



UNIVERSITY
of
GLASGOW

Brewster, S.A. (2002) Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing* 6(3):pp. 188-205.

<http://eprints.gla.ac.uk/3230/>

Overcoming the Lack of Screen Space on Mobile Computers

Stephen Brewster

Glasgow Interactive Systems Group
Department of Computing Science
University of Glasgow
Glasgow, G12 8QQ, UK
Tel: +44 (0)141 330 4966 Fax: +44 (0) 141 330 4913
Email: stephen@dcs.gla.ac.uk Web: www.dcs.gla.ac.uk/~stephen

Overcoming the Lack of Screen Space on Mobile Computers

Stephen Brewster

Glasgow Interactive Systems Group
Department of Computing Science
University of Glasgow
Glasgow, G12 8QQ, UK
Tel: +44 (0)141 330 4966 Fax: +44 (0) 141 330 4913
Email: stephen@dcs.gla.ac.uk Web: www.dcs.gla.ac.uk/~stephen

ABSTRACT

One difficulty for interface design on mobile computers is lack of screen space caused by their small size. This paper describes a small pilot study and two formal experiments that investigate the usability of sonically-enhanced buttons of different sizes. The underlying hypothesis being that presenting information about the buttons in sound would increase their usability and allow their size to be reduced. An experimental interface was created that ran on a 3Com Palm III mobile computer and used a simple calculator-style interface to enter data. The buttons of the calculator were changed in size between 4x4, 8x8 and 16x16 pixels and used a range of different types of sound from basic to complex. Results showed that sounds significantly improved usability for both standard and small button sizes – more data could be entered with sonically-enhanced buttons and subjective workload reduced. More sophisticated sounds that presented more information about the state of the buttons were shown to be more effective than the standard Palm III sounds. The results showed that if sound was added to buttons then they could be reduced in size from 16x16 to 8x8 pixels without much loss in quantitative performance. This reduction in size, however, caused a significant increase in subjective workload. Results also showed that when a mobile device was used in a more realistic situation (whilst walking outside) usability was significantly reduced (with increased workload and less data entered) than when used in a usability laboratory. These studies show that sound can be beneficial for usability and that care must be taken to do testing in realistic environments to get a good measure of mobile device usability.

Keywords

Earcons, mobile computers, PDAs, non-speech sounds, buttons, sonically-enhanced widgets, evaluation.

1. INTRODUCTION

Mobile computing devices are becoming extremely popular. Mobile telephones, Personal Digital Assistants (PDAs) and handheld computers are one of the fastest growth areas of computing. One problem with these devices is that there is a very limited amount of screen space on which to display information: the screen cannot physically be made bigger as the devices must be able to fit into the hand or pocket to be easily carried. Small screens can easily become cluttered with information and widgets (buttons, menus, windows, etc.) and this presents a difficult challenge for interface designers.

The graphical techniques for designing interfaces on desktop systems do not apply well to handheld devices. Screen resources are limited; often screens are black and white to reduce cost and power consumption. Memory and processing power are much reduced from desktop systems. However, in many cases widgets have been taken straight from standard desktop graphical interfaces (where screen space and other resources are not a problem) and applied directly to mobile devices. This has resulted in devices that are hard to use, with small text that is hard to read, cramped graphics and little contextual information. For example, early versions of Microsoft Windows CE used many features from Windows 95. As Zuberec [1] says:

“Windows CE designers have begun to realise that the device form factors and the tasks that the users need to accomplish should dictate the degree of the desktop behaviour retained”.

It is therefore clear that taking the desktop interface and implementing it on a mobile device does not work well; other methods must be investigated to make mobile interfaces more usable. One possibility to solve the problems with widgets could be to substitute non-speech sound for visual feedback so that smaller versions could be used. This would mean that the clutter on the display could be reduced and/or allow more information to be presented on the display. This must be done in a way that maintains usability otherwise these smaller widgets will render the device unusable. This paper presents initial work to investigate if sounds can be used in this way.

Another reason for using sound is that if users are performing tasks whilst walking or driving, they cannot devote all of their visual attention to the mobile device. Visual attentional resources must remain with the main task for safety. It is therefore hard to design a visual interface that can work under these circumstances. An alternative, sonically-enhanced interface would require less visual attention and therefore potentially interfere less in the main activity in which the user is engaged.

Many mobile computers can make simple sounds already – a selling feature of many mobile telephones is the range of ringing tones that they can make. Apart from the ringing tones, the audio capabilities are little used - many mobile telephones beep to confirm a button press and some PDAs do the same to confirm hitting buttons with a stylus. Sounds could be used for much more than this to improve the whole interaction with the device and the information it contains. This paper proposes some simple uses of sound that could significantly improve usability.

2. PREVIOUS WORK

There has been little previous research directly in the area of sound in mobile computing devices. Some of the research that has been done on different selection techniques using pen and a touch tablets applies to mobile interfaces. For example, Ren *et al.* [2-4] looked at six different selection strategies to see which was most effective (see).

	Strategy	Details
	Land-on 1	Target selected when stylus hits it
	Land-on 2	Similar to Land-on 1, but this time the stylus must start outside the target and then move into it

Take-off 1	Target is highlighted when stylus is touching it, selection is made when pen is taken off the target
Take-off 2	Similar to Take-off 1, but the target is selected when the stylus is removed from any point on the screen (either inside or outside the target area)
Space 1 Standard buttons no sound Small buttons with sound Small buttons no sound Small buttons with sound Small buttons no sound	Pen approaches from above, target highlights when stylus is within 1 cm above the target. Selection occurs as soon as the stylus lands on the target

Space 2	Similar to Space 1, except that the selection is made when the pen lands anywhere on the screen (either inside or outside the target area)
---------	--

Table 1 Table 1: Selection strategies from Ren *et al.* [3]. Table 1

In Ren *et al.*'s experiments, participants had to select individual targets that appeared in different locations on screen as fast as possible. Their results showed that the Land-on 2 strategy gave the best balance between speed of selection and error rate. This strategy has some problems in real situations as, if there were many targets close together, one might inadvertently select the wrong one by moving over it on the way to the target required.

Not all of these strategies are currently implemented on mobile devices (for example, the space strategies relied on an electromagnetic tablet which could sense the stylus when it was above the surface – no current mobile computers work in this way). Many mobile

devices implement just the Land-on 1 or Take-off 1 strategies (in the work described below the Take-off 1 strategy was used as the 3Com Palm III only provides this technique).

Ren *et al.* did not consider the effect of sound on selection strategies. Sound can improve targeting and therefore could affect the usability of the different selection strategies. Sound could help with highlighting when the stylus is on or off a target or when nearing a target (such as in the Space strategies). The work below will investigate the effect of sound in the Take-off 1 strategy to see its usefulness.

One problem when selecting items with a stylus is that there is no tactile feedback. With keys on a normal keyboard you can feel that you have pressed them, with a stylus or finger on a touch screen it is hard to know if you have hit the target item or not. Much previous work has shown that touch screens can be made more usable (users type faster and more accurately) if sounds are added to indicate key clicks (Lewis *et al.* [5] review this area). This suggests that sounds could be used to improve the performance of the stylus driven interfaces used by many mobile computers. This work did not address all of the information that could be presented in sound about a button – only the confirmation of key clicks. The work presented here will build upon this but present more information in sound as to the state of the button and whether the press was successful or not.

