THE SONIC ENHANCEMENT OF GRAPHICAL BUTTONS

Stephen A. Brewster[1], Peter C. Wright[2], Alan J. Dix[3] and Alistair D. N. Edwards[2]

[1]VTT Information Technology,
Tekniikantie 4 B,
P.O. Box 1203,
FIN-02044 VTT, Finland
Tel.: +358 0 456 4311
sab@hemuli.tte.vtt.fi

[2]Department of Computer Science,
University of York,
Heslington,
York, Y01 5DD, UK.
Tel.: +44 904 432765
[pcw, alistair]@minster.york.ac.uk

[3]School of Computing and Mathematics,
University of Huddersfield
Queensgate, Huddersfield,
HD1 3DH, UK

alan@zeus.hud.ac.uk

KEYWORDS: Auditory interfaces, multi-modal interfaces, earcons, sonification, sonically-enhanced widgets, buttons

ABSTRACT: Graphical buttons are common to almost all interfaces but they are not without problems. One common difficulty is slipping off a button by mistake and not noticing. Sonically-enhanced buttons were designed to overcome this problem and were experimentally evaluated. Timing, error rates and workload measures were used. Error recovery was significantly faster and required fewer keystrokes with the sonically-enhanced buttons than with standard ones. The workload analyses showed participants significantly preferred the sonically-enhanced buttons to standard ones. This research indicates that by simple addition of sound one of the major problems with graphical buttons can be overcome.

INTRODUCTION

One of the most fundamental widgets in all graphical human-computer interfaces is the graphical button (to avoid confusion, *graphical button* will here be used to refer to the button on the computer display and *mouse button* will be used to refer to the button on the mouse). Although they are very common they are not without problems. Dix, Finlay, Abowd & Beale (1993) and Dix & Brewster (1994) have described some of the problems of graphical buttons. One of the main difficulties is that the user may think the graphical button has been pressed when it has not. This can happen because the user moves off the graphical button before the mouse button is released. This is caused by a problem with the feedback from the graphical button (see Figure 1). Both correct and incorrect presses start in the same way (1A and 2A). In the correct case, the user presses the graphical button and it becomes highlighted (1B), the mouse button is then released with the mouse still over the graphical button, it becomes un-highlighted (1C) and the operation requested takes place. The button slip-off starts in the same way. The user presses the mouse button over the graphical button (2B), then moves (or slips) off the graphical button and releases the mouse button (2C), the graphical button becomes un-highlighted (as before) but no action takes place. The feedback from these two different situations is *identical*. This problem occurs infrequently but, as the error may not be noticed for a considerable time, the effects can be serious. With a one-step undo facility users must notice before the next action takes place otherwise they may not easily be able to correct the mistake.

The identical feedback would not be a problem if the user was looking at the graphical button to see the slip-off, but this is not the case (Dix & Brewster, 1994). Dix & Brewster suggest there are three conditions necessary for such slip-off errors to occur:

i) The user reaches closure after the mouse button is depressed and the graphical button has highlighted.
ii) The visual focus of the next action is at some distance from the graphical button.
iii) The cursor is required at the new focus.

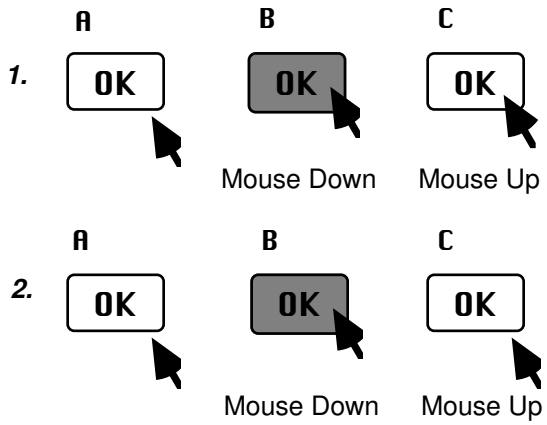

Figure 1: Feedback from pressing and releasing a graphical button. (1) shows a correct button selection, (2) shows a slip-off.

Closure occurs when the user perceives the task as being completed, which in this case is when the graphical button is highlighted (the mouse button is down). In reality, the task does not end until the mouse button is released. Because closure (i) is reached before this, the user starts the next task (mouse movement, iii) in parallel with the mouse up and a slip-off occurs. The user's attention is no longer at the graphical button (ii) so the feedback indicating a slip-off is not noticed. The problem occurs with expert users who perform many operations automatically and do not explicitly monitor the feedback from each interaction. This type of error is an *action slip* (Reason, 1990).

