

# The Use of Non-Speech Sounds in Non-Visual Interfaces to the MS-Windows GUI for Blind Computer Users

Helen Petrie and Sarah Morley

Sensory Disabilities Research Unit  
University of Hertfordshire  
Hatfield, Hertfordshire. AL10 9AB  
United Kingdom

## Abstract

Two studies investigated the use of non-speech sounds (auditory icons and earcons) in non-visual interfaces to MS-Windows for blind computer users. The first study presented sounds in isolation and blind and sighted participants rated them for their recognisability, and appropriateness of the mapping between the sound and the interface object/event. As a result, the sounds were revised and incorporated into the interfaces. The second study investigated the effects of the sounds on user performance and perceptions. Ten blind participants evaluated the interfaces, and task completion time was significantly shorter with the inclusion of sounds, although interesting effects on user perceptions were found.

keywords: blind computer users, graphical user interfaces, evaluation

## 1 Introduction

Non-speech sounds are still underused in computer interfaces, even though most current computer systems can easily generate a variety of sounds and numerous studies have shown that the addition of such sounds improves user performance [e.g. 2, 3, 6]. It is even more surprising that non-visual interfaces for blind and partially sighted computer users also underuse non-speech sounds, with some notable innovative exceptions [e.g. 9, 11]. Most interfaces for this user group use synthetic speech to convey both the contents of the application and the interface element. This is potentially confusing, a relatively slow method to communicate information, and given the current state of synthetic speech technology, not a particularly pleasant or clear method of receiving information for the users.

Appropriate methodologies are needed for the selection and evaluation of non-speech sounds for use in interfaces, be they for sighted or blind users. Mynatt [7,8] developed a technique for selecting sounds by asking people to listen to a range of sounds and attempt to identify the objects and actions involved. Her studies showed that sounds varied greatly in the accuracy with which they could be identified. The current study built on this methodology by asking both sighted and blind people to rate sounds both for their recognizability and for the appropriateness of their use to represent particular objects or events in the interface.

The GUIB Project (Graphical and Textural Interfaces for Blind People) [4] developed a number of interfaces for blind computer users to access the MS-Windows GUI. A multimodal approach was taken in the development of the interfaces, exploiting many aspects of the auditory and tactile routes for interaction. Three main interface configurations were developed: one based on synthetic speech output; one on refreshable Braille output; and one combining both synthetic speech and refreshable Braille. The use of non-speech sounds to increase the efficiency and pleasantness of the interface was explored extensively in the project. This paper will present two of the studies in which the use of non-speech sounds was evaluated: the first study evaluated potential users' perceptions of a selection of sounds in isolation; and the second study evaluated the effect of the use of a revised set of sounds on user performance in realistic use of the three interfaces.

## 2 Study 1: Non-Speech Sounds for a Non-Visual Interface: Recognisability and Mapping Judgements

An eclectic approach to the selection of sounds for possible inclusion in the interface was taken. Where possible *auditory icons* [5], sounds with some meaningful relationship to the interface object or event to which they corresponded, were used (e.g. a typewriter sound indicated an editfield was available for text entry, see sound 15 in Table 1). In cases where no appropriate auditory icon could be found, an earcon [1], usually a musical motif

was used (e.g. a harp motif indicated an application, see sound 6 in Table 1). Compound sounds were also used in the interface, so a spaceship door opening (sound 2) followed by a harp motif (sound 6) indicated an application opening. Sounds were also parameterised in a number of ways to add to their meaning (e.g. sounds were muffled to indicate background or inactive objects). The basic set of 21 sounds is shown in Table 1.

Code	Sound description	Referent object/event	Comments
1	tapping	icon	arbitrary sound
2	spaceship door opening	opening	auditory icon - if user recognises the StarTrek-type door sound
3	forklift truck lifting up an object	pick up (e.g. an icon)	auditory icon
4	forklift truck moving along with an object	drag	auditory icon
5	forklift truck putting down an object	drop	auditory icon
6	harp motif	application	earcon
7	boing	window border	auditory icon - if user understands the analogy of something bouncing off a border
8	spring	resizing window	auditory icon - if user understands the analogy of something springing in/out
9	submarine radar ping	edge of screen	arbitrary sound
10	doorbell	group	arbitrary sound (high pitch = open) (low pitch = closed)
11	heavy click	push button	auditory icon - although the use of different click sounds for the two kinds of buttons is arbitrary
12	short click	check box	auditory icon
13	multiple click	radio button	auditory icon
14	clock ticking	“please wait” icon	auditory icon
15	typewriter	editfield	auditory icon
16	paper riffle	listbox	auditory icon - if user understands the analogy of paper to indicate a list
17	people talking	dialog box	auditory icon - if user understands talking to indicate this type of “dialogue”
18	sitar motif	unknown object	earcon
19	scissors snipping	cut	auditory icon
20	piano chord	menu bar	earcon
21	piano scale	pull down menu	earcon

**Table 1: The 21 basic non-speech sounds used in the GUIB non-visual multimedia interface**

Eight blind and five sighted participants took part in the study. The blind participants were halfway through a course introducing them to computing with Windows, so had some familiarity with the kinds of objects and events to which the sounds referred. The sighted participants were all graduate students in occupational psychology and human-computer interaction who were regular users of Windows. The sighted participants were included in the study to investigate whether data from sighted computer users can reasonably be used in the design of interfaces for blind users, in a form of “empathic modelling” [10]. They were asked to imagine that they were blind and using an interface to Windows which gave them information in a combination of synthetic speech and sounds.

