THE POTENTIAL OF MULTIMODAL INTERFACES FOR THE BLIND: AN EXPLORATORY STUDY

Yacine BELLIK* & Dominique BURGER**
*LIMSI-CNRS, **INSERM, France

ABSTRACT
Multimodal interfaces are the subject of considerable research world-wide. This paper describes a joint project funded by two French research organisations, with the cooperation of an institute for the blind. This project applies multimodal interfaces to provide improved computer access for the blind. A multimodal text editor is described. It has been designed to provide enriched texts, direct manipulation and immediate feedback to text editing tasks, such as reading and modifying text attributes, selecting, copying, moving, or deleting parts of text, navigating within the text and getting additional information on words (definitions, annotations...).

KEYWORDS
Non Visual Interfaces, Multimodal Interfaces.

INTRODUCTION
The developments that have taken place in Human-Computer Interaction (HCI) over the past years have had contradictory effects on blind people (3). The advent of refreshable electronic Braille displays, and text-to-speech synthesisers in the late seventies opened up new opportunities for the social and professional integration of the blind. Many kind of jobs and activities became accessible thanks to workstations that were adapted to their special requirements. Despite certain difficulties in making the adapted interface as friendly as were screen-based interfaces, this was a time of great hope. The graphical user interface (GUI) revolution that has recently overtaken micro-computers has become a source of worry for the blind (4). One reason is that the bitmap display technology on which these interfaces are based means that access products based on the principle of reading characters from screen memory can no longer be used. But more fundamentally, blind people were led to believe that new generation software would result in the screen being covered with images, making its meaning impossible to decipher. "Because developers used a heavily visual design metaphor, it was assumed that blind people would never be able to access GUI systems" (1). The gradual evolution of software architecture, coupled with successes in research and development on HCI, now seems to be about to create new conditions for the design of non-visual human-computer interfaces that will provide a better match for needs of blind users (2). This study resulted in the establishment of a research and development program with the objective to explore the potential of multimodal solutions for constructing non-visual interfaces. This paper describes the initial results of the research program: a prototype application for text editing (Meditor). It is important to notice that this work does not deal with the problem of GUI accessibility by the blind. However, it aims at showing that multimodality constitutes an important factor which must be integrated in the solution of this problem, to have friendly non-visual interfaces.

MULTIMODAL INTERFACES
Human beings perceive the external world using the five senses of touch, hearing, sight, smell and taste. They interact with it by producing sounds, gestures, etc. These are referred to as communication modes, and they define the nature of the information used to communicate (visual mode, sound mode, gestural mode, etc.). A modality defines a particular form of a communication mode. Thus, noise, music, and speech are modalities of the sound mode. In most situations, natural communication between human beings is multimodal, as it combines several modes and modalities. Computer scientists are interested in producing multimodal interfaces that can integrate multimodal features and so make computer behaviour closer to human communication paradigms, and therefore easier to learn and use. This has been possible with the advent of new devices (speech recognizers and synthesizers, datagloves, touchscreens...). Multimodal interfaces are the ultimate HCI systems that can manage various and complex I/O information (5).

MEDITOR: A MULTIMODAL TEXT EDITOR
Meditor has been developed bearing in mind the many problems that blind people encounter when they have to access or manipulate textual material. Not only is text material deprived of some of its original features when it is presented in Braille on paper or by a refreshable Braille display, but manipulations as simple as underlining or annotating texts are a great challenge. This is particularly dramatic in education, where paper-based documents
The Potential of Multimodal Interfaces for the Blind

are fundamental for many exercises in which children construct, verify and reinforce forms of thinking by carrying out specific manipulations. We have consequently designed Meditor to overcome these difficulties, and to be an interactive support for the development of teaching and training material for blind students. The design work was carried out in close cooperation with two blind users, a teacher and a Braille proof-reader.

Hardware configuration
The computer environment in which Meditor was implemented is the basic workstation used by blind students at their institute. It includes a desktop PC, a Braille terminal and a French text-to-speech synthesiser. A speech recognition system has been added (fig. 1).

![Figure 1. Meditor: A multimodal text editor.](image)

Each cell of the Braille terminal presents a character on an 8-dot Braille matrix. The Braille terminal used, allows users to point at any character within the line by pushing a button located in front of the corresponding cell. The speech recognition system is speaker-dependent and allows recognition of continuous speech. A vocabulary of about 60 words is used.

Functionality
Meditor was designed to provide blind users with the main functions of a simple full-page editor. The user can perform the following tasks:

- read a text with embedded character attributes such as style, colour, font, or size
- select, copy, move, or delete parts of a text with immediate feedback
- modify text and/or attributes
- search strings or text with specific attributes
- read parts of text, using speech synthesis
- insert, consult and modify additional information on words (annotations).
- etc.