These problems occur in graphical buttons that allow a 'back-out' option: Where the user can move off the graphical button to stop the action. If the action is invoked when the mouse button is pressed down on the graphical button (instead of when it is released) then these problems do not occur as the user cannot slip off. These buttons are less common because they are more dangerous as users cannot change their minds.

Feedback is therefore needed to differentiate a successful and unsuccessful click and to indicate when a slip-off has occurred. In this paper we suggest auditory-feedback should be used for this. Why use sound to present any new information? A graphical method could be used instead. The drawback with this is that it puts a greater load on the visual sense. Furthermore, sound has certain advantages. For example, it can be heard from all around, it does not disrupt the user's visual attention and it can alert the user to changes very effectively. The problems due to closure discussed above could not easily be solved by adding more graphical feedback. The user is no longer looking at the button's location so any feedback given there will be missed. Feedback could be given at the mouse location but we cannot be sure the user will be looking there either. Sound is omni-directional and the user does not need to focus attention on any part of the screen to perceive it. It is for these reasons that we suggest sound should be used to enhance the graphical user interface.

The work reported here is part of a project looking at the best ways to integrate auditory and graphical information at the interface (Brewster, 1994). An earlier part of this work evaluated a sonically-enhanced scrollbar (Brewster, Wright & Edwards, 1994). The results showed that usability could be improved by the addition of sound. The experiment described here uses the same method of evaluation.

EXPERIMENT

An experiment was designed to test sonically-enhanced buttons to see if they would improve usability. An initial experimental design was piloted on five participants but failed to cause any mis-hit errors (Dix & Brewster, 1994). The experiment was then redesigned paying greater attention to the reasons why slip-off errors occurred. This is the experiment reported here.

Participants

Twelve participants were used. They were postgraduate students from the Department of Computer Science at the University of York. All had more than three years experience of graphical interfaces and buttons. Expert subjects were used because the type of error studied here is an action slip.

Task

Figure 2 shows a screen shot of the interface to the task. Participants were required to enter five digit codes via the on-screen keypad. The codes were randomly generated and displayed in the 'Code to type' field. The participants had to enter the code using the mouse and on-screen keypad. To enter a code participants had to press a number and then press the 'OK' button to accept it, then press the next number and so on. The numbers entered appeared in the 'Code' field above the keypad. When the code had been typed the 'Next' button was used to display the next one. This maximised the number of button presses and mouse movements that the participants

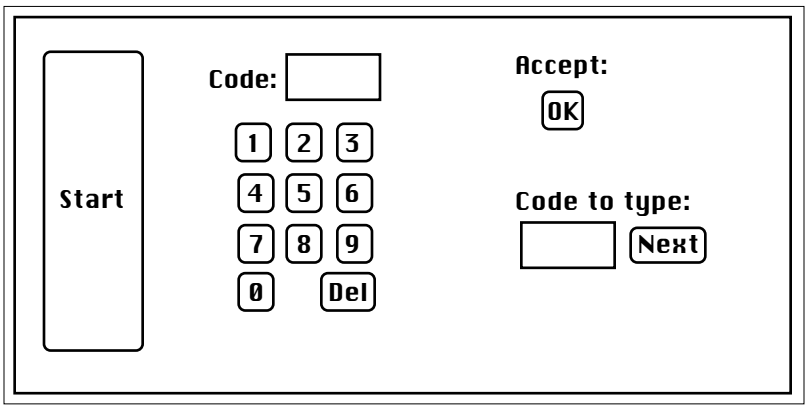

Figure 2: The button testing program (reduced in size).

had to make. In the visual condition the buttons acted like normal Macintosh buttons. In the auditory condition there was no visual highlighting (the buttons stayed white) but the buttons made the sounds described below.

The task was designed to be simple so that the participants could easily learn it and reach a level of automaticity in the task where slip-off errors would occur.

Sounds used

The sounds used were based around structured audio messages called *Earcons* (Blattner, Sumikawa & Greenberg, 1989 and Brewster, Wright & Edwards, 1993). Earcons are abstract, synthetic tones that can be used in structured combinations to create sound messages to represent parts of an interface. The sounds were created using the earcon guidelines proposed by Brewster (1994).

Two sounds were needed. One to be the auditory equivalent of the graphical highlight when the mouse button was pressed down on the graphical button. The other to indicate when a button was pressed correctly or when a slip-off occurred.

