

DESARROLLO DE HERRAMIENTAS DE ACCESIBILIDAD AL ORDENADOR Y A INTERNET BASADAS EN LA VOZ

Convocatoria XIII Premios de Accesibilidad
Universal 2010

Oscar Saz, William Ricardo Rodríguez, José Enrique García, Antonio Escartín, Eduardo Lleida,
Laura Abarca, Sara Mejuto, Verónica Bermúdez, Marta Peña, Elena Lafuente, José Félix Muñoz

DESARROLLO DE HERRAMIENTAS DE ACCESIBILIDAD AL ORDENADOR Y A INTERNET BASADAS EN LA VOZ

¹Oscar Saz, W.-Ricardo Rodríguez, J.-Enrique García, Antonio Escartín & Eduardo Lleida
ViVoLab, GTC, I3A, Universidad de Zaragoza

²Laura Abarca, Sara Mejuto, Verónica Bermúdez & Marta Peña
ASPACE-Huesca y CADIS-Huesca

³Elena Lafuente & José Félix Muñoz
Laboratorio Aragonés de Usabilidad, Walqa

RESUMEN

El acceso a las nuevas Tecnologías de la Información y Comunicaciones (TICs) supone una barrera que limita enormemente las posibilidades de desarrollo personal de personas con diversidad funcional. Mientras para la mayoría de la sociedad, obtener y mejorar el conocimiento del mundo a través del ordenador e Internet es algo habitual, para las personas con discapacidad supone un trabajo costoso debido a sus limitaciones. La necesidad de empoderar al usuario y proporcionar a estas personas acceso a los nuevos elementos de comunicación es un esfuerzo básico a la hora de pensar en la integración total de este colectivo (Abascal, 2008).

El trabajo que aquí se presenta desgana una doble vertiente de trabajos para mejorar este acceso a las nuevas tecnologías. La primera de ellas está orientada a proporcionar un modo alternativo de control a personas con una gran discapacidad física de los periféricos del ordenador (ratón y/o teclado) mediante entradas orales. La segunda describe una serie de herramientas para que los creadores de contenidos en Internet aprovechen las potencialidades de la síntesis y el reconocimiento de voz para proporcionar un acceso mejorado a personas con diferentes limitaciones: invidentes por un lado y personas con discapacidad física por otro.

Todo el núcleo del trabajo de desarrollo se ha llevado a cabo por parte de grupo de investigación en tecnologías del habla y el lenguaje ViVoLab, del Grupo de Tecnologías de las Comunicaciones (GTC) del Instituto de Investigación en Ingeniería de Aragón (I3A) de la Universidad de Zaragoza. Este grupo cuenta con experiencia en el trabajo para mejorar la calidad de vida de personas con discapacidad mediante el software para logopedia en "Comunica"¹ (Saz et al., 2009a), desarrollado en colaboración con el Colegio Público de Educación Especial "Alborada" y otras instituciones.

Para los trabajos aquí presentados se ha contado con la necesaria colaboración en el conocimiento de las necesidades de usuarios con discapacidad de ASPACE-Huesca y la Coordinadora de Asociaciones de Discapacidad de Huesca (CADIS-Huesca);

¹ <http://www.vocaliza.es>

y en el ámbito de la usabilidad y accesibilidad en la web del Laboratorio Aragonés de Usabilidad en el Parque Tecnológico Walqa. El enfoque multidisciplinar de todo el trabajo realizado es uno de los componentes básicos del mismo.

PARTE I. ACCESIBILIDAD AL ORDENADOR

I.1. INTRODUCCIÓN

La mayor dificultad que presenta una persona con una grave dificultad motriz, como por ejemplo alguien con parálisis cerebral, es la imposibilidad de manejar de forma eficiente elementos básicos en la interacción con el ordenador como el ratón o el teclado. Otras posibilidades de interacción deben ser, pues, estudiadas, adaptando la forma en que el usuario se relaciona con la máquina de acuerdo a sus limitaciones existentes. Los campos más novedosos tecnológicamente en los que se trabaja actualmente integran elementos como el seguimiento del iris, el reconocimiento de voz (Hawle et al., 2003), o los interfaces cerebrales (Iturriate et al., 2009), entre otros.

