

Gestural and Audio Metaphors as a Means of Control for Mobile Devices

Antti Pirhonen

Department of Computer Science and Information Systems
University of Jyväskylä
P.O. Box 35; 40351 Jyväskylä; Finland
+358 (0)14 260 3025
pianta@cc.jyu.fi

Stephen Brewster and Christopher Holguin

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

Abstract

This paper discusses the use of gesture and non-speech audio as ways to improve the user interface of a mobile music player. Their key advantages mean that users could use a player without having to look at its controls when on the move. Two very different evaluations of the player took place: one based on a standard usability experiment (comparing the new player to a standard design) and the other a video analysis of the player in use. Both of these showed significant usability improvements for the gesture/audio-based interface over a standard visual/pen-based display. The similarities and differences in the results produced by the two studies are discussed.

Keywords

Mobile computing, metaphor, gestures, non-speech audio, evaluation.

INTRODUCTION

Designing interfaces for mobile computers (personal digital assistants (PDAs), mobile phones, etc.) is problematic as there is a very limited amount of screen resource on which to display information and users' eyes are often needed on the environment rather than the interface (so that they can look where they are going) which makes input difficult.

Music players (e.g., CD, MP3 or MiniDisc) are very common mobile devices and present many interesting design challenges. They tend to be used for long periods of time whilst the user is doing other things: walking, running, exercising, etc. Many PDAs and mobile phones are also now starting to incorporate such music technologies. The latest version of Microsoft's PocketPC platform (www.pocketpc.com), for example, includes a version of the Windows Media Player (see Figure 1). One problem with PDA-based solutions is that they are very visually demanding – to use the interface in Figure 1 you would normally take the de-

vice out of your pocket, remove the stylus and then tap the appropriate on-screen button. The buttons are small and grouped close together so much visual attention is needed to use them. If users are using the device whilst walking or driving, they cannot devote all of their visual attention to it; attentional resources must remain with the main task for safety. It is therefore hard to design a visual interface that



Figure 1: The Windows Media Player interface for the Microsoft Pocket PC platform running on a Compaq iPAQ.

can work under these circumstances. Some dedicated audio players have textured and shaped buttons that make it easier to press them without looking. The device in Figure 1 also allows its physical buttons to be used to control the audio player, however they are designed to be general purpose so are not specific to the media player and also cannot be used for every different application. With the concentration of more applications into general devices like Figure 1 it is no longer possible to have specific buttons for each function of each application so an alternative

solution is required. The aim of the work presented here is to try and overcome this problem with a new, multimodal, design.

Non-speech audio has been shown to be effective in improving interaction and presenting information non-visually on mobiles [2, 20]. It allows users to keep their visual attention on navigating the world around them and allows information to be presented to their ears. This has resulted in significant usability advantages [3]. This solves some of the problems of output but there is still difficulty with input; using the stylus requires the user to look at the screen to press the small buttons. This is hard to do whilst mobile. One solution to this problem is to use gestures [14] for input that do not require the user to look at the screen. Harrison *et al.* [8] showed that natural gestures can be used in a range of different situations in mobile devices to simplify input. We used this multimodal approach to create an audio player that could be operated completely non-visually using gestures and sound.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2002, April 20-25, 2002, Minneapolis, Minnesota, USA. Copyright 2001 ACM 1-58113-453-3/02/0004...\$5.00.

It is very important that interfaces to mobile devices are tested in realistic settings so that real usage problems can be found. Just testing in a conventional usability lab is not likely to find the problems that would occur in real mobile usage [10]. As yet there are no well-developed methods for testing mobile devices, so, as part of the project developing the audio player, we decided to try two very different evaluation techniques to see what kinds of usability problems each brought up so that we could learn more.

Creating Metaphors

The concept of metaphor is widely used in the context of GUIs. However, the use of the concept is often behind its origin in linguistics. In its original use its power is seen in illustrating unfamiliar entities with analogies (e.g., [12]). This is often the case in UI design, too. But in addition, it is sometimes used imperfectly, for example, when imitating real world entities (in these cases, it is hardly question of metaphor but rather of simulation, as in [8]). Also, some remarkable attacks against the use of metaphors [7, 15] are based on an idea of metaphor as imitation of real world entities. In our study, metaphor and metaphorical refer to analogies, not simulation or imitation. We believe that by using metaphors in this way they still have a lot of power as design principles. Also, in this study, the success of our design highly depends on the appropriateness of the chosen metaphors. To make our interface easy to ‘pick up and use’ we wanted to design the interaction around metaphors with which our users would be familiar.