Some work has gone into the investigation of the size of targets on touch screens. Hall *et al.* [6] showed that for seated users to reach greater than 99% target selection accuracy required targets of 26mm^2 and for standing users 30mm^2 . This size was heavily affected by the size of users fingers; using a stylus (which has a much smaller tip than a finger) to select targets allows much smaller sized targets. One interesting thing to note is that users needed larger targets when standing than when sitting – it was more difficult to hit small targets when standing. This was probably because it was hard for users to rest and steady their arms when standing, but this is actually more realistic in many respects (for example, when pressing on the screen of an information kiosk or on the display of a mobile device). The work presented below investigates much smaller buttons than those investigated by Hall *et al.* – the size of buttons on the real display of a current mobile device were used.

2.1 Non-speech sounds in desktop interfaces

Much work has demonstrated that sound in desktop interfaces can be very successful. Beaudouin-Lafon and Conversey [7] showed that non-speech sounds could improve usability of scrollbars, Maury *et al.* [8] added sounds to improve menu selection times in drop-down menus.

Research at Glasgow has demonstrated the effectiveness of sound in a range of desktop widgets [9] but not so far in mobile devices. One successful sonically-enhanced widget was the on-screen (or graphical) button. Although buttons are very common they are not without problems [10-12]. One of the main difficulties is that the user may think the 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 (either on purpose – maybe the user changed his/her mind about pressing the button, or by accident – the user wanted to press it but slipped-off the button by mistake). This can also happen with a stylus interface.

The problem is caused by the feedback from the button (see Figure). Both correct and incorrect presses start in the same way (1A and 2A). Casting this in terms of a mobile device, in the correct case the user hits

the button with the stylus and it becomes highlighted (1B), the stylus is lifted off the button, it becomes un-highlighted (1C) and the operation requested takes place. The slip-off starts in the same way. The user hits the button with the stylus (2B), then moves (or slips) off the button and lifts off the stylus (2C), the button becomes un-highlighted (as before) but no action takes place. The feedback from these two different situations is *identical*. Correcting the problem (moving back to the button and re-pressing it) takes time and causes frustration. 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.

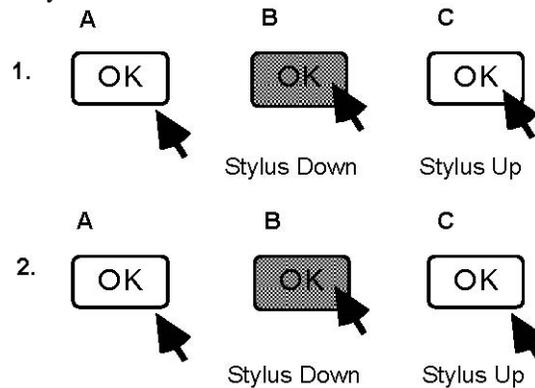


Figure 1: Feedback from pressing and releasing a graphical button with a stylus. (1) shows a correct button selection, (2) shows a slip-off.

To overcome the problems with buttons Brewster *et al.* added non-speech sounds [9]. A sound was played when the user moved the mouse over a button – this was a constant, quiet organ sound to indicate when the user was over a valid target (in this case a button). This was played for as long as the mouse was over the button. The next sound was played when the user pressed the mouse button down when on a graphical button – this was a louder, higher pitched sound. This was played for as long as the mouse button was down over the graphical button. The final sound was played when the user released the mouse button when over the graphical button – this was a brief, high-pitched ‘correct selection’ sound. No sound was played when the user slipped-off a button – the lack of a confirmation sound indicated the error.

Results for the desktop buttons showed a significant overall preference for the sonically-enhanced buttons over the standard silent ones. Users did not find the buttons annoying to use (the sounds were providing information users needed to operate them more effectively and were not just noise). Participants recovered from the slip-off errors significantly faster in the audio condition, the number of buttons clicks to recover from errors was significantly reduced (there was no difference in the number of errors made as the sounds only indicated when an error had occurred and allowed the user to fix it, they did not stop errors from occurring in the first place). There was no significant difference in the amount of data entered with the audio or visual buttons (one reason for this is that the visual feedback was removed from the buttons to see if it could be replaced by sound. This worked, but slowed down the interactions with the buttons. In the studies described below the visual feedback from the buttons was enhanced with sound – the visual feedback was still available).

It is suggested here that the advantages of sonically-enhanced buttons in the desktop could apply to mobile devices and therefore improve usability. The problems with button interactions are made worse by small buttons and, because their screens are small, these occur often on mobile computers. The tip of the stylus (or finger) can obscure the button underneath it. This makes it hard for users to know if they are on the button or not. The additional vibration cause by walking or other movement when using the device makes it even harder to keep on target. The visual feedback highlighting the button is the problem. More sophisticated visual feedback could be designed but this might be difficult given the limited resolution, screen space and available colours on many mobiles. If users are walking or moving then they will also not be able to look at the screen all of the time: they must look where they are going, therefore there is a chance that they might miss the feedback, however saliently it is presented visually.

3. GENERAL EXPERIMENTAL DESIGN

The pilot study and the two experiments on mobile device interfaces presented below used a similar testing strategy. They all used a within-groups design with a standard silent graphical interface compared against a sonically-enhanced one (in which the graphics were the same as the standard interface). To get a balanced view of the usability of the interfaces being tested a full range of quantitative and qualitative measures was taken [13]; the amount of data entered with the interface would show quantitative improvements and workload differences would show qualitative differences.

Hart and Wickens ([14], p 258) define workload “...as the effort invested by the human operator into task performance; workload arises from the interaction between a particular and task and the performer”. NASA [15] break workload down into six different factors: Mental demand, physical demand, time pressure, effort expended, performance level achieved and frustration experienced. The NASA-Task Load Index (TLX) was used for estimating these subjective workload factors in the studies here.

The basic six factors were used as described but a seventh factor was added: Annoyance. This is often cited as a reason for not using sound in interfaces as it is argued that continued presentation of sound would be an annoyance for the user. So, by adding this as a specific factor in the usability assessment it would be possible to find out if participants felt that sonic feedback was an annoyance. Participants were also asked to indicate overall preference: which of the two interfaces they felt made the task easiest. Participants had to fill in paper-based workload charts for all of the factors after each condition and treatment in each experiment.

3.1 Evaluation of mobile interfaces

It is important to test usability of any interface improvements to a mobile device outside the laboratory. The approach taken here was to initially test in the lab so that the fine-grained effects could be controlled and studied. The next step was to test outside in a much more realistic situation with users trying to use interfaces whilst performing other realistic tasks such as walking. It is important that mobile devices are not tested only in the lab as many potential problems may be missed. As Johnson [16] suggests:

“... the conventional usability laboratory would not be able to adequately simulate such important aspects as the weather and could not easily provide for the wide range of competing activities and demands on users that might arise in a natural setting”.