An electronic organ timbre was used for all the sounds. This was shown to be effective when sonifying a scrollbar (Brewster *et al.*, 1994). When the mouse button was pressed down over a graphical button a continuous sound at pitch C_3 (261Hz) was played. This continued for as long as the mouse button was down and the mouse was over the graphical button. If the mouse was moved off the graphical button the sound stopped. If the mouse was released over the graphical button then a success sound was played. This consisted of two notes, played consecutively, at C_1 (1046Hz) each with a duration of 2/60 sec. This success sound had to be kept short so that participants did not get confused as to which button the feedback was coming from: The audio feedback had to be able to keep pace with interactions taking place. To make sure that the number of sounds was kept to a minimum and speed maximised, if a participant quickly clicked the mouse over a graphical button the mouse down sound was not played; only the success sound. The mouse button down and success sounds differentiated a successful and unsuccessful mouse click.

The two sounds used a combination of pitch, duration and intensity to get the listener's attention. This meant that a lower level of intensity could be used, making the sounds less annoying for the primary user and others working nearby. Annoyance is most often caused by excessive intensity (Berglund, Preis & Rankin, 1990). It is important to note that intensity is not the only way to get the user's attention. The human perceptual system is good at detecting dynamic stimuli. As Edworthy, Loxley, Geelhoed & Dennis (1989) showed, attention-grabbing sounds can be created by varying other sound parameters.

All the sounds used were played on a Roland D110 multi-timbral sound synthesiser. The sounds were controlled by an Apple Macintosh via MIDI through a Yamaha DMP 11 digital mixer and presented to subjects by loudspeakers.

Experimental design and procedure

The experiment was in two halves and was a repeated-measures within-subjects design (see Figure 3). The order of presentation was counterbalanced to avoid learning effects. Training was given before each of the conditions so that participants could get used to the method of entering data. Each condition lasted 15 minutes and the participants had to type in as many codes as possible.

In order to get a full range of quantitative and qualitative results time, error rates and workload measures were used (Bevan & Macleod, 1994). Time and error rate reductions would show quantitative improvements and workload reductions would show qualitative improvements. The total number of codes typed and slip-off errors by each participant was recorded.

The NASA Human Performance Research Group (1987) break workload down into six different factors: Mental demand, physical demand, time pressure,

Participants	Condition 1	Condition 2
Six Participants	Auditory Buttons Train & Test	Visual Buttons Train & Test
Six Participants	Visual Buttons Train & Test	Auditory Buttons Train & Test

Figure 3: Format of the experiment.

effort expended, performance level achieved and frustration experienced. NASA have developed a measurement tool, the NASA-Task Load Index (TLX) for estimating these subjective factors. We used this but added a seventh factor: Annoyance. One of the main concerns of potential users of auditory interfaces is annoyance due to sound pollution. This is often given as a reason for not using sound at the human-computer interface. In the experiment described here the annoyance due to auditory feedback was measured to find out if it was indeed a problem. In addition to these seven factors we also asked our subjects to indicate, overall, which of the two interfaces they felt made the task easiest. Subjects had to fill in workload charts after both conditions of the experiment.

Experimental hypotheses

The extra feedback provided by the sounds should make it easier for participants to recover from errors. They will notice that the errors have occurred more quickly than in the visual condition. This should result in faster error recovery times in the auditory condition. More codes should be typed in the fifteen minutes due to less time being spent on error recovery.

The workload felt by participants should be reduced as the extra feedback would provide information that the participants needed. Participants should have to expend less effort recovering from errors. Physical demand and time pressure should be unaffected as they are unchanged across conditions. There should be no increased frustration or annoyance due to the addition of sound as the auditory feedback will provide information that the participants need.

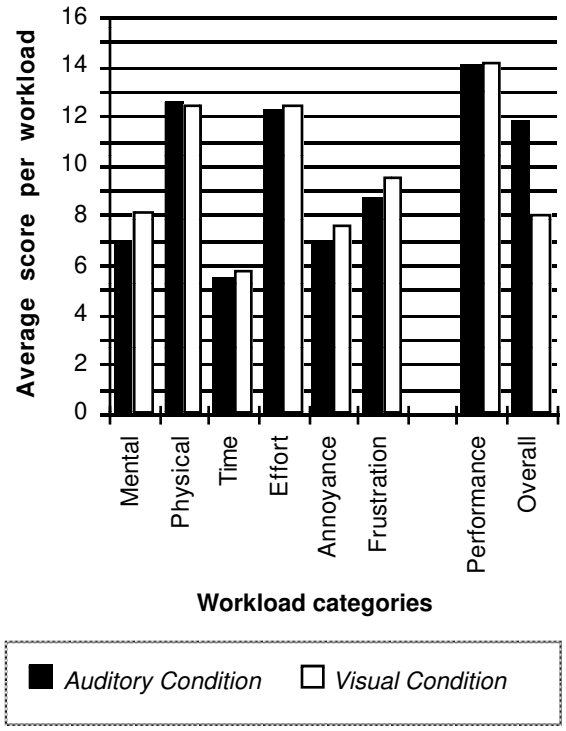

Figure 4: Average TLX workload scores for the auditory and visual conditions of the experiment. In the first six categories higher scores mean higher workload. The final two categories, performance and overall, are separated because higher scores mean less workload.