## 2.1 Method

Participants listened to each sound several times and were told what the sound was and its intended referent in the interface. They then rated the sound on a 7-point Likert scale (1 = very poor to 7 = excellent) on two judgements:

1. The Recognition Judgement: How recognisable and clear is the sound as an example of this class of sound (for example, if the sound is meant to be a clock ticking, is it immediately recognisable as the sound of a clock ticking)?
2. The Mapping Judgement: How clear is the use of this class of sound for the object or event in the interface which it is meant to indicate (for example, if a listbox is indicated by a paper riffle sound, how clear is the mapping from the idea of paper riffing to the idea of a listbox)?

After the ratings exercise was complete, sighted participants were asked to perform a second task which they had not been warned about (the Discrimination Task). They were given the name of an interface object or event (but no information about which sound represented this object/event). Then two sounds were played and participants decided which of the sounds corresponded to the object/event.

## 2.2 Results

On both Recognition and Mapping Judgements, sighted participants gave significantly higher ratings than blind participants ( $F = 18.76, p < 0.04$ ). There was also a significant interaction between participant group and rating ( $F = 6.08, p < 0.03$ ) with blind participants giving approximately equal mean ratings on both Judgements, but sighted participants giving much higher ratings on the Mapping Judgement than on the Recognition Judgement. The significant main effect and interaction both show that findings from sighted participants imagining themselves to be blind cannot be used as a substitute for data from participants who are actually blind.

For this reason, we will focus mainly on the results from the blind participants. In conducting the evaluation, it was decided to present only the first 13 sounds to the blind participants, as the reactions of the participants to the sounds were quite negative. For the same reason, the blind participants only undertook the Recognition Judgements. For each sound, the Recognition and Mapping Judgements ratings were assessed against a criterion that to be acceptable as a sound or as a mapping any sound ought to have a mean rating of greater than the mid-point of the overall scale (i.e. 4.00).

Only 7 of the 13 sounds (53.8%) passed the criterion for the Recognition Judgement. Only 6 of the 13 sounds presented (46.2%) passed the criterion for the Mapping Judgement, which is the more important criterion. It would be easy to find a new example of a harp sound if participants thought that the idea of using a harp to indicate applications was a good idea, but in fact, the mean rating for the harp-to-application mapping was only 3.25, so a new mapping needs to be found. Additional comments made by the participants were that the sounds were too long and too harsh, which would be intrusive when working with the computer.

Bearing in mind that the data from the Discrimination Task are for the sighted participants only, only 1 out of the 10 pairs of sounds presented posed any difficulty, this was the check box and radiobutton pair. These were two variations on a “click” sound, although the radiobutton was a longer, multiple click. Participants may have tried to memorise the sound as a click, which would then cause retrieval and matching problems when there were two different clicks amongst the sounds used. Although the random selection process used to produce the sounds for the Discrimination Task did not result in a pairing of the checkbox and push button sounds, these were two even more similar click sounds which may well have produced the same problem. During the Judgement phase of the study both blind and sighted participants also commented on how similar the three forklift truck sounds were, which might also cause similar kinds of discrimination problems.

In order to improve the sounds on both the Recognition and Mapping Judgements, various sounds were revised before inclusion in the GUIB interfaces: shorter sounds clips were made, softer examples were chosen and in some cases different instances of the sounds were used. Ideally, we would have repeated the above evaluation exercise with the revised set of sounds to empirically test whether their acceptability had indeed increased, but the need to develop full working prototypes for evaluation did not allow this.

### **3 Study 2: Effect of Non-Speech Sounds on User Performance and Perceptions in the GUIB Non-Visual Multimedia Interfaces**

#### **3.1 Method**

To evaluate the GUIB multimedia interface, an extensive user-based study was undertaken [results reported in 4]. Ten blind computer users took part, undertaking approximately 8 hours of training in the concepts underlying MS-Windows and the use of the different GUIB interfaces, and 2 hours of evaluation testing. Data collected was based on performance of realistic interface management and word processing tasks, ratings on 7-point Likert scales of the participants' perception of the usability, learnability and satisfaction with the interfaces and open-ended questions covering all aspects of the system.