I.2. EMULADOR DE PULSACIÓN BASADO EN VOZ

Los sistemas emuladores de teclado y ratón basados en barrido son, a día de hoy, uno de los elementos más usados por aquellos que trabajan con personas con diversidad funcional como la parálisis cerebral. El programa de barrido presenta de forma sucesiva al usuario las diferentes posibilidades (teclas en el teclado o direcciones de movimiento del ratón), y el usuario debe accionar un pulsador para activar la opción marcada en ese instante. El uso de pulsadores mecánicos accionados no sólo con la mano o brazo, sino con la cabeza o la pierna, es la situación más habitual en estos casos; aunque en muchas situaciones puede producir una sensación de incomodidad o cansancio al usuario debido a la necesidad de forzar la postura o repetir de forma continuada un mismo movimiento. La sustitución de estos pulsadores mecánicos por otros elementos de interacción basados en un interfaz más natural para la persona podría ser una solución para algunos de estos usuarios, simplificando y agilizando su trabajo con el ordenador.

De esta forma, el proyecto de colaboración “Sistemas De Interacción Oral Sobre Dispositivos Portables” entre CADIS-Huesca y la Universidad de Zaragoza vía ViVoLab (GTC/I3A) y auspiciado por la Diputación Provincial de Huesca (DPH), el Instituto Aragonés de Servicios Sociales (IASS) y la Caja de Ahorros de la Inmaculada (CAI) pretendía encontrar una solución a esta situación. El desarrollo del trabajo se llevó a cabo con la colaboración del personal del centro de día de ASPACE en Huesca. Dicho centro realiza actividades con usuarios con diferentes grados de parálisis cerebral, fomentando su capacitación y su autonomía personal. Entre las actividades existentes, se realizan aulas de informática, donde los usuarios practican diferentes tareas con el ordenador según sus posibilidades; desde acceder a Internet a escribir en un procesador de textos pequeñas historias o tarjetas de felicitación.

Después de evaluar las posibilidades que los interfaces orales podrían ofrecer a personas con parálisis cerebral severa como es el caso de ASPACE-Huesca se decidió que, dada la fuerte disartria que dificulta el habla de estas personas, se debía evitar el uso de sistemas basados en reconocimiento del habla, cuyas prestaciones decaen en presencia de fuertes trastornos del habla, buscando recoger emisiones más simples emitidas por los usuarios. Así pues, se empezó a plantear la forma de sustituir los pulsadores mecánicos existentes por pulsación por voz. Un grupo de usuarios fueron seleccionados por el personal de ASPACE-Huesca como sujetos de prueba durante el desarrollo de la ayuda técnica, presentando cada uno de ellos diferentes características en su capacidad cognitiva y movilidad.



Figura 1: Esquema de pulsación emulada por voz

El esquema de funcionamiento del software desarrollado al final del proyecto se basa en el esquema de bloques de la Figura 1. A través del micrófono, el usuario realiza una emisión oral en el momento en que quiera activar la pulsación. El sistema de captura de audio lleva incorporado un sistema de reducción de ruido para eliminar el ruido presente en el audio capturado. Posteriormente se estima la energía de dicha emisión, con el ruido ya suprimido, y cuando dicha emisión responde a la forma de un pulso cuadrado de una determinada duración, se lanza el evento de pulsación en el ordenador.

