Multimodal Interaction System for Disabled People

Use Case: Playback of multimedia content

Sebastián Sastoque H.
Universidad Militar Nueva Granada
Bogotá, Colombia

Soraya Colina
Clinical Phonoaudiologist and Researcher
Bogotá, Colombia

Marcela Iregui
Universidad Militar Nueva Granada
Bogotá, Colombia

ABSTRACT
In this paper, a multimodal interaction system for people with disabilities is presented. The system combines four different interaction techniques in a single graphical user interface for playback multimedia content. The combination of different interaction modalities aims to expand the number of potential users, by allowing the selection of appropriate interactions that fulfil their specific needs. In this work, we validate, for a study case, the idea that different methods of interaction allow a higher accessibility, this way, individuals experiencing permanent or temporary disabilities could benefit from technology applications.

Keywords
Human Computer Interaction, Interaction Techniques, Multimedia, Multimodal Interaction, Augmentative and Alternative Communication (AAC), Accessibility.

1. INTRODUCTION
People with disabilities face numerous daily challenges. Unfortunately the use of computers for these individuals is one of them. Although computers are becoming more advanced, people with disabilities encounter several problems that limit their access to technology. For a long time getting computers to be accessible for people with certain disabilities, was achieved. In many cases the approach was catering to the specific needs of a particular user. In that case, Human Computer Interaction practices focusing on people with disabilities, were made initially from handmade adaptations of surprising quality and usefulness. The disadvantage was being unable to perform effectively with others disabilities.

The mouse and keyboard, traditional devices for computer interaction, are useless for people with disabilities. For instance, Parkinson patients can hardly use a mouse because of the required precision to achieve any task [4]. Likewise, graphical user interfaces are usually very complicated for people with cognitive limitations, such as patients with aphasia who present problems interpreting complex instructions or processing the information in parallel. Interfaces with too many menus and buttons causes people with disabilities to refuse the use of computers because they get confused[8].

As Dumas et al [5] defined, multimodal systems interpret information from combining user input modes. They offer alternatives for human machine interaction that involve diverse and underserved users groups, achieving universal access. Therefore, different research projects, concerning multimodal interaction for people with disabilities, have been developed. The authors of *MailSaw* and *NaviSaw*[7] present a system for users with visual disabilities to browse the Internet, using voice synthesis, speech recognition and mouse commands to interact. In the project *BlindAid*[9], users interact with a haptic device in a 3D interface, through various unknown spaces, in order to carry out therapeutic activities for blind people rehabilitation. There are also several applications of unimodal interaction, such as mouse control from head movement[6], eye gaze[11] and hand gestures[12]. Some applications use virtual keyboards, speech recognition systems, communication boards and screen readers[1].

The applications referred above are based on the traditional model of HCI practices for people with disabilities, which focus on satisfying the needs of a specific group of users, limiting its use to a restricted group of users. For this reason, in this paper a multimodal interaction system is proposed, which allows to manipulate an application interface in different ways. It could be configured according to each user's preferences. The four methods of interaction implemented are: detection of head movements, image detection and recognition, trackball manipulation and command selection with an infrared pen (IRPen) in a GUI projection. The main goal of this work is to show, with a study case, that multimodal interaction improves applications usability, making them broadly accessible. The use case selected for validation consists of an application for playback multimedia content using the four methods of interaction mentioned above.

This paper is organized as follows: next section deals with the description of the use case. The third section focuses on the explanation of the four interaction techniques implemented, and their methods. The fourth section describes the proposed system with each of its components and their relationship. Finally, the last sections correspond to discussion of the results and conclusions of the work.

2. USE CASE

Multimedia technologies provide fuller entertainment experiences that generally are not accessible to disabled people. That is why, the selected use case consists on a playback multimedia content application, considering that many current players have usability limitations. This work is focused on two issues: The first one, concerning the design and implementation of accessible interaction techniques for people with sensory, motor or cognitive disabilities. The second one, concerning the software application design.