DESIGN OF THE TOUCHPLAYER

The first research task was to design and implement a music player controlled with simple gestures and feedback given via non-speech sound. We called this the *Touch-Player*. For our work we used a Compaq iPAQ 3630 Pocket PC with a pair of stereo headphones. We mounted the player at the user’s side, on the hip, with a belt clip so that the screen was vertical (see Figure 7) but facing either right or left (depending on whether the user was right or left handed). We concentrated our design on the main functions offered by the Windows Media Player shown in Figure 1. These were: play/stop, next/previous track and volume up/down. An iterative approach was used in the design of the interface to the mobile music player so that we could ensure our designs were effective. We took the method proposed by Harrison *et al.* [8], trying to map the natural gestures a user would make to the interface functions. Gestures were made by the user moving a finger across the screen. We could then recognise our gestures in the same way as normal stylus interactions on the screen. The key metaphors used in our application were related to the parallel between physical directions and logical order. After initial design discussions we decided upon:

- Sweep across screen left side → right side = next track (this could be reversed for left-handed users)
- Sweep across screen right side → left side = previous track

- Single tap = start/stop
- Sweep from bottom → top of screen = volume up
- Sweep from top → bottom of screen = volume down

In the use of gestural metaphors the location and position of the device were critical. The metaphor of forward a track was to go physically forward, i.e., the sweep ought to be from back to front. Therefore, the fixing point should be clearly on the side of the user. If it was in front of the user the sweeps would be in left-right axis, and the forward backward metaphor would not work. For the metaphor of going up and down (increase/decrease volume), it was essential that the device be fixed in vertical position.

The gestures were all simple to recognise (as they were single points or straight lines). For example, to recognise a *next track* gesture we simply compared the start (x,y) coordinates of the gesture with the end ones. If the user had moved more than 40 pixels in the x direction then the gesture was valid and if it finished within ±10 pixels of the start y coordinate it was considered a *next track*. If the user had not moved 40 pixels in x then the gesture was not valid. This was done to avoid inadvertent gestures that could occur when the user tapped against the screen by mistake when using the device on the move. If the user finished the gesture within ±10 pixels in the y direction then it was considered valid. This gave some flexibility so that the user did not have to do a perfectly straight sweep – a small drift in direction would still generate a *next track* gesture. Again this was important as the hand and the device could both move about a lot when the user was mobile.

One question was: in which direction should the volume sweep be? Bottom to top for volume up seemed natural. However, we also considered that a top to bottom ‘push’ might have been more natural (like a fader on a mixing desk, which is pushed away from the user to increase volume). We carried out an initial think aloud trial with seven users to investigate all the gestures developed. Users wore the device on their hip and had to perform the gestures with a finger. This allowed us to see if the gestures were appropriate. We gave the users sets of gestures to perform and recorded the correctness of the gestures made. We then interviewed them to see what they felt about the gestures and if they were appropriate. Users performed the gestures well with very few instances of mistakes. The users rated the gestures as very natural and none felt that the ‘push’ gesture was appropriate for volume increase.

The sounds used to give feedback on the gestures were earcons [1, 4] as these have been shown to be effective in improving the usability of mobile devices [2]. It was important to provide feedback on the gestures as users would not see anything but would need to know the state of the device. The sounds were pre-recorded stereo sound samples played through the headphones. We used sounds to give feedback on the gestures made so that users knew if the gesture had been recognised or not. For example, start and stop sounds were simple beeps taken from a desktop

PC. Next track had a sound that increased in pitch and previous track a sound that decreased in pitch. We did not give any explicit audio feedback for volume changes, as the change in the music playing would indicate this change implicitly. The sounds were spatialised (using stereo panning) on a horizontal line in front of the user to represent the display of a CD player control panel to reinforce the metaphor we were using. The sound for next was played on the right, start/stop in the centre and previous on the left. Twelve audio tracks were included on the device so that it functioned like the standard Pocket PC Media Player.

EVALUATION OF THE INTERFACE

One aim of our work was to compare two radically different evaluation techniques in a mobile setting to see what results were obtained and how the two techniques were similar and different.

Study 1: Formal usability experiment

For the experiment we developed ideas that had been proposed by Petrie *et al.* [17] and used successfully by Brewster [3]. The aim of these was to test a mobile device interface whilst the user was mobile in a fairly realistic environment but to keep enough control so that measures could be taken to assess performance.