Each participant evaluated two different interfaces which varied on a number of dimensions such as whether only synthetic speech or Braille output was available or both for example. One dimension was also whether the non-speech sounds were included. Unfortunately, given the large number of dimensions involved, a fully factorial design could not be implemented (in which all combinations of configurations were tested), but some comparisons on the effects of the sounds are possible.

#### **3.2 Results**

The total time taken to complete a set of word processing tasks showed a significant overall difference due to the inclusion of sounds ( $F = 5.11$ ,  $p = 0.05$ ) with shorter times achieved with the sounds present. The participants' ratings of the overall usability of the interfaces did not vary significantly due to the inclusion of sound. However, in rating more specific components of usability, there was a significant interaction between whether sounds were included and the specific interface being evaluated ( $F = 3.12$ ,  $p = 0.01$ ). For the synthetic speech interface, ratings of ease of learning, efficiency and acceptability were higher for the interface when the sounds were included (although, the rating for ease of use did not differ between these two conditions). For the Braille interface and the combined synthetic speech and Braille interfaces, there were no differences in these ratings due to the inclusion of the sounds.

Another interesting significant difference due to sound was found. Participants were also asked to rate the ease of performing different types of interactions with the interfaces; inputting commands, exploring the interface, and finding information in the interface. This produced a significant difference due to the inclusion of sound ( $F = 11.04$ ,  $p = 0.002$ ), but in this case, perhaps surprisingly, ratings were higher for the versions of the interfaces without sound than those with sound. This may simply be an anomaly due to the small number of participants used, but may also indicate that perceptions of interactions do not always match performance data, or that users do not always prefer the most efficient method of interaction.

### **4 Conclusions**

The first study demonstrated a methodology for preliminary evaluation of non-speech sounds before inclusion in an interface which was very simple to conduct, but included a number of dimensions of the appropriateness of the sounds and could also easily be extended to include other relevant dimensions. Ideally this methodology could be performed iteratively to optimise the sounds to be used. However, the study did suggest that judgements from sighted users imagining themselves to be blind cannot be used as a substitute for data from participants who are actually blind.

The second study showed that the inclusion of sounds improved the performance of tasks with the interface, although the differences in ratings were not as great as might have been expected. One reason for this might have been that the use of a mixture of auditory icons and earcons created a complex and somewhat inconsistent set of sounds. This mix was used to explore the effects of each type of sound (and will require further data analysis), but made the users' task of mapping between sound and interface object/event more complex. Nonetheless, some participants found the sounds to be more useful than they had expected, especially

once they had learned the mappings, and expressed views such as “they are a much quicker way of taking in information than the speech or Braille output - they’re not intuitive - but once you’ve learned what they mean they’re very useful for confirmation, information, and orientation” (participant 9).

## 5 References

1. Blattner, M.M., Sumikawa, D.A. and Greenberg, R.M. (1989). Earcons and icons: their structure and common design principles. *Human-Computer Interaction*, 4, 11 - 44.
2. Brewster, S.A., Wright, P.C. and Edwards, A.D.N. (1994). The design and evaluation of an auditory-enhanced scrollbar. *Proceedings of CHI '94 (Boston, MA)*. New York: ACM Press.
3. Brewster, S.A. and Crease, M.G. (1997). Making menus musical. *Human Computer Interaction: INTERACT '97. (IFP TC13 International Conference on Human Computer Interaction. Sydney, Australia)*. London: Chapman and Hall.
4. Emiliani, P.-L. (1995). Final report of the Textual and Graphical User Interfaces for Blind People Projects. London: Royal National Institute for the Blind.
5. Gaver, W.W. (1989). The SonicFinder: an interface that uses auditory icons. *Human-Computer Interaction*, 4, 67 - 94.
6. Gaver, W., Smith, R. and O’Shea, T. (1991). Effective sounds in complex systems: the ARKola simulation. *Proceedings of CHI '91 (New Orleans)*. New York: ACM Press.
7. Mynatt, E. (1994). Designing with auditory icons. *Proceedings of the Second International Conference on Auditory Display (ICAD '94)*.
8. Mynatt, E. (1997). Designing auditory cues for Mercator. *Human Computer Interaction: INTERACT '97. (IFP TC13 International Conference on Human Computer Interaction. Sydney, Australia)*. London: Chapman and Hall.
9. Mynatt, E. (1997). Transforming graphical interfaces into auditory interfaces for blind users. *Human Computer Interaction*, 12, 7 - 45.
10. Poulson, D., Ashby, M. and Richardson, S. (1996). *Userfit: a practical handbook on user-centred design for assistive technology*. Brussels-Luxembourg: European Commission.
11. Stevens, R.D., Edwards, A.D.N. and Harling, P.A. (1997). Access to mathematics for visually disabled students through multimodal interaction. *Human Computer Interaction*, 12, 47 - 92.