3. INTERACTION TECHNIQUES

The interaction methods implemented in this work, i.e. detection of head movements, image detection and recognition, trackball manipulation and command selection with an IR pen in a GUI projection, were designed based on the requirements of users with reduced motor skills and difficulties to follow complex instructions.

3.1 Head Movements Detection (HMD)

For users with reduced motor skills (i.e. quadriplegics), detection of head movements provides the possibility to control a software application. For the implemented use case, the instructions are identified by tracking the direction of the head movements, as shown in figure 1. That is, horizontal movement of the user's head is used to navigate between functions and vertical movements to perform a selected action.

The method implemented for motion detection is based on the *Frame Difference*[10] algorithm. Once the frame differences are performed, the movement areas are detected and the centers of these two bounding boxes are calculated to find the angle between these two points. According to the angle value the movement is associated with a direction between up, down, left or right. This process is repeated for the next 15 frames (1 second). The final decision is determined by calculating the more frequent direction.

3.2 Image Detection and Recognition (IDR)

The use of symbols or pictograms is broadly used in Augmentative and Alternative Communication (AAC) systems, specially for people with communication disorders (i.e. aphasia, autism) [2]. That is to say, images can be easily associated with an idea and then with an instruction.

Fifteen (15) standard symbols were stamped in plastic cards, each one related with an action on the playback multimedia system. Once a symbol card is placed in front of the camera, the system recognizes the symbol printed and performs an

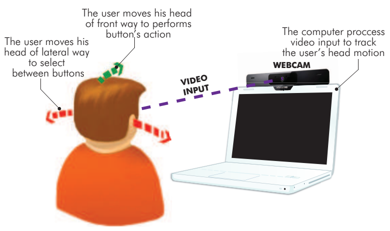

Figure 1: Head Movements Detection Interaction

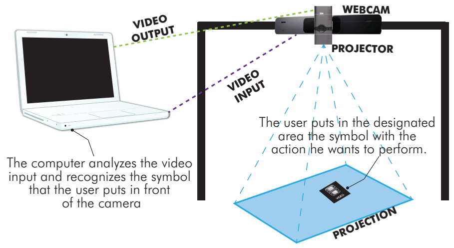

Figure 2: Image Detection and Recognition Interaction

action according to the associated instruction, as shown in figure 2.

When the video detects a symbol card the system proceeds to the recognition. The method used for image classification is based on the well known emphSpeeded Up Robust Features (SURF)[3] algorithm. In that method, feature extraction is performed by doing a scale-space representation of the image, followed by detection and description of the interest points. Secondly, the classification is performed by features matching using the nearest neighbor criteria.

3.3 Command selection with an IRPen (CSIR)

For some users it is difficult to use a mouse in order to interact with a graphical user interface, due to the precision required for that task (i.e Parkinson). The method proposed to overcome that limitation is the use of an infrared pen to select the action from a projection of the GUI in a work surface. This method requires a particular arrangement of devices consisting of a projector, an IRPen and a Wiimote disposed as shown in the figure 3. Once the IR pen touches the surface, it emits a signal which is detected by the Wiimote and then transmitted to the application via bluetooth. The IR pen coordinates are used to select the action from the GUI.

3.4 Trackball Manipulation (TM)

Some users, particularly those with visual limitations and motion reduced upper limbs endure problems with traditional interaction systems [4]. Interaction based on movement fits these kind of users (see figure 4).

This interaction is based on gesture characterization from the movements performed by the user in a trackball device. These gestures are grasped from the momentum and the mo-

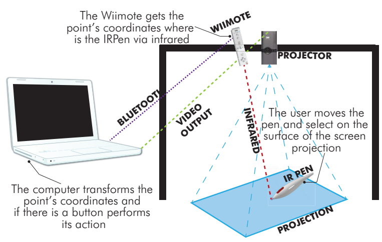

Figure 3: Command selection with an IRPen Interaction

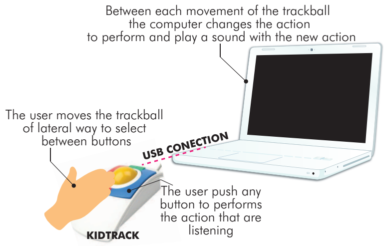

Figure 4: Trackball Manipulation Interaction

tion direction. The Euclidean distance between the starting point and the ending point is calculated to determine the displacement and the angle formed by these two points to establish the direction.