Experimental design

A 2-condition within-groups design was used with interface type as the independent variable. An instrumented analogue of the standard PocketPC Media Player interface was used as the control (called *MediaPlayer*) and compared to the gestural/audio interface above. Fifteen participants took part; all were students at the University of Glasgow (none took part in the other study below).

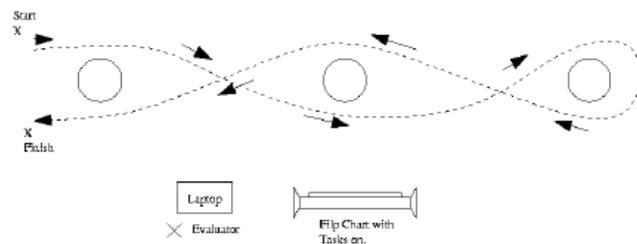


Figure 3: Route walked by the participants.

Participants wore the device on their hip with the iPAQ sitting vertically against the leg (but could remove and hold it if needed) and performed both the *MediaPlayer* and *TouchPlayer* conditions in counterbalanced order. Users had to walk seventeen 8 metre laps of a corridor at the University whilst performing a set of tasks (see Figure 3) in each condition. This allowed us to take measurements of the usage of the device whilst the users were mobile but was not as formally controlled as a laboratory study, which would lack realism and ecological validity. Interruptions were possible as other people used the corridor but these would affect all users in a random way. A brief training session was given before each condition and questionnaires

completed afterwards. The whole experiment took approximately 20 – 25 minutes.

Users walked a series of laps of the corridor whilst performing a set of 17 tasks per condition (one set per lap). The tasks were presented on a flip-chart (so that users could see them as they walked along) and consisted of instructions such as: “Find the song Wonderwall” or “Move forward two tracks and increase the volume”. Users were instructed to read the list of tasks from the chart at the start of each lap and then perform them (the tasks were simple and could easily be completed in one lap). The experimenter recorded lap times on a laptop and turned over the flip chart to present the next set of tasks at the end of each lap. Other timing and error data were recorded on the iPaq.

To get a balanced view of the usability of the interfaces being tested a full range of quantitative and qualitative measures was taken. Overall time taken to complete the experiment, time taken to complete the individual tasks, errors in the gestures made and subjective workload (NASA TLX) [9] were all recorded. Workload is important in a mobile setting as users must monitor their surroundings and navigate, therefore fewer attentional resources can be devoted to the computer. An interface that reduces workload is likely to be successful in a real mobile setting. We added an extra factor to the standard TLX test: 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. We also recorded percentage preferred walking speed (PPWS) [17]. PPWS has been used as an evaluative measure for electronic travel aids for blind people [5] but has not been used before with standard mobile devices. As Petrie *et al.* [17] say “The rationale underlying the use of this measure is that all pedestrians have a walking speed which they prefer and that this appears to be the speed, which for them, is the most physically efficient. The ability of any mobility aid to allow the pedestrian to walk at this preferred walking speed (PWS) is therefore argued to be a measure of its effectiveness”. This measure would allow us to assess the impact of our device on the participants: the further below their normal walking speed that they walked the more negative the effect the device was having on them. Before the experiment began participants walked 10 laps of the corridor and the times of the middle 5 laps were recorded and averaged so that we could calculate their standard PWS without using the device.

The main hypotheses for the experiment were that the *TouchPlayer* interface would improve usability when on the move. Users would experience a lower subjective workload, should be able to complete tasks more quickly and walk closer to their normal walking speed.

Results & discussion

All the data were analysed using within-groups *T*-tests. Subjective workload was reduced in the gesture/audio condition. Overall workload was significantly reduced in the TouchPlayer condition as compared to the standard MediaPlayer (mean overall workload in TouchPlayer condition = 25.4 (SD=38.6), MediaPlayer = 43.8 (SD=59.5), $T_{14}=6.41$, $p<0.001$). Figure 4 shows the results for the individual workload categories. In all categories (except Performance Achieved) the workload for the TouchPlayer interface was significantly lower. In particular Mental Demand reported by participants was significantly lowered by the new interface ($T_{14}=6.54$, $p<0.001$).

