# Sound, Space, and Metaphor: Multimodal Access to Windows for Blind Users

**Aude Dufresne,** *Centre for Information Technology Innovation*
**Odile Martial,** *Centre for Information Technology Innovation*
**Christophe Ramstein,** *Centre for Information Technology Innovation*
**Philippe Mabilleau,** *Centre for Information Technology Innovation*

As Graphical User Interfaces (GUIs) are now the standard in human-computer interaction, developing alternate modes of interaction for the blind becomes increasingly important. The goals of the PC-Access project were not only to provide blind users with direct access to GUIs within existing applications but also to allow blind users to reap the benefits of direct manipulation of these interfaces. To keep up with the spread of GUI applications, commercial screen readers are being adapted to these new interfaces (ANPEA, 1996; Andrews, 1995). These screen readers provide access to some of the graphical information on the screen by using voice-synthesis or braille and keyboard commands. These non-visual interfaces generally provide a concise verbal description of graphical elements along with linear access to the screen contents through commands and tabs. Prior to forming a mental image of concepts such as spatial arrangement and elements like icons, menus, window boxes and borders, non-visual interfaces can be very difficult to understand and manipulate. While GUIs may have made using computers easier for sighted people, the blind user remains at a disadvantage.

On the other hand, some research projects, such as GUIB and System 3, try to enrich the interface by using direct manipulation (Petrie & Gil, 1993; Vanderheiden, 1991). The Mercator System, on the other hand, moves away from spatial analogies and tries to redefine non-spatial audio interface for the X Windows system (Edwards, Mynatt, & Stockton, 1995; Mynatt & Webber, 1994). Unlike most current research projects, which try to improve either direct manipulation or earcons for visually impaired users, the PC-Access1 interface provides blind users with a multimodal access to Windows 3.11 that includes both earcons and direct manipulation (ASSET Proceedings, April 11-12, 1996; Ramstein et al., 1996). Following a brief description of the PC-Access interface, this article describes the design process used to define the interface's earcons. We have already completed several phases of the evaluation, and our results show that PC-Access is an efficient non-visual interface.

## The PC-Access Interface

This study focuses on three basic premises: (1) the use of metaphorical symbols is important in reducing the cognitive load, (2) direct manipulation will provide a better access to the spatial organization of the screen, and (3) multimodality is essential when substituting for sight. Two issues raised by this system are the importance of restoring the spatial properties of the screen and pointing (Martial & Dufresne, 1993). PC-Access was designed to integrate both modalities in order to provide barrier-free access to the graphical information in Windows 3.11 (Ramstein et al., 1996).

Using these principles, two versions of the PC-Access interface were in order to give blind users access to GUIs. Both versions include a mouse and a tablet onto which the screen is projected in absolute coordinates (relative exploration is impossible in the absence of feedback), and both versions use earcons to translate the screen's graphical elements (icons, window borders, buttons, menus, etc.). One version uses a commercial drawing pad and mouse (Martial & Dufresne, 1993). In the second version, force feedback is introduced by enhancing the mouse through the addition of a Pantograph (Ramstein, 1995). The blind user can sense the position of the mouse relative to the screen and thus retains the position of, and locates specific objects on the desktop more easily. Earcons, along with voice synthesis, were used to identify objects and events (Ramstein et al., 1996).

## Designing a Multimodal Interface for Windows

Designing a multimodal interface for a system as complex as Microsoft Windows is not a trivial task. Methodologies, general guidelines, and evaluation grids have been proposed to help design usable interfaces. But virtually nothing exists that can translate graphical information into a non-visual form for blinds users. Because the PC-Access interface is only a transposition, our objective was to try to provide blind users with access to the same feedback that a sighted person has. Internal and external coherence were also important as was reducing memory requirements by limiting the number

of lexical elements. Making use of perceptive and cognitive considerations that are specific to blindness when designing the multimodal interface allowed us to enrich the analysis by situating it within a semiotic framework.

## Perceptual and Cognitive Considerations

To replace sight, which is a rapid and precise means of conveying graphical information about many objects simultaneously, it seems essential to investigate the use of many interactive modalities. While voice synthesis and braille can be used for textual information, we propose facilitating access to graphical information and allowing direct manipulation through the use of tactile and auditory modalities. Several studies have explored the use of "earcons" to complete the graphical metaphor (Blattner, Sumikawa, & Greenberg, 1989; Buxton, 1989; Brewster, Wright, Dix, & Edwards, 1995; Gaver, 1989; Leinmann & Schulze, 1995). In the case of the blind, we found it difficult to preserve the diversity of the graphical display while transposing it to auditory stimulation. One dimension that was especially difficult to transpose was the simultaneous perception of elements, for exploration without the benefit of sight remains sequential. Nevertheless, the use of a pointer, spatial organization, and exploration with both hands (manipulating the pointer and sensing the tablet) should translate into a net improvement over screen-reading when using only the keyboard.