4. SYSTEM DESCRIPTION

The playback of multimedia content system, shown in figure 5, has a step by step choice where the final result is the reproduction of media. The system was designed to enhance the usability for people with limitations.

The system uses the same GUI for the different interaction modalities. The user first selects a content type and the *Interaction Analysis Module* interprets the user's input. Subsequently, the *Content Selection Module* returns the list of media for the selected type. To select the option, the Interaction Analysis Module interprets the user's action and finally, the *Content Playback Module* decodes the file and displays the content. The considered actions for this use case are playing, pausing and stopping the current content; looking for the next or previous content; and selection of music, text, images or video.

Summing up, the system is composed of three main modules, which are:

1. **Interaction Analysis Module:** The IAM applies algorithms and methods to grasp the user preferences from the different interaction possibilities and translates them to system instructions.
2. **Content Selection Module:** The CSM searches the indexed content locations for a specific content type.
3. **Content Playback Module:** The CPM opens and controls the media file. Mainly, it decodes the file and displays the content.
4. **Multimedia Content Indexes:**The structure of indexes of media files is organized in an XML file into four categories, one for each type of content (audio, video, text or image). Each indexed file in a category has the *ID*, *Title*, *Author*, *Duration* and *Location* attributes.

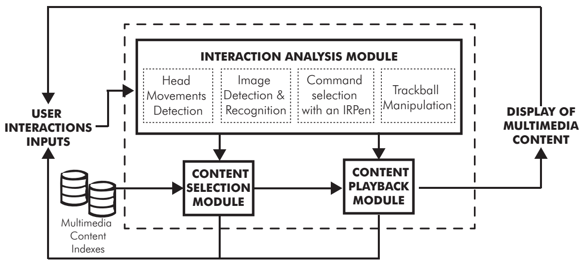

Figure 5: General description of the system

Figure 6: 55 years old woman with Broca's aphasia using the system

5. RESULTS

A case study of a 55 years old woman with Broca's aphasia due to stroke in the Sylvian fissure of the left middle cerebral was considered. This person has deficits in language production and comprehension, she has gross motor skills, and she had previous contact with technology therefore knew how to use a computer. Before the evaluation, an explanation on the using of the system was given to the patient (see figure 6).

The following activities were performed with an allowed maximum of five (5) attempts: Play a song, play a video, find an image of your home, play the second video from the list, display a text, navigate in a text, show the duration of a song, show a song artist and show the title of a song.

Each activity was performed by using isolated interactions, and using multimodal interaction (see figure 7). For each trial, the time undertaken for each activity was measured, from the moment when the system shows the options for choose until the person achieve the task. When the user failed to do the task, the time is measured again from the beginning in a new attempt. The results reported in table 1 is the time of the first attempt that the user achieve the task.

A comparison was made with certain players using the mouse for the interaction, for the same tasks and test conditions mentioned above, obtaining the results shown in table 2.

Figure 7: Use of the proposed interaction

Table 1: Comparison between Interaction techniques results (seconds to achieve a task)

Task	HMD	IDR	CSIR	TM	Multimodal
Play a song	23s	21s	20s	28s	20s
Play a video	42s	38s	39s	40s	37s
Find an image of your home	78s	67s	72s	69s	65s
Play the second video from the list	39s	37s	35s	40s	35s
Display a text	31s	28s	38s	32s	30s
Navigate in a text	21s	18s	23s	20s	15s

Table 2: Comparison between players (seconds to achieve a task)