The approach to evaluation taken here was inspired by work from Petrie *et al.* [17] on interfaces to mobile devices for blind people. They have begun to investigate different evaluation techniques that give some measure of usability in realistic, real-world situations.

4. PILOT STUDY

A pilot study was needed to see what effect sound would have on the usability of a PDA display. For the experiment a 3Com Palm III was chosen because this is the biggest selling palmtop computer of 1998-1999. On-screen buttons were chosen as these are one of the most common widgets found in the displays of mobile devices and previous results showed that sonic-enhancement could provide usability benefits. Two different button sizes would be compared to see if the smaller ones with sound would be made as effective as the larger ones without sound. The sounds used would be based on those already in the Palm III and then enhanced ones that were shown to be beneficial in previous experiments [18].

4.1 Hypotheses

The main hypotheses for the pilot study were that the sounds would improve the usability of buttons of both sizes. This breaks down as follows:

- The amount of data entered with the sonically-enhanced buttons would be increased for both button sizes. The smaller buttons would benefit more as they would be harder to use. This would be demonstrated by the small buttons with sound becoming as effective as the large buttons without sound.
- The basic sounds would be less effective than the enhanced sounds. The analysis of button actions performed in previous work [18] suggested that the basic sounds provided by the Palm III would not be enough to solve usability problems. The enhanced sounds would solve these problems and therefore

- allow more data to be entered and workload to be reduced.
- The sounds would reduce the workload required to operate the buttons as the extra feedback provided would help participants use the buttons more effectively. This would be demonstrated by reduced TLX workload scores.
- The sounds would not be annoying to the participant – they would provide information to allow participants to use the buttons better; they would not be just noise but information.
- User preference would be for the buttons with sounds for the same reason – the sounds would help them use the buttons more effectively.

4.2 Design and procedure

The pilot study used a two-condition, within-groups design (see Table 2). There were two button conditions: standard size (16x16 pixels) and small (4x4 pixels). 16x16 pixel buttons are common on the Palm III and so were used for the control condition. 4x4 buttons were as small as it was possible to go in the widget builder used to construct the software for the study. This really made them the worst-case situation but one in which adding sound could have a large benefit as the visual presentation was so limited. There are few buttons of this size on most mobile devices but they are possible and the study would show their usability.

There were three 7-minute treatments in each condition: silent, basic sound and enhanced sound. The conditions and treatments were not fully counterbalanced for this pilot study – the aim of the study was to quickly investigate the general effects of sound. Conditions and treatments were carried out in the order shown in . This could cause learning effects but would give general information on the effectiveness of sound that could be followed-up by more formal studies. Seven minutes of training was given before each participant started the first condition. This ensured that all participants were able to perform the task effectively. The task was designed to be very simple to learn. The experiment was conducted in a usability lab with participants carrying out the tasks on their own sat at a desk.

Condition: Button size / Treatment: Sound type

16 x 16 pixel	Silent (C1.1)	Basic sound (C1.2)	Enhanced sound (C1.3)
4 x 4 pixel	Silent (C2.1)	Basic sound (C2.2)	Enhanced sound (C2.3)

Table 2: Format of the pilot study.

The independent variables were button size (two levels: small and standard) and sound type (three levels: silent, basic and enhanced sound). The dependent variables were number of codes typed and the NASA TLX workload scores.

4.3 Participants

Twelve participants were used. All were postgraduate students in the Department of Computing Science at Glasgow. Nine were male and three female with ages ranging from 22 to 35. They all had experience of graphical user interfaces and buttons, but not of the Palm III interface. Only twelve were used as this was to be an initial pilot investigation of the effects of sound. If this demonstrated interesting results then bigger, more formally controlled studies could be undertaken to investigate the effects in detail.

4.4 Task

The experiment was run on a 3Com Palm III handheld computer with input via a stylus. The task the participants had to perform was similar to that of Brewster *et al.* [10]. This task and interface had proved effective in the previous research and so was used again. Participants had to enter a series of five digit codes (shown in the target windows in Figure 2) using the numeric keypad. These codes were randomly generated and the same set of codes was used by each participant in each condition (the same set of codes was used for each of the studies presented here). After each digit had been tapped the OK button had to be

pressed to confirm it. The numbers entered appeared in the window labelled 'Current'. This maximised the number of button taps and stylus movements that the participants had to make. In the visual condition the buttons acted like normal buttons. In the auditory condition the buttons made the sounds described below. Training was given before the experiment began and TLX workload charts were completed after each treatment.

7

The standard sized buttons (16x16 pixels) became highlighted (in reverse video) when pressed and un-highlighted when not pressed. They were approximately 5mm² in size. The small buttons did not highlight – with only 4 x 4 pixels (1.25mm²) there were no pixels available for highlighting.

4.5 Sounds used

Simple *Earcons* [19, 20] were used for the audio feedback. Earcons are abstract, synthetic tones that can be used in structured combinations to create sound messages to represent parts of an interface. Earcons are constructed from motives. These are short rhythmic sequences that can be combined in different ways. The simplest method of combination is concatenation to produce *compound earcons*. By using more complex manipulations of the parameters of sound *hierarchical earcons* can be created [21] which allow the representation of hierarchical structures. Detailed investigations of earcons by Brewster, Wright & Edwards [19] showed that they are an effective means of communicating information in sound. The sounds were designed using earcon construction guidelines proposed by Brewster *et al.* [22], given the restrictions imposed by the Palm III. The Palm III is only able to present very simple sounds – it has a simple tone generator that can only change frequency and duration. This restricted the sophistication of the sounds that could be used in the experiment. In previous studies of sonically-enhanced widgets, sounds were produced from a MIDI synthesiser and were of much higher quality (mobile devices that can play these kinds of sounds will be available soon).

For the basic sounds the key-click provided as standard on the Palm III was used. This indicated that the pen was released on a button and that it was selected correctly (example 1C in Figure 1). The enhanced sounds used the basic ones but added a sound for pen down on a button. This indicated that users had hit the target (especially important for small targets when the pen tip can obscure the target that you are trying to hit, or when you are walking so that the device is not still and the vibration makes it hard to get (and stay) on target) and was a higher pitched version of the basic sound. Finally, a sound was played if the user slipped off a button [10] (example 2C in Figure 1). This was a lower pitched version of the basic sound (the lower pitched sound had a more 'negative' feel).

Figure 2: Screen shots of the small and standard sized buttons used in the pilot study.

These sounds were very similar to those used in previous studies [9] except that a sound was used to indicate a slip-off. In the desktop buttons experiment this was indicated by a lack of sound. Here a sound was added to make it very clear to the user when this had happened. In a mobile device the user might be doing other things whilst using the device, for example walking, and so would need a more salient cue to indicate that a slip-off had occurred.

