas

1. Grau. "Tecnología y discapacidad visual. Necesidades tecnológicas y aplicaciones en la vida diaria de las personas con ceguera deficiencia visual". Madrid. Fundación UCM-ONCE. 2004 -
2. Luque Parra, Rodriguez Infante."Tecnología de la información y comunicación aplicada al alumnado con discapacidad: Un acercamiento docente". Revista Iberoamericana de Educación, ISSN: 1681-2653. 2009.
3. ACM. Asociación de Maquinaria Computarizada (ACM). <http://www.acm.org>. Consultado en 2012.
4. Centro para el Estudio de Librerías Digitales(CSDL) de la universidad AM de Texas, Curso de HCI (CPSC 436). <http://www.csdl.tamu.edu/leggett/courses/436/part1/sld015.htm>. Consultado en 2008.
5. Karray, Milad, Abou, Arab. Human-Computer Interaction: Overview on State of the Art. <http://www.s2is.org/Issues/v1/n1/papers/paper9.pdf>. 2008. Recuperado en 2012.
6. Pictogramroom. http://fundacionorange.es/areas/22_proyectos/proy_230.asp. Consultado en 2012.
7. NAVI. www.webayunate.com/ojos-virtuales-para-personas-ciegas-gracias-a-kinect. Consultado en 2012.
8. 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.
9. Madrid Vivar. <http://www.tecnoneet.org/docs/2002/2-82002.pdf>. 2002 . Recuperado en 2012.
10. JClíc, sitio oficial. <http://clíc.xtec.cat/es/jclíc/index.htm>. Consultado en 2012.
11. Moralejo Lucrecia, Ostermann Stefania, Sanz Cecilia, Patricia Pesado. "Adaptación mediante comandos de voz, para ámbitos de la Educación especial.". Congreso Argentino de Ciencias de la Computación CACIC 2010.
12. CMU Sphinx, Sitio Oficial - <http://cmusphinx.sourceforge.net/sphinx4/>. Consultado en 2011.