Task	iTunes	Windows Media Player	VLC	Our System
Play a song	50s	57s	63s	20s
Play a video	48s	60s	53s	37s
Find an image of your home	Player doesn't have this function			65s
Play the second video from the list	45s	60s	Don't achieve	35s
Display a text	Player doesn't have this function			30s
Navigate in a text	Player doesn't have this function			15s
Show the duration of a song	10s	12s	8s	4s
Show a song artist	15s	22s	20s	3s
Show the title of a song	20s	17s	25s	6s

6. DISCUSSION

Analyzing the obtained results in the study case, it is straightforward to affirm that the proposed system is usable for people with disabilities related with aphasia. This is due to several facts:

- The way in which the options are sequentially presented, since the user must not process a high number of information to decide what content to play, although this can become rather uncomfortable and unnecessary for a user who doesn't have a cognitive disability.
- Each interaction evaluated in an isolated manner improves usability due to the implementation of accessible interfaces adapted to specific user disabilities..
- The possibility of having multimodal interfaces allows fast interactions, because the user has different alternatives to attain the control of the application. Unlike isolated interactions that provide easier interactions for some options but not for the whole set.
- The proposed system is adequate and efficient for users with disabilities compared to other systems of the same kind. This is confirmed by the shorter amount of times used to successfully utilize applications as reported by the proposed system. This, being a result of accessible interfaces and simpler graphical interfaces. Moreover the implemented application presents a use case richer in options because it is adapted to different media as text, music, video and images.

7. CONCLUSIONS

This work has demonstrated that an application, with user driven interfaces for disabled and multimodality, enhances the accessibility and as well as the user's experience. In doing so, people experiencing a permanent or temporary situation of disability could benefit from the use of entertainment technology.

The development validates for a study case the fact that multimodal techniques permit the appropriation of technology and could be considered to be a applied in software applications.

In order to have more facts allowing the generalization of the conclusions, further work must be accomplished to develop an extended quantity of use cases that enable the evaluation of user experience variables in different scenarios. Additionally, this work must be validated by people with different disabilities and the population must be extended for each condition.

8. ACKNOWLEDGEMENTS

This work has been funded by the Universidad Militar Nueva Granada in the framework of their PIC projects program.

We thanks Helioth Sánchez and Eduard Sierra for useful discussions and the comments provided during the development of this project.

9. REFERENCES

[1] J. Abascal and R. Moriyón. Tendencias en interacción persona computador, 2002.

[2] C. Basil, E. Soro-Camats, and C. Rosell. *Sistemas de signos y ayudas técnicas para la comunicación aumentativa y la Escritura*. Masson, 2004.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346 – 359, 2008.

[4] A. M. Cook, J. Polgar, and S. Hussey. *Cook and Hussey's assistive technologies: Principles and practice*. Mosby Elsevier, St. Louis, 2008.

[5] B. Dumas, D. Lalanne, and S. Oviatt. Multimodal interfaces: A survey of principles, models and frameworks. In *Human Machine Interaction*. Springer Berlin / Heidelberg, 2009.

[6] J. Gips, M. Betke, and P. Fleming. The camera mouse. In R. Press, editor, *Proceedings of RESNA 2000*, pages 98–100, 2000.

[7] J. González, F. Sánchez, A. Lozano, M. Macias, and F. Sanchez Herrera. Navisaw y mailsaw herramientas multimodales para el acceso a internet para usuarios con discapacidad visual. Technical report, Universidad de Extremadura, 2007.

[8] M. Iza. Tecnología computacional en afasia. *Revista de Psciología General y Aplicada*, 56(1):101–123, 2003.

[9] O. Lahav, D. Schloerb, S. Kumar, and M. Srinivasan. Blindaid: A learning environment for enabling people who are blind to explore and navigate through unknown real spaces. In *Virtual Rehabilitation*, 2008.

[10] P. Rivas Perea and M. Chacón Murguía. Evaluación de métodos de detección de movimiento. In *ITCH - ELECTRO 2005*, 2005.

[11] L. E. Sibert and R. J. Jacob. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 281–288. ACM, 2000.

[12] X. Zabulis, H. Baltzakis, and A. Argyros. Vision-based hand gesture recognition for human-computer interaction. In *The Universal Access Handbook*. LEA, 2009.