To design an effective sound interface, it is important to consider the perceptual differences between sound and vision. It is difficult to substitute for vision, which is highly parallel and which makes it easy to combine different characteristics (open the folder that is selected). Sounds, however, are difficult to differentiate and combine effectively. Sounds can only be accessed dynamically through either the user's exploration as commands or as system feedback. Because they are dependent on the user's or on the system's activities, sounds are harder to interpret than are visual cues. The context in which sounds are to be used must be considered, and their timing must be designed to coincide with the timing of activities. Long-lasting continuous sounds are annoying, especially for activities such as writing. Hence, continuous sounds can be used only if objects are small. On the other hand, in order to perceive and manipulate small objects, it is important that they be associated with longer sounds. Pitch and stereo effects are interesting, but not everyone is equally sensitive to them; relative differences such as increasing or decreasing pitch are more easily perceived.

## A Semiotic Framework

Interfaces equipped with icons and direct manipulation have now become real tools for communication. In this context, we adopt a semiotic framework that helps us describe the "Visual Rhetoric" of GUIs (Marcus, 1991; Familiant & Detweiler, 1993). Our objective is to translate graphical elements into significant non-visual information for blind users. The debate surrounding the design of interfaces focuses on whether icons should be iconic at all. In their relationship to their metaphorical referents, the screen's graphical symbols can be classified as icons, indices, or symbols. The three semiotic levels of analysis (semantic, syntactic, and pragmatic) correspond, respectively, to metaphor, spatial organization, and direct manipulation of the interface's objects. To translate a graphical symbol into a meaningful earcon, it is necessary to grasp its essential meaning; this theoretical framework has helped us in our analysis.

## Three Design Principles and Sounds

Having presented our approach, which is based on a realistic design, we can now define how sounds are used in respect to the three main principles of the PC-Access interface: metaphor, direct manipulation, and multimodality.

### Sound and Metaphor

The PC-Access design requires that the Microsoft Windows interface be transposed onto a new interface that includes earcons. We first have to analyze the relationship between the graphical symbol and its referent through the desktop metaphor (at the semantic level). While some authors have proposed transforming metaphors for the blind, our observations of blind users at work revealed that, in practice, the blind rely even more heavily on the physical properties of objects than do sighted users (Martial & Dufresne, 1993). Blind users define a place for everything, order their diskettes, put recognizable cues into documents, etc. We found no reason to believe that a new metaphor would be more useful to them than the physical and the spatial metaphor. The notion of professional integration also leads us to think that it is important for the blind user's environment to be as similar to the sighted user's environment as possible.

### Sound and Direct Manipulation

In the PC-Access interface, the same device (a mouse) provides information to the user and serves as the means of manipulation. Direct manipulation is thus a very important aspect of the interface. When the user explores the screen, the tablet provides spatial feedback, which is needed to understand the placement and scale of objects (on the syntactic level) and to allow for motor perception and manipulation. Spatial information (iconic or graphic) is perceived directly and therefore does not have to be translated into sound (except when generated by the system--e.g., in describing the position of the dialog box).

Two-handed exploration was used, which is important in situations requiring scaling and rotation. When modifications to the screen occur (e.g., dialog box or error), the user is informed of the modification through a voice-synthesis messages, which guides the exploration. Every action and event also has its auditory translation on the pragmatic level.

## Sound and Multimodality

The PC-Access system is essentially multimodal as it uses sounds, voice synthesis, and spatial information. In designing the system, it is thus important to optimize the use of multimodality. Too much stimulation can create interference as well as increase physical and cognitive stress. Based on previous experiments, we tried to use each of the various modalities for the task to which it seemed best suited. For those cases where multimodal stimulation is available, we intend to analyze the use and the performance of each of the existing modalities.

## Creating the Sounds of PC-Access

The debate surrounding the design of interfaces focuses on whether icons should be iconic at all. Many types of relationships may exist between a sign and its referent, and diverse taxonomies have been proposed to distinguish, in particular, between icons that bear a resemblance to the referent and more arbitrary symbols. Familiant et. al. also propose to differentiate signs along other dimensions: direct reference vs. indirect reference, graphic vs. alphanumeric, and part-part vs. part-whole (Gaver, 1989). In order to simplify the system, it is also important to structure stimuli by creating hierarchical families of easy-to-learn symbols (Blattner, Sumikawa, & Greenberg, 1989). Various criteria for designing icons can be applied to the design of earcons: closure, continuity, symmetry, simplicity, unity, clarity, and consistency.