4.6 Results from the pilot study

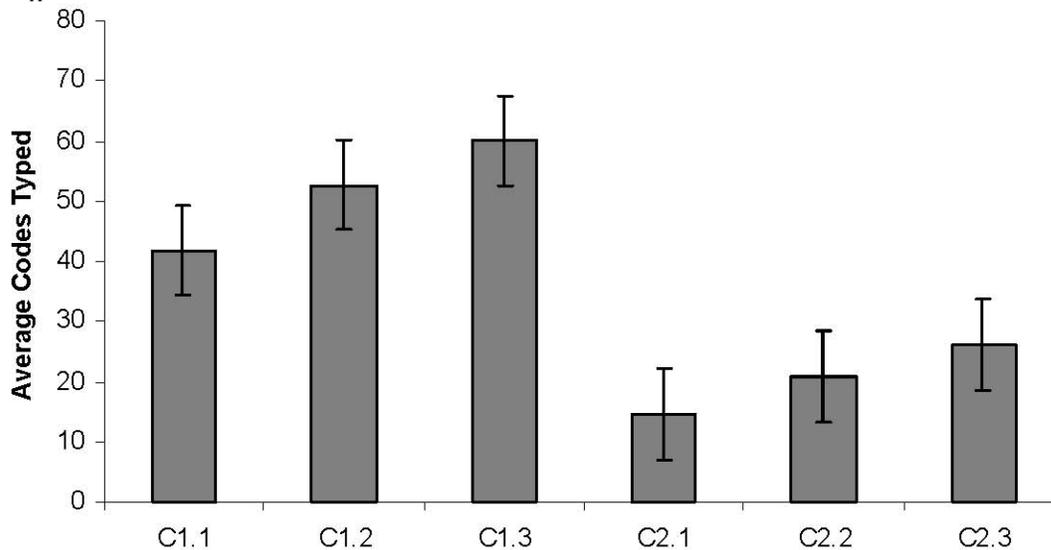
One-factor ANOVA tests were conducted across the three sound treatments for each button size. Workload categories were scored out of 20. For the standard sized buttons there was a significant reduction in overall workload ($F_{2,15}=15.44$, $p=0.0002$) across the three treatments (see Figure 3) confirming the hypothesis. Overall workload is the average of the standard six TLX factors (mental demand, physical demand, time pressure, effort expended, frustration experienced and performance level achieved). Tukey HSD tests showed that both audio treatments significantly lowered workload over the silent buttons (but no difference was shown between the two audio treatments except for mental demand. Here there was a significant reduction between the silent and the enhanced treatment only - $F_{2,33}=3.42$, $p=0.004$, $Q_{33}=2.58$, $p=0.05$).

There was a significant overall preference for both of the audio treatments and there was no significant effect on annoyance due to the sounds.

Figure 4 shows the results of the workload tests for the small buttons. There was again a significant decrease in overall workload across the three treatments ($F_{2,33}=11.11, p=0.0001$) confirming the hypothesis. Tukey HSD tests again showed that both audio treatments significantly reduced workload. There was also a significant reduction in each of the workload categories (apart from physical demand and time pressure) for the audio treatments. In all cases, the enhanced sound was equal to or better than the basic sound. For example, mental workload was significantly reduced with the enhanced sound only ($F_{2,33}=5.98, p=0.006, Q_{33}=4.25, p=0.05$). Performance was significantly higher in the enhanced sound treatment than in either the basic or silent ones ($F_{2,33}=16.5, p=0.00006$, enhanced vs silent: $Q_{33}=3.2, p=0.05$, enhanced vs basic: $Q_{33}=1.54, p=0.05$). Participants again preferred the sonically-enhanced buttons.

Figure 3: Mean workload results for the standard sized buttons in the pilot study. For the first six categories higher scores mean higher workload, for the final two categories (performance and overall preference) higher scores mean lower workload. Standard error bars are shown.

The numbers of codes typed in each condition can be seen in Figure 5. For the standard buttons there was a significant increase in the number of codes entered between both of the audio treatments and the silent one ($F_{2,33}=15.41, p=0.00002$), confirming the hypothesis. No significant differences were found between the audio treatments. For the small buttons there was again an overall increase in the number of codes typed in both of the audio treatments over the silent ($F_{2,33}=13.53, p=0.00005$), confirming the hypothesis. A T-test indicated that the enhanced sound significantly increased the number of codes typed over the basic sound ($T_{11}=5.12, p=0.0003$).



The standard buttons with no sound (C1.1) allowed significantly more codes to be typed than the small buttons with enhanced sound (C2.3) ($T_{11}=6.26, p=0.00006$), failing to confirm the hypothesis. Participants typed on average 42 codes with the standard size silent buttons and only 26 with the small buttons with sound.

Condition/Treatment

Figure 4: Mean workload results for the small buttons in the pilot study. For the first six categories higher scores mean higher workload, for the final two categories (performance and

overall preference) higher scores mean lower workload. Standard error bars are shown.

Figure 5: Mean number of codes typed per condition/treatment in the pilot study. Standard error bars are shown.

4.7 Discussion

The results of this pilot study generally indicate that enhancing the buttons on a handheld computer such as the Palm III with sound could significantly increase their usability. This was only a pilot study and its aim was to initially explore the possibilities of sound and therefore was not fully controlled, but some broad conclusions can be drawn. The sounds were, in general, useful with the enhanced sounds the same or better than the basic ones. In particular, the buttons with sound allowed more data to be entered than those without, and the small buttons with enhanced sound significantly more than the silent or basic sound versions.

The small buttons with enhanced sounds were not found to be as usable as the standard buttons without sound in the data entry task here. The problem was that the small buttons were just too small. The stylus tip almost completely obscured the buttons when pressing them and the buttons did not highlight as there were not enough pixels to put into reverse video. These buttons are perhaps too small to be used reliably in a data entry application such as the one used on here, and whatever was done to them would not make them effective. That the small buttons were very difficult to use is vividly demonstrated in the workload data in Figure 3 and Figure 4. The average overall workload scores of the three treatments in each condition (taking an average of the overall workload for each button size) were very significantly higher for the small buttons (small buttons workload = 12.05, standard buttons = 5.6, $T_5=13.33$, $p=0.00004$). This suggested that buttons of this size are really not practical in the type of interface tested here. This would be even worse if participants were trying to operate the interface on the move and were therefore not able to hold the device steady. Button size is an important factor in the design of the display of a mobile device – if a designer makes them too small this could have a major impact on usability.

This time the experiment used a more formal, fully counterbalanced, two-condition, within-groups design. The independent variables were again button size and sound type. There were two conditions: standard (16x16 pixels) and small (8x8 pixels) buttons (see Figure 6). Both of these sizes are commonly found on PDA's. There were two 7-minute treatments in each button size condition: visual only buttons and visual plus sound. The experiment was designed to be as similar as possible to the pilot study so that the results could be compared.

The dependent variables were subjective workload (using the standard NASA TLX workload tests) and number of codes entered. Training was given before each part of the experiment began and NASA TLX workload scales were completed after each treatment. The experiment was carried out in a usability lab as before.