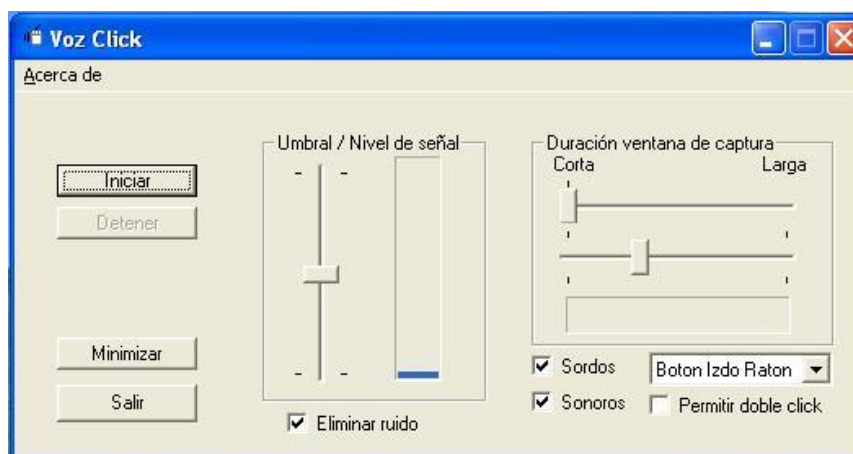


Figura 2: Interfaz visual de "VozClick"

El interfaz de usuario del software "VozClick" se puede ver en la Figura 2. Está pensado para ser configurado por la persona que trabaje con el usuario, y una vez iniciado permite a dicho usuario emular la pulsación con su habla. La ventana de aplicación contiene todos los elementos requeridos para su funcionamiento, como la intensidad mínima que tiene que alcanzar la emisión del usuario para activar la

pulsación, así como la duración de la emisión de voz que el usuario tiene que realizar para que el sistema se active. Otros elementos secundarios de configuración de la aplicación proporcionan las siguientes funcionalidades: Distinguir emisiones sonoras (articulaciones vocálicas) de sordas (soplidos); activar o no la supresión del ruido según el entorno en que se trabaje, minimizar la ventana en la bandeja de Windows o habilitar/deshabilitar las pulsaciones consecutivas en un corto espacio de tiempo, para permitir la emulación de la característica de “doble clic”.

I.3. EMULADOR DE RATÓN BASADO EN LA PRODUCCIÓN DE VOCALES

Tras la finalización del proyecto que dio lugar a “VozClick” se siguió manteniendo abierta la línea de trabajo con CADIS-Huesca y ASPACE-Huesca. En este sentido, se estudió la posibilidad de eliminar completamente el uso de sistemas de barrido de ratón en el ordenador a través de interfaces orales. Como se ha comentado anteriormente, el uso de reconocimiento del habla no es suficientemente robusto en presencia de usuarios con disartria muy profunda, por lo que se desestimó el reconocimiento de palabras completas como “arriba”, “abajo”, “izquierda” o “derecha”. Aunque se sigue investigando en este aspecto, todavía no es posible el desarrollo de herramientas reales que apliquen este conocimiento.

Según la experiencia de otros grupos de investigación de reconocido prestigio (Bilmes et al., 2006)(Harada et al., 2008) se observó que el control de la producción de vocales en individuos afectados por disartria era mejor que la calidad e inteligibilidad del habla. Por ese motivo, se pensó en la posibilidad de emular las cuatro direcciones básicas de uso de un ratón mediante la producción de vocales sostenidas por parte del usuario. Según esta idea, se desarrolló una nueva versión de “VozClick”, llamada “VocalClick”, que funcionara bajo esta premisa.

El diagrama de funcionamiento de “VocalClick” se presenta en la Figura 3. El audio es capturado por el micrófono del ordenador, la estimación de los formantes vocálicos se realiza en los segmentos sonoros de la voz mediante algoritmos de procesamiento robusto de señal, y estos formantes se trasladan al mapa formántico del español para detectar de forma automática la vocal pronunciada. Según la vocal reconocida, se realiza el movimiento del ratón en la dirección asignada. (“a” para la derecha, “i” para la izquierda, “e” para arriba, “o” y “u” para abajo). La pulsación del botón izquierdo del ratón se emula de la misma manera que en “VozClick”, mediante un golpe de voz sonoro, además, permite emular la doble pulsación repitiendo este golpe de voz en un breve periodo de tiempo, que también es configurable según las capacidades del usuario..