RESULTS

TLX Results

Figure 4 shows the average score for each of the workload categories. They were scored in the range 0-20. Paired T-tests were carried out on the auditory versus visual conditions for each of the workload categories. An analysis of the individual scores showed that none were significantly different between conditions. However, the sonically-enhanced buttons were given a significantly higher overall preference rating ($T(11)=5.14$, $p=0.0003$). Here, participants were asked to rate which type of button made the task the easiest. This strongly significant results seems to indicate that the participants found the task easier with the sonically-enhanced buttons but this did not affect the workload required for the task.

There was no significant difference in terms of annoyance. Four participants rated the auditory condition more annoying than the visual but five

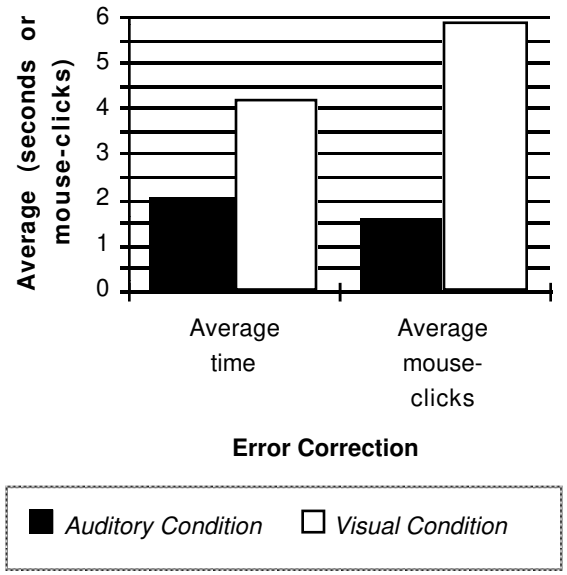

Figure 5: Error recovery in the buttons experiment. The graph shows average error recovery times and the average number of mouse clicks needed for recovery.

participants rated the visual more annoying than the auditory. This indicated that the participants did not find the sound feedback annoying to use.

Timing and Error Results

Figure 5 shows the results of error recovery. The time to recover from each slip-off was calculated. This was taken as the time from when the slip-off occurred until the user pressed the mouse button down on the correct graphical button again. It was found that participants in the auditory condition recovered from slip-off errors significantly faster than in the visual condition ($T(12)=3.51$, $p=0.004$). Average recovery times ranged from 2.00 seconds in the auditory condition to 4.21 seconds in the visual condition. The number of mouse button downs and button ups taken to recover from slip off errors was also significantly reduced in the auditory condition ($T(12)=4.40$, $p=0.0008$). The average number of clicks to recovery was 1.5 in the auditory condition and 5.89 in the visual. In the auditory condition the participants recognised an error had occurred and often fixed it by the next mouse button down. In the visual condition it took nearly six button ups and downs before the participants recovered from an error. These results confirmed the hypothesis that sound can help users recover from slip-off errors more quickly.

The auditory condition had an average of 6.6 slip-off errors per participant and the visual condition 3 per participant. There was no significant difference between these scores ($T(11)=2.03$, $p=0.067$). There was no significant difference in the total number of codes typed in the two conditions ($T(11)=0.401$, $p=0.696$). The average number of codes typed per participant in the auditory group was 64.5 and in the visual 65.5.

DISCUSSION

The workload analysis showed that there were no significant differences between the conditions on any of the factors. This showed that the sonic enhancements did not reduce the workload of the task. However, the participants very strongly preferred the sonically-enhanced buttons to the standard ones. This may have been because the auditory buttons allowed participants to recover from errors with less effort. It is unclear why this was not reflected in the workload scores. It may be that recovering from errors was seen as a separate activity from the main task and therefore did not figure in workload estimates but might have affected preference ratings

The sonically-enhanced buttons did not increase the annoyance or frustration felt by the participants. This, along with results from sonifying a scrollbar (Brewster *et al.*, 1994), gives strong evidence to suggest that if sounds provide useful information they will not be perceived as annoying to the primary user of the computer. The sounds used here were kept at a low intensity so that other users nearby would not be annoyed by them.