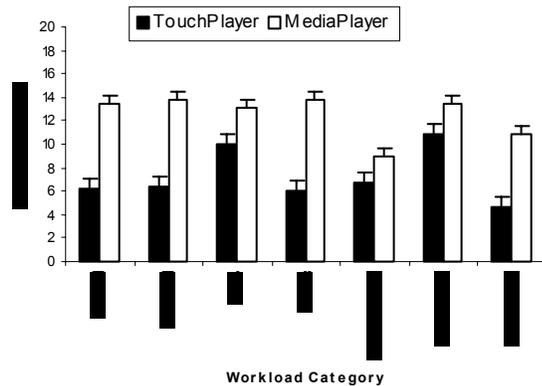


Figure 4: Average Workload scores for the two conditions. Standard error bars are shown.

There was a significant reduction in the overall time taken for the experiment. The average time taken in the TouchPlayer condition was 143.3 secs. and in the standard condition 164.7 secs. ($T_{14}=5.00$, $p=0.0001$). The interfaces had a significant effect on PPWS (see Figure 5). In all cases the TouchPlayer was closer to the standard preferred walking speed than the MediaPlayer. The average PPWS in the TouchPlayer condition was 109% (SD=45.6) than the MediaPlayer 126% (SD=196.7) ($T_{14}=5.1$, $p=0.0001$). As can be seen from the standard deviation results, the variation in the MediaPlayer was much greater than with the TouchPlayer. There was no difference in the number of errors made with input in the two interfaces. The mean number of errors in the TouchPlayer condition was 3.6, versus 3.2 in the MediaPlayer condition. The main source of error with the gestures in the TouchPlayer condition was with the start/stop gesture (which accounted for 60% of the errors). It was too easy to do a single tap by mistake as the device and users hand were both moving and it was easy for a finger to inadvertently tap the screen.

The results from this evaluation showed some significant usability benefits from the use of the gesture/audio interface. There was a major reduction in subjective workload and, in particular, mental demand. This is an important factor as it deals with the amount of cognitive resource the user needs to dedicate to using the interface. In a mobile

setting users must firstly deal with the environment, therefore fewer resources can be devoted to the computer. A reduction in mental demand is therefore likely to make a big difference in real-world usage.

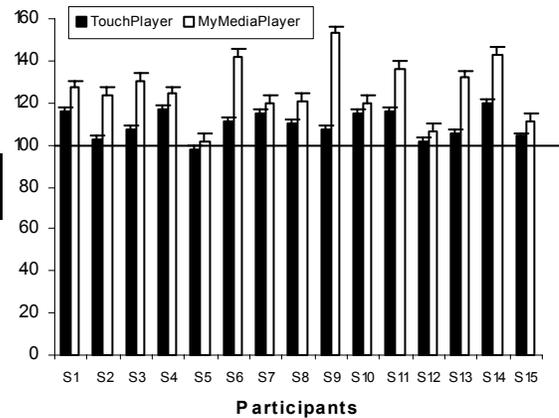


Figure 5: Percentage Preferred Walking Speeds for the participants in both conditions. The 100% line shows the preferred walking speed.

Participants completed the experiment more quickly in the TouchPlayer condition but did not make more errors. The problem with the errors in the play/stop gesture has now been corrected by making the gesture a double-tap. This is far less likely to happen by chance from a finger banging against the screen when on the move, but is still easy to do. Another important result is that users walked closer to their normal preferred walking speed when using the TouchPlayer. This suggests that the interface put much less load on the user than the standard one and that PPWS is a useful measure of mobile usability. This evaluation showed the new multimodal interaction techniques are a very effective way of improving the usability of mobile devices.

Study 2: Video analysis of the device in use

In this study the experimental focus was on the behaviour of the users. The method used was highly grounded in the data. This meant that experiment sessions were videoed, and the way the video data were utilised in the analysis was decided on the basis of the quality of the collected material [11]. The research task was to investigate 1) the usability of the application and 2) the applicability of the method. The results should be the basis for further development of the application and the methodology.

Experimental design

The design of the study had two, somewhat contradictory requirements: Firstly, the experiment should resemble field conditions as much as possible. Secondly, the setting should allow high quality video filming with two cameras (close-up and over-all views) for good quality data collection. As the use of our application in mobile conditions would have made it impossible to get good video, we conducted the experiment in a laboratory. However, the

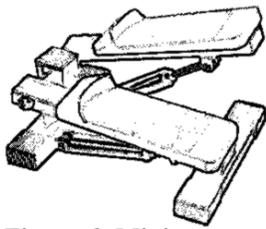


Figure 6: Mini-stepper used in Study 2.

evaluation was arranged to resemble field conditions as much as possible by using a simple mini-stepper (Figure 6). Users walked on the mini-stepper to simulate walking in the real world. The height of the step could be adjusted so that the movement of user's feet combined the properties of walking on a flat surface and

walking up stairs. The iPaq was worn as before, thus keeping the device in a stable position. It was operated as before, with the same gestures and sounds.