To define these earcons, we had to describe the elements of the Windows interface systematically. This description followed a semantic scheme whereby each element was described according to its class, parts, and properties. Actions were also described for GUI manipulation and text reading, and a structured set of stimuli was designed for each modality of the new interface. Table 1 presents examples of the transposition of some Windows elements and actions. The arrangement of sounds was adjusted to ensure discrimination as well as to resolve interference, timing, and volume problems.

| Objects | | Sound | Force-feedback |
|---|---|---|---|
| Scroll bar | | | |
| part | arrow top/bottom | scroll - short high or low note | |
| part | arrow left/right | scroll - short note from left or right | |
| part | scrolling box | scroll - short sound | |
| part | scrolling bar top/bottom | scroll - short relative high or low | Scroll border |
| part | scrolling bar left/right | scroll - relative left or right | Scroll border |
| Reducing box | | reduce - short sound | Box border |
| Title bar | | move - short sound | Title border |
| **Actions** | | | |
| Scrolling | | | |
| up/down | | scroll - scale rising or falling | |
| left/right | | scroll - stereo going from left to right or reverse | |
| Reducing | restoring | reduce - longer with more echo or same reverse | |
| Moving | | move - longer with more echo | Elastic force to origin |

**Table 1. Examples of the transposition of graphic elements onto earcons**

**Semantic associations:** Sounds are used to facilitate the identification of objects and events. To make learning easier, a hierarchical structure using a limited number of different stimuli (24 earcons) was created that used variations in terms for grouping. In the final paper, a complete table will present the different earcons as they are organized in terms of objects and of actions or events.

**Families:** The sounds of objects are heard when the mouse passes over them (passing over a border, a title bar, or a selected window). Objects and actions are linked because some objects are also tools that perform different actions (reduction box or scroll bars). In these cases, the sound of the object was designed to suggest the action being undertaken. For example, the title bar sound might be a short "movement" sound while movement involving the title bar might be represented by the same sound played for a longer period of time.

**Variations:** The state of an object (whether it is selected or shaded) is conveyed by variations in volume and intensity. Selected options are indicated

by an added tone.

**Symmetry:** To help the user rearrange the position of windows and dialog boxes, it is important to determine if the mouse has entered or exited the objet. Because passing over a border is a very short event, a symmetrical sound could not be perceived (in greater than out vs. out greater than in), so an "exiting sound" was added to the border sound to indicate that the user had exited the object. This sound also serves as an alarm in cases where the function that has been selected is not something that the user wishes do (e.g., in scrolling menus or when browsing in a dialog box). Symmetry was translated by reversing the melody (opening or closing or scrolling menus up or down).

**Looping:** Small objects are associated with looping (repeated) sounds, so these objects can be found and manipulated. An exception is made when small objects are numerous and the use of sounds could distract the user from the written content (as in the FileManager where the user has to explore directories and documents).

**Stereo:** Stereo sounds are used to translate horizontal scrolling bar properties. Stereo sounds are not used to translate spatial position because perceiving subtle differences in stereo sounds is difficult and because the tablet could be used for this purpose. As for symbols, variations in pitch were also kept to a minimum as non-musicians have difficulty perceiving them. These sources are not really used for identification; they instead translate relative modifications: menus or windows scrolling up or down, longer or shorter menus (shorter scale), etc.

### Several Steps in the Evaluation Process

During our design phase, we carried out a four-step users' testing process involving a total of 39 blind persons using Windows 3.11. The results of these tests are very positive. After a two hour training period, we found that users could understand Windows and perform simple tasks at the desktop level or text editing by using the sound-only version of PC-Access (Dufresne, Martial, & Ramstein, 1995; Martial & Garon, 1996). The system seems promising as most blind users like both the sounds and the possibility of spatial exploration. These blind users seem to learn the system quickly, and they understand and use the mouse extensively to explore text and to manipulate menus. Though blind users are accustomed to learning sequences, it seems that PC-Access

allowed test subjects to understand the desktop metaphor and are then able to transfer this knowledge to a command mode of operation. As traditional non-visual keyboard-based interfaces require more memory with numerous commands and a complex syntax, blind users found PC-Access to be easier to remember and efficient to use after only a short training period.

We are currently testing the system for a period of three weeks. These tests involve asking blind persons to use the sound-only version of PC-Access for normal tasks in conjunction with voice synthesis and braille readers. We wish to evaluate how efficiently these blind users can access the GUI's elements, measure the effects of learning and fatigue, and evaluate the user's satisfaction and performance after an extended period. We will take the results of these experiments into account when designing the later earcons model.