The main hypothesis, that the addition of sound would speed up error recovery, was proved correct by the experiment. Time to recover from errors and the number of keystrokes needed were both significantly reduced. These results indicate that, if the simple sound-enhancements suggested here are used, slip-off problems can be dealt with very effectively.

There was no significant difference between the total number of codes typed in either condition. Even though the participants in the auditory condition recovered from errors more rapidly they did not type more codes. The participants in the auditory condition made more slip-off errors than in the visual (although this difference was not significant). The auditory condition made, on average, 6.6 slip-off errors per participant and each of these took, on average, 2 seconds to recover from, making 13.2 seconds spent

on error recovery. In the visual condition there were, on average, 3 slip-off errors per participant taking 4.2 seconds to recover from, making 12.6 seconds spent on error recovery. These times were not significantly different indicating why there was no difference in the number of codes typed. The fact that the participants made more errors in the auditory condition wiped-out the advantage gained in recovering from errors more quickly. It was as if participants became more careless with their clicking because they knew that they could recover from errors with little cost. However, recall that the difference in total errors was not significant so we cannot make any strong conclusions about the number of errors that would be observed in a real interface. It is hoped that when sonically-enhanced buttons are used in real interfaces users will not make more errors but that they will make the same number of errors and recover more quickly.

Another important result from the experiment was that visual feedback was removed and replaced with more effective auditory feedback. The auditory feedback was not displayed redundantly with the graphical, it replaced it. This is important because it shows that sound can be used to present information that is currently graphical. The high overall preference for the sonically-enhanced button shows that participants did not miss the graphical feedback and preferred the auditory.

FUTURE WORK

Buttons are used in many areas of the interface, from icons to menus. The advantages demonstrated here could be used in these other widgets to overcome their similar problems.

CONCLUSIONS

Sonically-enhanced buttons were shown to be effective at reducing the time taken to recover from slip-off errors. The number of keystrokes necessary to recover was also reduced. The sonically-enhanced buttons were strongly preferred by the participants over standard visual ones. These results suggest that the introduction of such buttons into human-computer interfaces would improve usability but not at a cost of making the interface more annoying to the user.

The results showed that sound could be used to replace visual feedback. In the auditory condition there was no graphical feedback to indicate that the button had been pressed: It was done purely in sound. This proved to be effective and users preferred it. This leads the way to removing other graphical feedback and replacing it with more effective auditory feedback, leaving the visual sense free to concentrate on the main task the user is trying to accomplish.

ACKNOWLEDGEMENTS

Thanks go to Jon Watte for his help with Macintosh programming. This work was supported by DSS funding.

REFERENCES

Berglund, B., Preis, A., & Rankin, K. (1990). Relationship between loudness and annoyance for ten community sounds. Environment International, 16, 523-531.

Bevan, N., & Macleod, M. (1994). Usability measurement in context. International Journal of Man-Machine Studies, 13(1 & 2), 123-145.

Blattner, M., Sumikawa, D., & Greenberg, R. (1989). Earcons and icons: Their structure and common design principles. Human Computer Interaction, 4(1), 11-44.

Brewster, S. A. (1994) Providing a structured method for integrating non-speech audio into human-computer interfaces. PhD Thesis, University of York.

Brewster, S. A., Wright, P. C., & Edwards, A. D. N. (1993). An evaluation of earcons for use in auditory human-computer interfaces. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel, & T. White (Eds.), Proceedings of INTERCHI'93, (pp. 222-227). Amsterdam: ACM Press, Addison-Wesley.

Brewster, S. A., Wright, P. C., & Edwards, A. D. N. (1994). The design and evaluation of an auditory-enhanced scrollbar. In B. Adelson, S. Dumais, & J. Olson (Eds.), Proceedings of CHI'94, (pp. 173-179). Boston, Massachusetts: ACM Press, Addison-Wesley.

Dix, A., Finlay, J., Abowd, G., & Beale, R. (1993). Chapter 9.4 Status/Event Analysis. In Human-Computer Interaction (pp. 325-334). London: Prentice-Hall.

Dix, A. J., & Brewster, S. A. (1994). Causing trouble with buttons. In Ancillary Proceedings of HCI'94, . Sterling, UK: Cambridge University Press.

Edworthy, J., Loxley, S., Geelhoed, E., & Dennis, I. (1989). The perceived urgency of auditory warnings. Proceedings of the Institute of Acoustics, 11(5), 73-80.

NASA Human Performance Research Group (1987). Task Load Index (NASA-TLX) v1.0 computerised version. NASA Ames Research Centre.

Reason, J. (1990). Human Error. Cambridge, UK: Cambridge University Press.