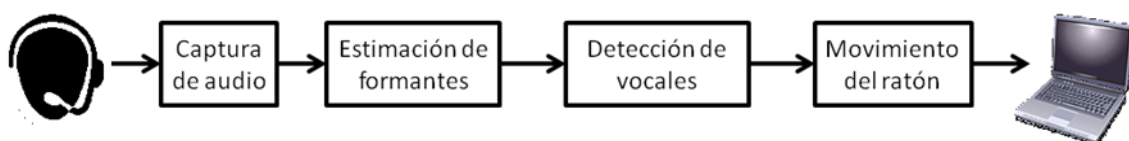


Figura 3: Esquema de movimiento del ratón con vocales

El interfaz visual de “VocalClick” se puede ver en la Figura 4. Permite la selección de diferentes elementos internos de la aplicación, así como adaptarlo a las características físicas del usuario (sexo y altura), ya que esta adaptación permite una mejor estimación y normalización de la vocal pronunciada. Al igual que “VozClick”, una vez iniciado el programa, puede minimizarse mientras sigue recogiendo la producción de voz del usuario para realizar el movimiento del ratón, haciendo más cómodo el trabajo con el ordenador al usuario.

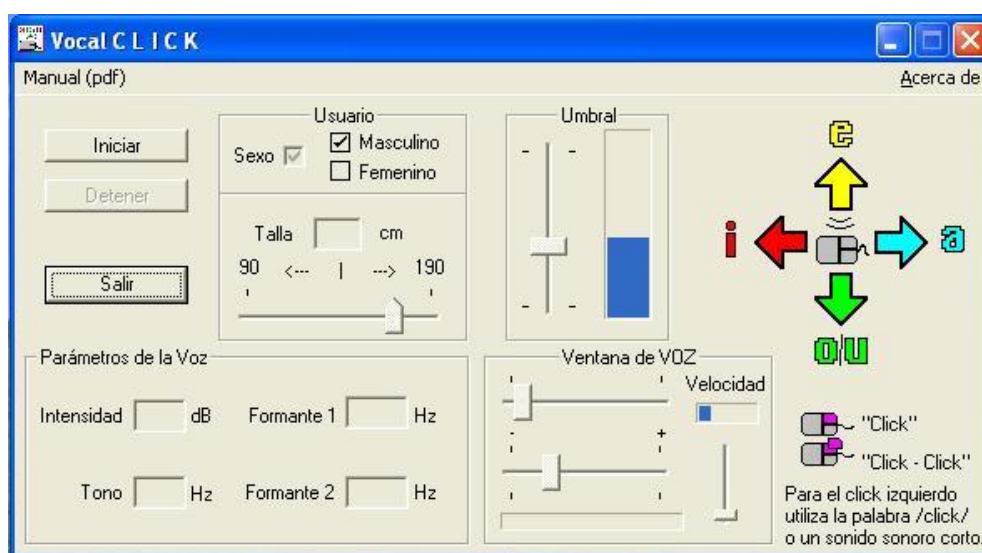


Figura 4: Interfaz visual de “VocalClick”

I.4. RESULTADOS Y CONCLUSIONES

El trabajo realizado ha supuesto una línea de trabajo constante de CADIS-Huesca y ASPACE-Huesca con el equipo investigador de la Universidad de Zaragoza durante los últimos 2 años. Este trabajo ha dado lugar a las ayudas técnicas propuestas en esta memoria con una buena acogida en el entorno de uso de las mismas. La herramienta “VozClick” es utilizada actualmente por usuarios de ASPACE-Huesca, habiéndose obtenido un gran resultado especialmente con una usuaria que ha visto acelerado su trabajo con procesadores de textos y navegadores web gracias a la herramienta desarrollada.