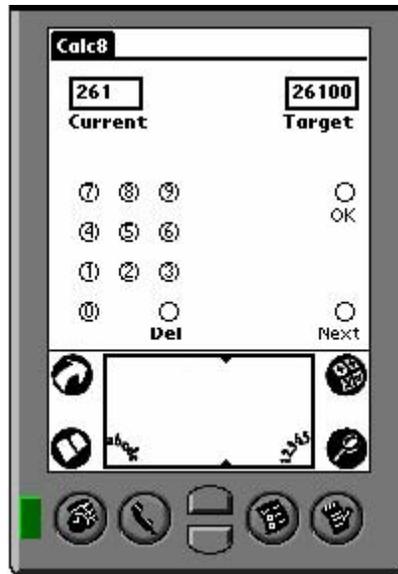


Figure 6: Screenshot from the small button (8 x 8 pixels) condition in Experiment 1. The standard sized buttons were the same as the pilot study (see the right hand side of Figure 2).

5.3 Participants

Sixteen participants from the Computing Science Department at Glasgow University were used. The selection was made up of postgraduate and undergraduate students and staff, nine male and seven female, all between the ages of 21 and 45. They all had experience of graphical user interfaces and buttons but not the Palm III interface. None of the participants had been used in the pilot study.

5.4 Task

The experiment was run on the same 3Com Palm III handheld computer as before. The task the participants had to perform was similar to the pilot study except that there was no basic sound treatment. Participants had to enter the same series of five digit codes (shown in the target window in Figure 6) using the numeric keypad.

The quantitative results can be seen in Figure 8. This shows the number of 5-digit strings that were entered in each of the treatments. The results show that in both of the sound treatments performance was very significantly improved (standard buttons: $T_{15}=9.22$, $p<0.001$, small buttons: $T_{15}=7.79$, $p<0.001$), confirming the hypothesis.

The standard sized silent buttons again allowed significantly more data to be entered than the small sonically-enhance ones ($T_{15}=7.79$, $p<0.001$), again disproving the hypothesis in this area. This time, however, the actual difference in the number of codes typed was only 3, with 44 entered with standard buttons and 41 with the small ones.

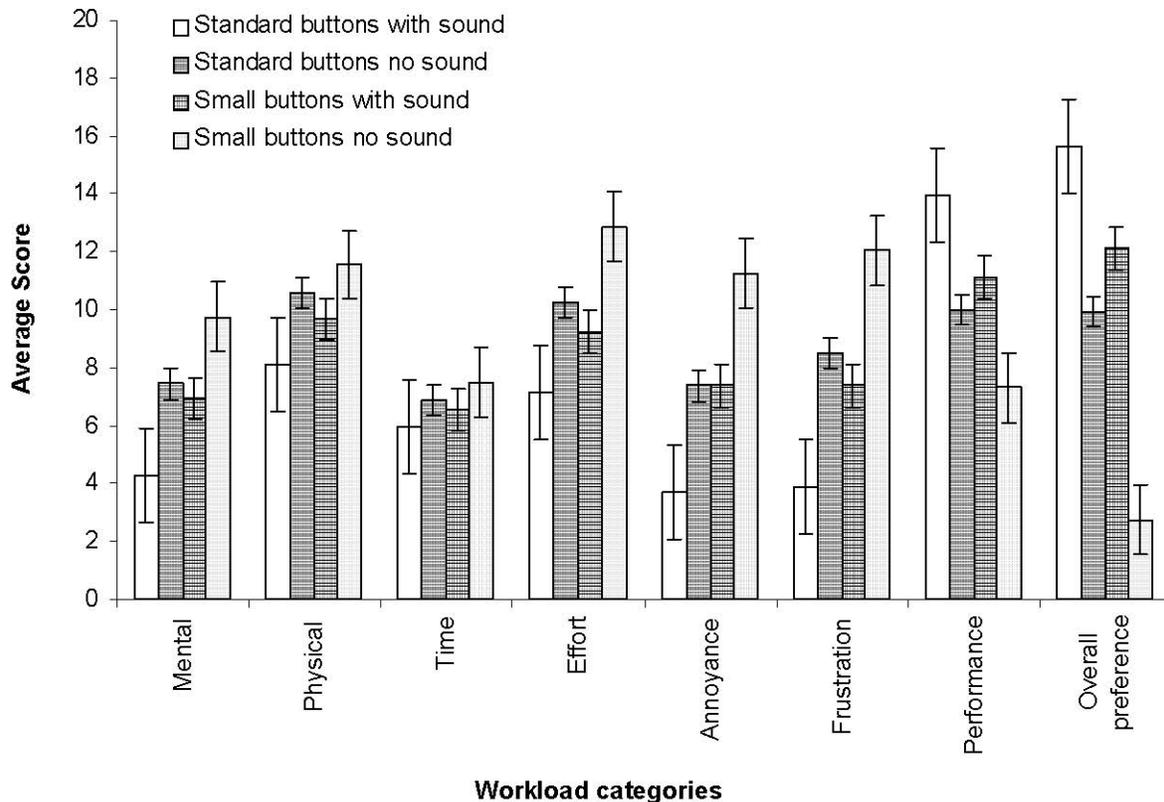


Figure 7: Mean workload scores for Experiment 1. For the first six categories higher scores mean higher workload, for the final two categories (performance and overall preference) higher scores mean lower workload. Standard error bars are shown.

5.6 Discussion

The results from the experiment show that sound can have important effects on usability in mobile computing devices. The results broadly confirmed those of the pilot study but now under more formal experimental conditions. The qualitative results showed that sound had a big effect on workload for both button sizes in the experimental task. In almost all of the categories workload was significantly reduced when sound was present. The results also showed that the sounds did not annoy the participants, in fact they rated the annoyance as significantly less when sounds were present. They also rated the sounds as preferable to the silent buttons. This indicated that sound can improve the qualitative experience users have with mobile devices.

With the small buttons no difference in subjective physical effort between the sound and no sound conditions was detected by the experiment (as in the pilot study). With the standard buttons sound reduced the physical effort required for the task. Any reduced effect with the small buttons may perhaps be explained by the fact that they were harder to press with the pen *because* they were small and addition of sound did not help. The effect of size out-weighed the effect of sound. This would suggest that it would be very difficult to make these buttons easier to use as usability problems are dominated by size of target (as would be suggested by studies into Fitts Law, for example).

The quantitative results also back up the hypothesis that sounds improve usability as participants were able to enter significantly more codes when sounds were present for both button sizes in the experimental task. The sounds helped users target the buttons better; know when they had been pressed correctly and when they had been mis-hit. This made them significantly easier to operate. A difference between the number of codes entered with standard sized silent buttons and small sonically-enhanced ones was still found but it had been reduced to just 3 codes over the 14 minutes of the two trials. Depending on the situation in which a mobile device might be used, this difference could be acceptable given the amount of screen space that

would be released for more information.

Figure 7: Mean workload scores for Experiment 1. For the first six categories higher scores mean higher workload, for the final two categories (performance and overall preference) higher scores mean lower workload. Standard error bars are shown.

These results give interface designers a simple way to increase the usability of buttons (one of the most common widgets) of different sizes in mobile devices in the data entry type of task studied here. They also show that using small buttons can reduce the usability of an interface. Designers must take care over the buttons (and other targets) that users must select – if they have to be small then sounds could be used to raise usability levels.