The number of subjects was originally 10, but due to technical problems the data of four sessions were discarded making the final number of sessions analysed 6. All of the participants were Computing students at the University of Glasgow. Three were male and three female. The backgrounds of the students were extremely varied. Two of the participants were undergraduates; the others were doing a Masters conversion course in information technology.

All sessions consisted of five parts. Firstly, there was a brief interview (about 5 minutes) in which relevant information about the background of a participant (field of study, previous experience of portable music players, etc.) was discussed. The participant was then given very brief instructions about the use of the TouchPlayer. The interview was recorded in the same way as the actual test. The second part was a trial session, in which the participant became familiar with the functions of the application. The session was recorded as shown in Figures 7 and 8. The researcher asked the participant to walk slowly on the stepper while using the TouchPlayer (the MediaPlayer interface was not used in this study). The participants were given instructions on what to do with the player, for example "play", "stop", "next track", "previous track", "find the first track", "find the last track", "find your favourite track", and "count the number of tracks". These were given by showing the participants instructions written on a card. Throughout the experiment, the participants stepped slowly on the stepper, listened to the music through the headphones, and followed the instructions given. This part took about 10 minutes, during which the number of user actions ranged from 80 to 160. There was no fixed set of instructions, instead the researcher decided at each point of time what the next task should be. The reason for this was that it was important that each task made sense for the participant. For example, if the participant had performed "find your favourite track", and it happened to be the first one, it would not have been appropriate to next ask: "find the first track".

Two video cameras were focused on the participants: one for close-up view of the device and the finger controlling it and one for an overall view. These videos were mixed and stored on a digital video (DV) recorder (see Figure 7). The speech from the participants and experimenter, along with

audio from the TouchPlayer, were also stored on the DV (see Figure 8).

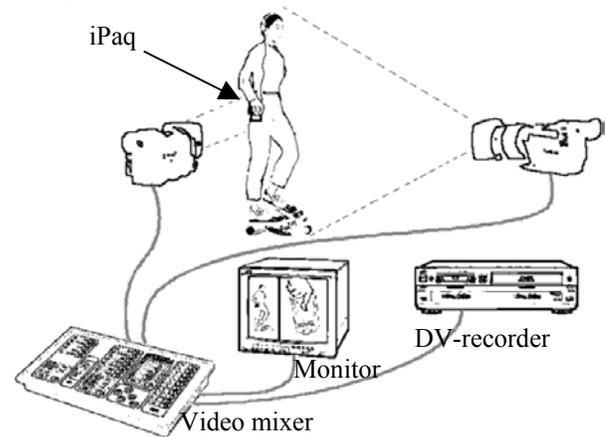


Figure 7: Collection of video information.

The third part was another interview, in which first reactions to the device were discussed (5-10 min). The interview was again recorded.

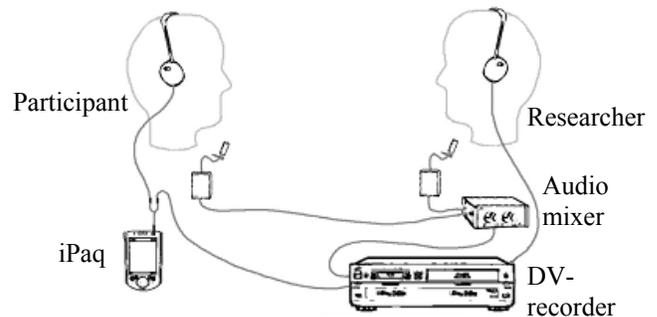


Figure 8: Collection of audio information.

In the fourth part each participant again performed the given instructions. This time, the feedback sounds varied at certain intervals (either panning, pitch, or both). These changes were programmed in the application as a special 'experiment mode'. First, the set of feedback sounds was as in the trial session (set 'A', Table 1). After 20 actions, the set changed to 'B', which meant that the panning of forward and backward sounds changed (Table 1).

Table 1: Changes in the audio feedback used in Study 2.