El trabajo con “VozClick” en el proyecto “Sistemas De Interacción Oral Sobre Dispositivos Portables” fue presentado el 20 de abril de 2009 a la finalización del proyecto bajo el auspicio de la DPH, el IASS y la CAI (Figura 5). Posteriormente, se divulgó la situación del mismo en foros del ámbito de la comunicación alternativa y las ayudas técnicas (Saz et al., 2009b)(Saz et al., 2009c) con una buena acogida en estos ámbitos. Ambos programas, “VozClick” y “VocalClick” se encuentran disponibles como software libre en la sección “Otras Ayudas Técnicas”² de la página de “Comunica” .

² <http://www.vocaliza.es/cadis.html>



Figura 5: Demostración en vivo de “VozClick” durante su presentación.

PARTE II. ACCESIBILIDAD EN INTERNET

II.1. INTRODUCCIÓN

Una de las aplicaciones más demandadas por usuarios con diversidad funcional es el acceso a la información en Internet. En el interés del grupo ViVoLab (GTC/I3A) por acercar las nuevas tecnologías a estos usuarios, se identificaron dos grupos de usuarios con diferentes problemáticas respecto a este punto: Por un lado, invidentes y deficientes visuales no pueden acceder a la información de la web debido a la gran importancia que se da al componente visual en la misma; y, por otro lado, usuarios con discapacidad física con dificultad en el manejo del ratón navegan de una forma muy lenta debido a que el ratón es la herramienta principal para la navegación en Internet y el acceso a los diferentes enlaces dentro de la red. La importancia de proporcionar inclusión a personas con diversidad funcional en la web ha llevado al World Wide Web Consortium (W3C) a definir los puntos clave para facilitar esta interacción (W3C, 2005).

En los dos casos comentados, existen soluciones generales de accesibilidad que también permiten el acceso a Internet, como los programas JAWS Screen Reader o IBM Home Page Reader y otros lectores de pantalla para invidentes; o elementos de barrido para el manejo del ratón, como se ha visto para personas con discapacidad motriz. Sin embargo, las posibilidades que ofrece el lenguaje HTML (HyperText Markup Language), que es el lenguaje de diseño web, para incluir múltiples elementos simplemente mediante “tags” o etiquetas hizo pensar en la posibilidad de crear una especificación de diseño de “web accesible” mediante síntesis y reconocimiento del

habla que fuese específica para la tarea de navegación y que pudiese ser incluida fácilmente por cualquier diseñador web.

Una de las especificaciones de diseño que se vio como requisito indispensable para la creación de esta ayuda técnica, era que se simplificase su uso lo más posible en el ordenador del usuario. Esto puede llevarse a cabo mediante tecnología cliente-servidor en arquitecturas distribuidas. Como muestra la Figura 6, en el sistema propuesto, el usuario accede a la página web mediante los protocolos habituales en la World Wide Web (WWW). La página puede seguir manteniendo su misma apariencia y diseño usando los lenguajes usuales de Internet, HTML, PHP, Javascript; lo que hace que la persona que no requiere la ayuda de accesibilidad no vea modificada su experiencia de navegación en la página. Si la ayuda de accesibilidad es requerida, la página se comunica con el servidor remoto y hace de puente para que el usuario pueda usar reconocimiento y síntesis de voz en su navegación por la página.