6. EXPERIMENT 2

This final experiment built on the results from Experiment 1. Improvements had been shown by the use of sound in buttons for the experimental task, but this had only been demonstrated in the usability laboratory with the user sitting in silence using the device. This was a very controlled environment – it was quiet, participants did not move around. These are key differences to the use of mobile devices in the real world. People use them whilst on the move and in noisy environments. An experiment was needed to see if the sonically-enhanced buttons would have the same benefits when used in a more realistic setting. To do this some of the control gained from testing in a laboratory had to be given up, but the realism and generality of the results would be greatly enhanced. The participants performed the same task as in the previous experiment but had to walk laps up and down an indicated outdoor route at the same time.

6.1 Hypotheses

In general, performance would be expected to be worse in all areas as participants had to walk at the same time as entering data, whereas in Experiment 1 they only had to enter data so could concentrate all of their efforts on this. The main hypotheses in this experiment were:

- That the use of sound should improve performance in real-world, outdoor use of a mobile computer. This would be demonstrated by an increased number of codes entered when sound was present for both of the button sizes.
- The reduced subjective workload due to the sounds found in Experiment 1 should cause the distance walked by each participant to increase. Participants would not have to concentrate on the screen of the Palm III as hard because some of the information needed was presented in sound and therefore they would be able to look where they were going – allowing them to complete more laps.
- Subjective workload should be decreased in the audio treatments for both button sizes, as in Experiment 1.
- Participants should prefer the buttons with sound and not find them annoying to use, again as in Experiment 1.

Figure 8: Mean number of codes entered in Experiment 1. Standard error bars are shown.

6.2 Participants

The sixteen participants used were again from the Department of Computing Science. Eleven of these had used some form of mobile device before (mostly mobile telephones or handheld games). None had used the Palm III before and none had taken part in either of the previous two studies.

6.3 Task

The task used in this experiment was exactly the same as in Experiment 1. Exactly the same software and hardware were used and exactly the same sounds were played. The only difference was that the study took place outside and the participants had to walk up and down a special route whilst performing the task. This

allowed the direct comparison of the results obtained here with those from Experiment 1. The route was reasonably quiet and shady but was not controlled as part of the experiment, thus simulating a more realistic usage scenario.



Figure 9: Looking each way down the street where Experiment 2 was conducted.

The participants had to walk up and down a pathway in the University of Glasgow (see Figure 9). The route was 10m long and straight. The route to be walked was clearly marked. Once a participant reached the end he/she turned and walked back to the start, continuing to do laps and entering codes until the time ran out.

6.4 Design

The design was exactly the same as in Experiment 1 and the order of presentation was fully counterbalanced in the same way. The same measures were taken, but, in addition, the number of complete laps walked by each participant was also recorded by the experimenter.

	Standard buttons with sound	Standard buttons no sound	Small buttons with sound	Small buttons no sound
Experiment 1 (Indoor)	5.9 / 1.6	8.9 / 1.5	8.1 / 1.3	11.0 / 2.0
Experiment 2 (Outdoor)	7.7 / 1.1	9.2 / 1.4	11.0 / 1.5	12.8 / 2.1

6.5 Results

The results of the NASA TLX workload tests for the standard and small buttons can be seen in Figure 10. Paired T-tests were used for the analysis. For the standard sized buttons sound significantly reduced workload in terms of mental demand ($T_{15}=2.99$, $p=0.009$), physical demand ($T_{15}=2.16$, $p=0.047$), effort

expended ($T_{15}=5.37$, $p=0.00007$), annoyance experienced ($T_{15}=2.62$, $p=0.019$) and performance level achieved ($T_{15}=2.1$, $p=0.05$). Participants significantly preferred the buttons with sound to those without ($T_{15}=6.83$, $p=0.0002$).

For the small buttons sound again had a positive effect. Sound significantly reduced workload for mental demand ($T_{15}=3.94$, $p=0.0001$), physical demand ($T_{15}=4.92$, $p=0.0001$) and effort expended ($T_{15}=2.68$, $p=0.017$). Participants significantly preferred the buttons with sound to those without ($T_{15}=5.46$, $p=0.0009$).

There was again a significant difference between the workload scores for the small and standard buttons (see). The average overall workload for the standard buttons was 8.5 and for the small buttons was 11.9 ($T_5=7.36$, $p=0.0007$).

	Standard buttons with sound	Standard buttons no sound	Small buttons with sound	Small buttons no sound
Experiment 1 (Indoor)	5.9 / 1.6	8.9 / 1.5	8.1 / 1.3	11.0 / 2.0
Experiment 2 (Outdoor)	7.7 / 1.1	9.2 / 1.4	11.0 / 1.5	12.8 / 2.1

Table 3: Mean overall workload score / Standard Deviation for Experiment 1 and Experiment 2.

Results for the numbers of codes typed with the different buttons sizes can be seen in Figure 11. There were significantly more codes entered with the sonically-enhanced buttons for both button sizes (standard buttons: $T_{15}=4.94$, $p=0.0002$, small buttons: $T_{15}=4.35$, $p=0.0006$). In this experiment no significant difference was detected between the small buttons with sound and the standard silent buttons ($T_{15}=2.08$, $p=0.055$). On average 28 codes were entered with the small buttons with sound and 32 with the large silent buttons (see Ta).

Table 4: Mean number of codes typed / Standard Deviation for Experiment 1 and Experiment 2.

Figure 10: Mean workload scores for Experiment 2. For the first six categories higher scores mean higher workload, for the final two categories (performance and overall preference) higher scores mean lower workload. Standard error bars are shown.

Experiment 2 (Outdoor)	42 / 9.5	32 / 7.3	28 / 6.2	18 / 6.7
	Standard buttons with sound	Standard buttons no sound	Small buttons with sound	Small buttons no sound

Experiment 1 (Indoor)	58 / 7.2	45 / 8.9	42 / 6.6	30 / 8.4
Experiment 2 (Outdoor)	42 / 9.5	32 / 7.3	28 / 6.2	18 / 6.7

The number of laps completed can be seen in Figure 12. Significantly more laps were walked using the standard sized buttons with sound than without ($T_{15}=2.67$, $p=0.018$). There was no significant difference in the number of laps walked between the small buttons with sound and those without ($T_{15}=1.10$, $p=0.287$). There was no difference shown between the small buttons with sound and the standard buttons without ($T_{15}=0.88$, $p=0.388$). In fact, the differences between the numbers of laps walked was small, only varying between 11 (standard buttons with sound) and 9 (small buttons without).

6.6 Comparing inside in the laboratory to walking outside

As we now have data on the same experiment conducted sat at a desk in a usability lab and walking outdoors it is possible to see the effect that this had. From the data on the number of codes entered (see Table and Figures 8 and 11) it can be seen that being outdoors and walking had a negative effect. To assess this a two-factor ANOVA was carried out on the indoor/outdoor data and the button size/sound treatment data. The results showed a significant main effect for indoors versus outdoors ($F_{1,120}=96.48$, $p<0.001$), a significant main effect for button size/audio treatment ($F_{3,120}=62.96$, $p<0.001$) but no interaction between the two ($F_{3,120}=0.31$, $p=0.81$).