### Conclusion

Our results show that PC-Access is an efficient teaching interface and a useful tool for blind users. PC-Access helps blind users build a mental image of the screen, to memorize the Windows syntax and symbolism, and to control interaction (Martial & Garon, 1996). Through this research project, we have learned several interesting things about multimodal interaction. For example, while some earcons are audible to certain Pantograph users, the significance of these sounds is only obvious when the same user is working with the drawing pad. These results open new avenues for research into multimodal interface design. As semantic associations and attitudes depend of the user's personality, we can also imagine studying the effects of customized audio styles (cartoons, romantic, etc.) in different contexts (learning, working, leisure, etc.) and for different categories of users (blind persons, children, etc.).

### References

Andrews, D. (Ed.). (1995). National center for the blind, screen review programs - comprehensive reviews of speech access programs for the blind. Baltimore, MD.

ANPEA, (1996). L'accès aux applications de windows à partir d'un terminal braille. Paris:ANPEA.

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. In Human-computer interaction (pp. 11-44). Hillsdale, NJ: Lawrence Erlbaum

Buxton, W. (1989). In Human-computer interaction (pp. 1-10). Hillsdale, NJ: Lawrence Erlbaum

Brewster, S. A., Wright, P. C., Dix, A. J., & Edwards, A. D. N. (1995). The sonic enhancement of graphical buttons. In Proc. of IFIP International Conference Interactions '95 Lillehammer, Norway.

Dufresne, A., Martial, O., & Ramstein, C. (1995). Multimodal user interface system for blind and "visually occupied" users: Ergonomic evaluation of the haptic and auditive dimensions. (pp. 163-168). In Proceedings of IFIP International Conference Interactions '95 Lillehammer, Norway.

Edwards, W. K., Mynatt, E. D., & Stockon, K. (1995). Access to graphical interfaces for blind users. (pp. 54-67). In Proceedings of IFIP International Conference Interactions '95 Lillehammer, Norway.

Familiant, M. E., & Detweiler, M. C. (1993). Iconic reference: evolving perspectives and an organizing framework. International Journal of Man-Machine Studies, 39, 705-728.

Gaver W. (1989). The SonicFinder: An interface that uses auditory icons. Journal of Human Computer Interaction, 4, 67-93.

Leinmann, E., & Schulze, H. H. (1995). Earcons and icons: An experimental study. In Proceedings of IFIP International Conference Interactions '95 (pp. 49-54). Lillehammer, Norway.

Marcus, A. (1991). Sign language: Graphics designers try to break the language barrier with icons. Unixworld Magazine, Special Report: International Unix, pp. 63-67.

Martial, O., & Dufresne, A. (1993). Audicon: Easy access to graphical user interfaces for blind persons-designing for and with people. In Human computer interaction (pp. 808-813). Orlando: Elsevier.

Martial, O., & Garon, S. (1996). How the visually impaired learn to work with windows, new technologies. In The education of the visually handicapped

John Libbey Eurotext.

Mynatt, E. D., & Weber, G. (1994). Nonvisual presentation of graphical user interfaces: Contrasting two approaches. In Proceedings of the ACM CHI'94 Conference on Human Factors in Computing Systems (pp. 166-172). Boston: ACMCHI.

Petrie, H., & Gill, J. (1993). Current research on access to graphical user interfaces for visually disabled computer user. European Journal of Special Needs Education, 8 (2), 153-156.

Proceedings of ASSET'96, ACM/SIGCAPH, 2nd Annual ACM Conference on Assistive Technologies (April 11-12, 1996). Vancouver, BC, Canada.

Ramstein, C. (1995). MUIS: Multimodal user interface system with force feedback and physical models. In Proceedings of IFIP International Conference Interactions '95 (pp. 157-162). Lillehammer, Norway.

Ramstein, C., Martial, O., Dufresne, A., Carignan, M., Chassé, P., & Mabilleau, P. (April 11-12,1996). Touching & hearing. GUIs: Design issues for the PC-Access system, ASSET '96, ACM/SIGCAPH. In 2nd Annual ACM Conference on Assistive Technologies (pp. 2-9). Vancouver, BC, Canada.

Vanderheiden, G. (summer 1991). Graphic user interface: A tough problem with a net gain for users who are blind. Presented at Technology and Disabilities '91.

Aude Dufresne, Odile Martial, Christophe Ramstein, and
Philippe Mabilleau
Centre for Information Technology Innovation
1575, Chomedey Blvd
Laval, Qc
H7V 2X2 Canada