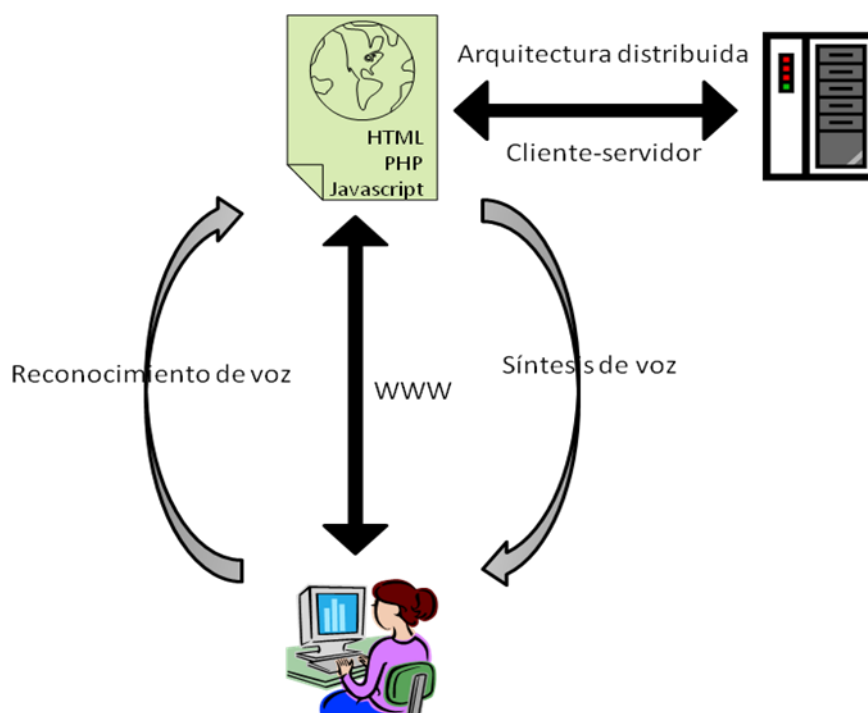


Figura 6: Esquema de uso de reconocimiento y síntesis de voz de forma remota para navegación web.

II.2. ACCESIBILIDAD WEB PARA INVIDENTES Y DEFICIENTES VISUALES

La definición de página web accesible surgió inicialmente como una propuesta llevada a cabo desde el grupo ViVoLab (GTC/I3A) junto con el Laboratorio Aragonés de Usabilidad para proponer la generación de una página web accesible a invidentes y deficientes visuales por parte de Aragón Radio, la cadena de radio de la Corporación Aragonesa de Radio y Televisión (CARTV), dentro del programa AVANZA por el que la Corporación lanzó la página web de su segundo canal de radio en Internet, Aragón

Radio 2³, una radio on-line que buscaba fomentar la participación y la inclusión de toda la comunidad aragonesa.

La herramienta principal que se pone en mano del diseñador web es el applet de síntesis/reconocimiento del grupo ViVoLab (vivoreco). Este applet de Java permite incluir de forma directa esta tecnología sin necesidad de instalar ningún elemento externo en el ordenador del usuario, simplemente mediante comunicación con un servidor que realiza todo el procesado (véase Figura 6). Desde ViVoLab se han creado una serie de reglas referentes a las etiquetas HTML a usar por el desarrollador web para definir elementos que serán sintetizados según el usuario navega por la página. Estas definiciones de etiquetas, junto con el manual de uso del applet, son los únicos requerimientos y conocimientos que un desarrollador web necesitaría para proporcionar accesibilidad por voz a su web (Saz et al., 2010).

Por parte del usuario, la navegación por esta página accesible se basa en el uso de teclas y comandos de reconocimiento de voz. Mediante el uso de las combinaciones de teclas definidas en la Tabla 1 se avanza o retrocede por las diferentes cabeceras y elementos sintetizables definidos en la página.

Combinación de teclas	Acción
Ctrl+a y Ctrl+d	Activar/Desactivar la lectura de la página
Ctrl+q y Ctrl+w	Lectura de la ayuda de síntesis/reconocimiento
Ctrl+r	Activación del reconocimiento de voz
Ctrl+(dcha) ó Tab	Lee la siguiente cabecera o elemento
Ctrl+(izda) ó May-Tab	Lee la cabecera o elemento anterior
Ctrl+Inicio	Lee la primera cabecera de la página
Ctrl+Fin	Lee la última cabecera de la página
Ctrl+(abajo)	Repite el último elemento leído
Enter	Accede al último enlace sintetizado

Tabla 1: Comandos básicos de navegación por la web accesible.