Design rules
Concerning inputs, the general principle consists in selecting an entity and then apply a command to it. By adding speech, several types of selections can be made through the same pointing operation. The user has just to say the name of the entity while clicking on any character of this entity. For instance, to select a sentence he has to say "sentence" while clicking on any character of the desired sentence. To select a paragraph he has to say "paragraph" instead of "sentence". If a command is executed without any explicit selection, it is applied to the word in which the pointing character belongs. The standard keyboard or the Braille keyboard are used to enter text, file names, searched strings etc.

Concerning outputs, the Braille display is used for text presentation, while auditory messages (speech and beeps) are used to deliver feedback information about user commands, and information about the displayed text (current line number, number of pages, etc.). Attributes are presented through 3 levels. When a line containing particular attributes is displayed, a beep is produced. This is the first level. It indicates that particular attributes are present within the current line. The eight-dot matrix of the two first characters of the concerned entity are then set up\(^1\). This is the second level. It informs the user more precisely about the location of the particular attributes. If the user wants more information about the type of attributes (colour, style, etc.) and the type of the concerned entity (word, sentence, paragraph...) he has just to click. A speech message containing the whole information is then synthesised (fig. 2). This is the third level.

![Figure 2. Attributes presentation.](image)

Interaction examples
To put a word in italic

\(^1\)Giving that blind are used to read 6 dots Braille, only the two first characters are marked to ensure that the information will be detected without disturbing the reading process.
The user says "italic" while clicking on any character of the word on the Braille terminal. **To place a character into an exponent position**

The user says "character exponent" while clicking on the corresponding character. **To delete a part of the text**

1) The user says "begin selection" while clicking on the first character of the string to be deleted, 2) then says "end selection" while clicking on the last character, and 3) says "delete" to complete the command. The message feed-back "selection deleted" is then generated by the speech-synthesiser.

First stage evaluation

The application is currently being thoroughly evaluated and the results will be soon available. We describe here the preliminary evaluation that has been done with two blind subjects.

Evaluation conditions.

The first stage evaluation of Meditor was based on three simple but realistic tasks carried out by the two blind subjects during two trial sessions of two hours each. The following three tasks were performed:

**Task 1**: An exercise in grammar. The subjects had to colour code words of sentences according to their grammatical class. Verbs were made red, while nouns were made green.

**Task 2**: To read a text on prehistory. The definitions of some specific words were available on request, via the speech synthesiser. A special attribute was used to mark these words. The command to access the definition of a word is: say "definition" while pointing at the word.

**Task 3**: Text-editing. The subjects were asked to correct spelling errors in a text and to reorganise it using "Cut", "Copy" and "Paste" commands.

Results

Both subjects agreed that the interface was intuitive. The combination of auditory and tactile information to translate the attribute information did not disturb their reading of the text. They found no difficulty with the use of visual terms for colours that were considered as particular attributes. The combination of pointing gestures and spoken commands appeared to be quite natural. Only a few minutes explanation was necessary for them to understand and memorise the commands, the vocabulary and the principles of interaction. Nevertheless, they were given a four-page Braille direction for use to help them remember the vocal commands. It was almost never used, indicating that the memory load was minimal. Undoubtedly, more efforts would have been required to memorise keyboard functions and combining keys on the keyboard, as when people can not use a mouse. The control by speech allowed the subjects to concentrate on tactile information and to perceive immediate feed-back of commands. It also alleviated the workload as physical actions are minimised. The interaction principles seem quite close to direct manipulation in a GUI, as users can directly act on the objects through pointing operations. The principle of combining speech and pointing gestures, considerably increases the power of expressions and the interaction rapidity. For instance, a command such as "Sentence bold italic" while pointing at a character of the sentence, represents a contraction of several commands in a classical text editor.

CONCLUSION

This paper has described a study exploring some aspects of a multimodal solution to the problem of computer accessibility by the blind, through a concrete application. This study has indicated that multimodal interfaces represent a promising solution to improve computer access by the blind. An interesting perspective now is to study how multimodality can be integrated in the problem of GUI accessibility. The study has also revealed the need for a multidisciplinary approach in which blind users, specialists in fields, such as cognition, education, ergonomics, computer science and neuroscience, cooperate to create the models and guidelines necessary to design non-visual interfaces.

ACKNOWLEDGEMENTS

We wish to thank Y. Felix, J.F. Colas, C. Coudert and A. Junior from the INJA institute.

REFERENCES