This result strongly shows that being outside and walking reduced the amount of data that could be entered with the buttons in the application used, knocking an average of 32% off the numbers of codes entered (see Table 3). The data analysis in the section above has already described the effect for button size and audio treatment on the number of codes typed. The lack of interaction shows that walking outdoors did not significantly affect the smaller buttons more than the bigger ones, nor the sonically-enhanced ones more than the silent ones.

For the subjective workload data walking outside increased the workload by an average of 17% when using the data entry application (see Table 3). Again, walking and trying to perform the data entry task outdoors was much more difficult than indoors. This is not surprising but it is important to be clear what effect a more realistic situation has. This experiment has demonstrated that using a data-entry type of application on a mobile device in a more realistic environment can significantly change the interaction and this must be taken into account when designing and testing mobile devices. Testing in the laboratory on its own is not enough. Further research is needed to develop evaluation techniques that can test devices in more realistic situations.

Figure 11: Mean number of codes entered in Experiment 2. Standard error bars are shown.

6.7 Discussion

The results for Experiment 2 are generally in-line with those of Experiment 1. Again sound had a positive impact on the quantitative and qualitative aspects of usability for both button sizes. The small buttons were again the hardest to use. In this experiment there was no measured difference in the number of codes entered between the standard buttons with no sound and the small buttons with sound (on average 28 codes were entered with the small buttons with sound and 32 with the large silent buttons). This suggests that the small buttons with sound may have reached a similar level of performance to the large silent buttons for data entry. A further study would be needed to investigate this in more detail but it begins to show that the use of sound can offset a reduction in widget size. However, the workload scores were still significantly

higher for the small buttons, so this result must be used with care. It does indicate that sound can improve the performance of small buttons when on the move, but users have to invest more effort to use the small buttons at this higher level.

Why did the participants not walk further when using the small buttons with sound? The number of laps walked was not a very sensitive measure (as mentioned above, there was only a difference of two laps between the best and worst condition) perhaps because the length of each lap was long compared to the time taken for the experimental tasks. It proved to be useful in the more real world effects it had on workload and performance but did not cause much differentiation between participants. Further investigation of this as a measure must be undertaken to understand it more. One other possibility is that the participants had to balance two tasks – walking and data entry. They may have invested more effort in walking (and so walked at almost the same rate as with the standard buttons) and consequently less in the data entry (and so recorded lower scores than in Experiment 1). This is realistic as, if users were using their PDA's whilst crossing roads, avoiding cars, etc. in everyday life, they must, for safety, concentrate on the walking and not on their PDA.

7. GENERAL DISCUSSION

The results of the studies here show that, if sound is added, the size of buttons in a mobile text entry type application on a handheld computer such as the 3Com Palm can be reduced from 5mm² to 2.5mm² and the amount of data entered with them will not be greatly affected. This will, however, be at the cost of increasing subjective workload. Moving to 4x4 pixels (1.25mm²) makes buttons very hard to operate in this type of task and the use of sounds cannot offset this (although this may be acceptable under some circumstances where it is important to get as much information on the screen as possible and so lower usability is acceptable). It is clear that sound should be used in all cases of buttons on PDA's as it always increased performance. Designers in general should avoid very small buttons (or targets) as they will significantly reduce usability, but if they must be used then they should have sounds.

Figure 12: Mean number of laps completed in each condition in Experiment 2. Standard error bars are shown.

It seems that to maintain the same level of quantitative performance as a 5mm² button, one cannot go much smaller than 2.5mm² on a Palm. This size seems to be around the borderline – in Experiment 1 a significant difference in the amount of data entered was found and in Experiment 2 it was not. We know that going to 1.5mm² had a very definite negative impact. It is therefore suggested here that, for the type of interaction examined in this paper, buttons should be a minimum of 2.5mm² and no smaller. This research confirms the hypothesis that sound can be used to reduce the size of items in the interface of a mobile device, but care must be taken as this will increase the workload required to operate it. The Palm's touchscreen and stylus type of interaction is common across many handheld devices. If technology changes then this might well have an impact on the usability of buttons. More precise pointing devices might allow smaller buttons to be usable. This work provides a benchmark against which any new design can be compared.

It is hoped that the results here can be generalised from just buttons to other widgets. In most widgets the user has to click to activate (menus, scrollbars, tool palettes, etc) and so must target the widget and then stay on it whilst interacting with it. The sounds described here could help users know when they are on a scrollbar thumb wheel, when they are dragging it (and when they slip off by moving too far to either side of the scrollbar and out of its 'hot spot' [9]), for example. Sound added to a range of desktop widgets caused a general increase in usability [9] and the results on buttons here would suggest that the same could be true for other mobile widgets. We are now developing interfaces that use a full set of sonically-enhanced widgets to assess usability [24].

The results on annoyance here only looked at a single user of the system and not others nearby for whom the sounds might be annoying to overhear. The sounds are currently quite quiet so do not travel far (the

intensity level of sound being one of the major causes of annoyance [23]). This is not a full solution to the problem. However, the solution may already be in place – many people use hands-free kits with their mobile telephones (especially as more of these are incorporating MP3 players). This allows the sounds to be presented without annoying others nearby by using an ear-piece. Many people also use personal stereos – again the sounds are delivered over headphones so reducing the amount of annoyance to others. Many new mobile devices are also including this type of functionality and so will be used with headphones (for example, Microsoft's Pocket PC operating system for handheld devices includes a copy of Windows Media Player for playing audio files).

The walking task from Experiment 2 gave some idea of what could be expected to happen to usability when a mobile device such as the Palm was used in a more realistic situation. The walking task was still quite controlled – users walked along a reasonably quiet, straight path. This is less realistic than the tasks that users might perform in their everyday lives – crossing roads, negotiating busy, noisy streets. This is likely to have an even greater impact on performance than the walking task reported here. This means that it is vitally important that mobile interfaces are tested in real situations to assess usability. Further development of realistic evaluation techniques for mobile devices is needed so that designers can be sure the devices and interfaces they create are usable.

8. FUTURE WORK

The next step in this research is to look at presenting much more complex, dynamic information in sound. Many mobile devices display only static data – address books, calendars, etc. There is much potential for presenting more dynamic data using sound, for example stock market information or monitoring remote information and devices by telemetry [25]. Users would then be able to get at this from wherever they were.

Presenting dynamic information on devices with small displays is difficult. Techniques for large displays again do not transfer well to small ones. For example, on a large graphical display a strip-chart or line graph might be used so that users can see a change in one or more variables over time. This is very effective as the user can see general trends, compare one graph against another and also find current values. Utilising this type of presentation on a small display device is difficult as the screen is too small to provide much history so that trends are hard to follow. If the user has the device to his/her ear then no visual display is possible at all.

One way to present this type of information is using sound. Research has shown that presenting graphs in sound (for example mapping the Y axis value to pitch and X to time – this then allows you to hear the graph as a variation in pitch over time) for blind people allows them to extract much of the information needed from the graph [26] – maximum and minimum points, rates of change, etc. We are working on solutions based on non-speech sounds to present the trend information with speech to present current values when required. This will allow mobile devices to expand beyond their current limitations of static information to become devices that can present useful dynamic data.