II.3. ACCESIBILIDAD WEB MEDIANTE BARRIDO

Por otro lado, CADIS-Huesca, consciente de la necesidad de incluir a las personas con discapacidad, ha solicitado al Ayuntamiento de Huesca que, de cara al nuevo diseño de su página web⁴ tenga en cuenta estas consideraciones e incluya contenidos de interés para la comunidad de personas con discapacidad y elementos de accesibilidad universal en, al menos, algunas secciones de la misma. Nuevamente pensando en las posibilidades que las tecnologías del habla podrían aportar en este caso, tras contactar con el grupo de investigación ViVoLab (GTC/I3A) se inició la creación de un estándar para la generación de páginas accesibles mediante síntesis de

³ <http://www.aragonradio2.com>

⁴ <http://www.ayuntamientohuesca.com>

voz que pudiesen ayudar a personas con discapacidad severa a seguir y navegar más fácilmente por la web.

La definición de página accesible para estos casos parte de la definición dada en la página web para invidentes, pero simplificando la interacción con la página, adaptándola a las mayores dificultades en la interacción de estos usuario. En este caso, la página web va leyendo de forma secuencial las cabeceras más relevantes de la página, volviendo a empezar una vez se llega al final de la página. Con una simple pulsación de la tecla “Enter”, el usuario puede pasar a leer los elementos dentro de una determinada cabecera, volver a leer las cabeceras o acceder a un enlace. El reconocimiento de voz

Esta forma de leer la página es similar a los programas de barrido tan habituales en el trabajo con gente con parálisis cerebral, por lo que es de fácil comprensión para ellos y sólo requiere de un pulsador para acceder a las páginas y a la información. Por esta misma razón, esta definición de “página accesible” permite la utilización del software “VozClick” en el caso de que se quiera utilizar pulsación emulada por voz en vez de pulsación mecánica.

II.4. RESULTADOS Y CONCLUSIONES

El trabajo en la definición y creación de páginas web accesibles mediante el uso de síntesis de voz y reconocimiento del habla para los grupos de trabajo mencionados está todavía en proceso y supone el trabajo de diferentes grupos (ViVoLab, Laboratorio Aragonés de Usabilidad) e instituciones ya mencionadas.

Dos demostradores se encuentran actualmente funcionando para dar a conocer las posibilidades de esta tecnología. El primero⁵, mostrando las potencialidades para invidentes, es una versión estática de la página web de Aragón Radio 2. Este demostrador permite navegar por varias secciones de la página (inicio, noticias, servicio público y podcasts), escuchar los dos canales de Aragón Radio y acceder a una selección de podcasts. La página web se usará para realizar una evaluación con usuarios reales en colaboración con la delegación provincial de la ONCE en Zaragoza. El segundo demostrador se ha desarrollado con las propuestas realizadas para el uso por usuarios con gran discapacidad física⁶. Este prototipo muestra las posibilidades que ofrece la definición de “página accesible” con una versión estática de la propia web del Ayuntamiento de Huesca alojada en los servidores del grupo ViVoLab.

La potencialidad futura de este trabajo y su necesidad y utilidad para la inclusión y accesibilidad en Internet, hacen de esta línea de trabajo una prioridad en el trabajo e investigación de todas las personas implicadas hasta ahora en los desarrollos realizados.

⁵ http://dihana.cps.unizar.es/~alborada/ar2/ar2_frames.htm

⁶ <http://dihana.cps.unizar.es/~alborada/ahuesca/ahuesca.php>

DISCUSIÓN FINAL

La colaboración entre las diferentes instituciones continúa de forma constante desde los proyectos iniciales y es la intención de todos los que han participado seguir con el mismo, mejorando las funcionalidades de los elementos ya existentes y abriendo vías en nuevas posibilidades que permitan acercar las TIC al colectivo de personas con diversidad funcional. Los resultados obtenidos hasta el momento impulsan esta colaboración y se espera que el trabajo fructifique en nuevos elementos e ideas que sigan mejorando la capacidad autónoma de estas personas.