An important future experiment would be to run a Fitt's law study of the effects of sound on target selection time. Sound is likely to change performance, for example the sound played when the stylus is on the button should aid targetting. The effects of sound could be quantified by such a study, as there is currently no information on this.

9. CONCLUSIONS

The pilot study and experiments described here have looked in detail at the use of sound in data entry type applications on mobile computer interfaces with different button sizes. Lack of screen space is a major interface design problem for mobile computers and if sound can improve performance then it is one way of overcoming this. The results have shown that adding sound to buttons on a 3Com Palm III can, in general, significantly improve usability by decreasing workload and increasing the amount of data that can be entered.

The results showed that if sound is added to buttons then they can be reduced in size from 5mm² to 2.5mm²

without much loss in quantitative performance when used for data entry tasks. This reduction in size, however, has a significant effect on the qualitative aspects of the interaction by increasing subjective workload. Going to buttons less than 2.5mm² is likely to cause a fall in quantitative performance for data entry as well as a further significant increase in workload. As a general design guideline buttons should not be reduced to below 2.5mm² and sound must be added to offset the size reduction.

If small buttons are to be used in the display of a mobile device then sounds could be included to raise levels of usability. If sounds are not used then performance will be significantly impacted. This is an important but simple design guideline that mobile device designers can use to improve usability and quality of their systems.

10. ACKNOWLEDGEMENTS

This work was part funded by EPSRC grant GR/L79212. Thanks go to Peter Cryer, Martin Crumley and Paul Blakey who all helped in this work as part of their undergraduate and masters projects in the Department of Computing Science. Thanks to the reviewers for excellent comments on the paper.

11. REFERENCES

References by Brewster can be found at <http://www.dcs.gla.ac.uk/~stephen>

1. Zuberec, S. The design of Microsoft Windows CE. In Proceedings of Human-Computer Interaction International'99 (Munich, Germany) Lawrence Erlbaum Associates, 1999, pp. 1148-1152.
2. Ren, X. and Moriya, S. A state transition model representing pen-based selection strategies. In Proceedings of IFIP INTERACT '99 (Edinburgh, UK) British Computer Society, 1999, pp. 57-58.
3. Ren, X. and Moriya, S. The best among six strategies for selecting a minute target and the determination of the minute maximum size of the targets on a pen-based computer. In Proceedings of IFIP Interact'97 (Sydney, Australia) Chapman & Hall, 1997, pp. 85-92.
4. Ren, X. and Moriya, S. The influence of target size, distance and direction on the design of selection strategies. In proceedings of BCS HCI'98 (Sheffield, UK) Springer, 1998, pp. 67-82.
5. Lewis, J., Potosnak, K. and Magyar, R. Chapter 54 - Keys and keyboards. In Helander, M., Landauer, T. and Prabhu, P. (ed.) Handbook of human-computer interaction, North-Holland, Amsterdam, NL, 1997, pp 1285-1315.
6. Hall, A., Cunningham, J., Roache, R. and Cox, J. Factors affecting performance using touch-entry systems: tactual recognition fields and system accuracy. Journal of applied psychology, 1988; 73: 711-720.
7. Beaudouin-Lafon, M. and Conversy, S. Auditory illusions for audio feedback. In ACM CHI'96 Conference Companion (Vancouver, Canada) ACM Press, Addison-Wesley, 1996, pp. 299-300.
8. Maury, S., Athenes, S. and Chatty, S. Rhythmic menus: toward interaction based on rhythm. In Extended Abstracts of ACM CHI'99 (Pittsburgh, PA) ACM Press, 1999, pp. 254-25.
9. Brewster, S.A. The design of sonically-enhanced widgets. Interacting with Computers, 1998; 11, 2: 211-235.
10. Brewster, S.A., Wright, P.C., Dix, A.J. and Edwards, A.D.N. The sonic enhancement of graphical buttons. In Proceedings of IFIP Interact'95 (Lillehammer, Norway) Chapman & Hall, 1995, pp. 43-48.
11. Dix, A.J. and Brewster, S.A. Causing trouble with buttons. In Ancillary Proceedings of BCS HCI'94 (Glasgow, UK) Cambridge University Press, 1994.
12. Dix, A., Finlay, J., Abowd, G. and Beale, R. (eds.) Chapter 9.4 Status/Event Analysis. Human-Computer Interaction, Prentice-Hall, London, 1993, pp 325-334.
13. Bevan, N. and Macleod, M. Usability measurement in context. Behaviour and Information technology, 1994; 13, 1 & 2: 132-145.
14. Hart, S.G. and Wickens, C. Workload assessment and prediction. In Booher, H.R. (ed.) MANPRINT, an approach to systems integration, Van Nostrand Reinhold, New York, 1990, pp 257-296.

15. Hart, S. and Staveland, L. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Hancock, P. and Meshkati, N. (eds.) Human mental workload, North Holland B.V., Amsterdam, 1988, pp 139-183.
16. Johnson, P. Usability and Mobility; Interactions on the move. In First workshop on human-computer interaction with mobile devices (Glasgow, UK) Glasgow University, 1998.
17. Petrie, H., Furner, S. and Strothotte, T. Design Lifecycles and Wearable Computers for Users with Disabilities. In First workshop on human-computer interaction with mobile devices (Glasgow, UK) Glasgow University, 1998.
18. Brewster, S.A. Using Non-Speech Sound to Overcome Information Overload. *Displays*, 1997; 17: 179-189.
19. Brewster, S.A., Wright, P.C. and Edwards, A.D.N. An evaluation of earcons for use in auditory human-computer interfaces. In Proceedings of ACM/IFIP INTERCHI'93 (Amsterdam, Holland) ACM Press, Addison-Wesley, 1993, pp. 222-227.
20. Blattner, M., Sumikawa, D. and Greenberg, R. Earcons and icons: Their structure and common design principles. *Human Computer Interaction*, 1989; 4, 1: 11-44.
21. Brewster, S.A. Using Non-Speech Sounds to Provide Navigation Cues. *ACM Transactions on Computer-Human Interaction*, 1998; 5, 3: 224-259.
22. Brewster, S.A., Wright, P.C. and Edwards, A.D.N. Experimentally derived guidelines for the creation of earcons. In Adjunct Proceedings of BCS HCI'95 (Huddersfield, UK), 1995, pp. 155-159.
23. Berglund, B., Preis, A. and Rankin, K. Relationship between loudness and annoyance for ten community sounds. *Environment International*, 1990; 16: 523-531.
24. Crease, M., Gray, P. and Brewster, S.A. Resource Sensitive Multi-Modal Widgets. In Volume II of the Proceedings of INTERACT '99 (Edinburgh, UK) British Computer Society, 1999, pp. 21-22.
25. The Economist. *Ring in the new*. The Economist Newspaper Limited, 9/10/99, 1999.
26. Edwards, A.D.N. and Stevens, R.D. Mathematical representations: Graphs, curves and formulas. In Burger, D. and Sperandio, J.-C. (ed.) Non-Visual Human-Computer Interactions: Prospects for the visually handicapped, John Libbey Eurotext, Paris, 1993, pp 181-194.