Uno de los principales resultados de la colaboración llevada a cabo hasta ahora es la unión de diversas personas de muy diversos campos, para llevar a cabo un trabajo con un enfoque multidisciplinar. Estos campos, recogidos en el grupo de trabajo incluyen grupos tecnológicos (ViVoLab, GTC/I3A), grupos expertos en usabilidad (Laboratorio Aragón de Usabilidad) e instituciones asistenciales (CADIS-Huesca, ASPACE-Huesca). Por la experiencia adquirida durante este tiempo, sólo a través de la este tipo de equipos multidisciplinarios se pueden obtener resultados que sean novedosos a nivel tecnológico y útiles para usuarios reales que enfrentan día a día grandes retos para avanzar en su autonomía personal.

Por último, en todo momento se ha trabajado pensando en hacer llegar las tecnologías de accesibilidad al mayor número posible de personas, y por eso todos los desarrollos realizados están disponibles de forma libre y gratuita mediante Internet. Al igual que con Comunica, se cree que el acceso a los elementos que ayuden a integrar al colectivo de discapacitados en la sociedad debe permitirse sin ningún tipo de restricción al ser un derecho fundamental de estas personas.

REFERENCIAS

Abascal, J. (2008). No hay razón para la e-exclusión. *XI Jornadas Estatales de Terapia Ocupacional*. Madrid.

Bilmes, J., Malkin, J., Li, X., Harada, S., Kilanski, K., Kirchhoff, K., y otros. (2006). The Vocal Joystick. *2006 International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (págs. 625-628). Toulouse.

Harada, S., Landay, J., Malkin, J., Li, X., & Bilmes, J. (2008). The Vocal Joystick: evaluation of voice-based cursor control techniques for assistive technology. *Disability and Rehabilitation: Assistive Technology*, 22-34.

Hawley, M., Enderby, P., Green, P., Brownsell, S., Hatzis, A., Parker, M., y otros. (2003). STARDUST: Speech Training And Recognition for Dysarthric Users of Speech Technology. *7th Conference of the Association for the Advancement of Assistive Technology in Europe (AAATE)*. Dublin, Irlanda.

Iturriate, I., Antelis, J., & Minguez, J. (2009). Synchronous EEG brain-actuated wheelchair with automated navigation. *IEEE International Conference on Robotics and Automation*. Kobe, Japón.

Saz, O., García, J. E., & Lleida, E. (2010). *ACCESIBILIDAD EN LA WEB MEDIANTE VOZ A CIEGOS Y PERSONAS CON DEFICIENCIAS VISUALES*. Informe técnico, Universidad de Zaragoza, Zaragoza, España.

Saz, O., Lleida, E., Abarca, L., & Mejuto, S. (2009). MOUSECLICK: ACCESO AL ORDENADOR A TRAVÉS DE LA VOZ. *II Congreso Nacional de Comunicación Aumentativa: Comunicación y Tecnologías para la Vida*. Zaragoza.

Saz, O., Lleida, E., Abarca, L., & Mejuto, S. (2009). MOUSECLICK: ACCESO AL ORDENADOR A TRAVÉS DE LA VOZ. *IV Jornadas Iberoamericanas de Tecnologías de Apoyo a Discapacidad "Las Tecnologías De Apoyo En Parálisis Cerebral Y Alteraciones Afines"*. Madrid.

Saz, O., Yin, S.-C., Lleida, E., Rose, R., Vaquero, C., & Rodríguez, W. (2009). Tools and Technologies for Computer-Aided Speech and Language Therapy. *Speech Communication*, 948-967.

W3C. (5 de Mayo de 2005). *How People with Disabilities Use the Web*. Obtenido de <http://www.w3.org/WAI/EO/Drafts/PWD-Use-Web/>