Set	Function	Pitch	Panning
A	forward	increasing	right
	backward	decreasing	left
B	forward	increasing	left
	backward	decreasing	right
C	forward	decreasing	right
	backward	increasing	left
D	forward	decreasing	left
	backward	increasing	right

After certain number of actions (varying from 16 – 20 depending on the set), the set changed again. The sequence of

sets was A, B, A, C, A, D, A. Participants were asked to hold up a hand whenever they noticed a change in the audio feedback. In this way we tried to find out whether the participant consciously utilised different properties of the feedback sounds or not. This part took 10 minutes. Finally, we were interested in the participants' conceptions of the device – to see if the metaphors developed by the designers supported these user conceptions. Therefore, the participant was asked to sketch rapidly a visual counterpart to the user-interface of this application. The researcher asked questions about the drawing in order to complement the information from the actual drawing.

The data in DV format made it possible to analyse the data in many ways. Firstly, it enabled us to make interpretations of the subjective qualities of user behaviour on the basis of interviews and overall observation. Secondly, accurate statistical analysis could be made of details of the behaviour by encoding the events with the help of DV's time code.

Results & discussion

The results of part 2 of the study showed relatively uniform learning curves across the participants. A typical example is shown in Figure 9. The figure shows the learning curve of one participant learning the function "previous track". The Y-axis shows the percentage of the last 5 attempts that have been successful. In this case, 37 attempts are analysed (the actual number of attempts was 42, but counting starts from the fifth attempt). In the figure, the first bar refers therefore to the fifth attempt). It can be seen, that after 9 attempts the function was learned so well that there had been five successive successful actions (100% success). The curve thus describes the progress of learning better than, for example, the overall error rate, which is highly effected by the first, unsure attempts. In most cases the learning curve followed the same pattern as in the case illustrated in figure 9. Initially the use of a function was learned rapidly, but later the performance deteriorated 2-3 times. This was particularly the case with the functions that had clear feedback sounds ("next track", "previous track"). Even the periodical deterioration of performance may indicate an increased feeling of control: after a certain amount of trust in the functioning of the application, users may act more carelessly if they know that an error can be easily

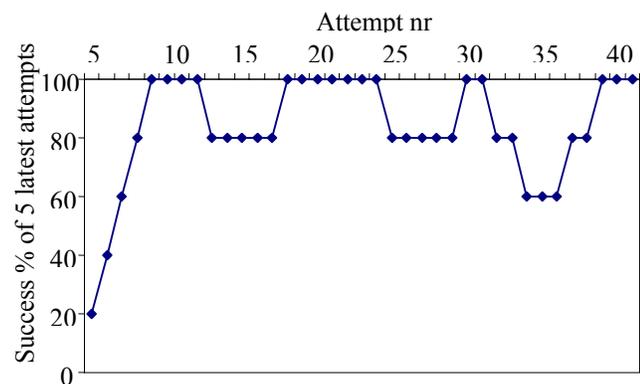


Figure 9: Typical progress of learning the *previous track* command.

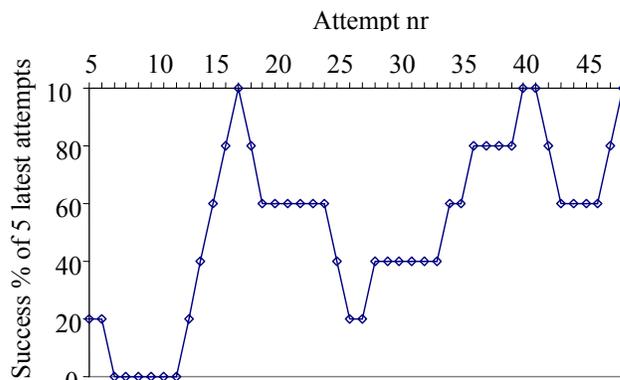


Figure 10: Typical progress of learning the *decrease volume* command.

corrected. The learning of the functions that lacked clear and immediate feedback ("increase volume", "decrease volume") was much slower. Figure 10 shows the learning curve of the same participant learning the "decrease volume" function.

The overall success percentage was also significantly higher in the functions with immediate, clear feedback. In forward and backward sweeps the mean percentages were 86.2% (SD 6.9) and 89.0% (SD 6.5), whereas in increasing and decreasing volume the mean percentages were as low as 64.0% (SD 13.9) and 43.8% (SD 16.1). The video analysis showed that in the case of error the participants corrected it very rapidly if they got immediate audio feedback. However, the functions lacking clear feedback caused uncertainty and frustration. These functions were used more carefully during the experiment, seemingly because of the fear of error. The problem with the up/down gesture has now been corrected by the addition of an explicit volume change sound (a short beep played at the new volume) when the volume has changed. This provides the explicit feedback that the users needed.

On the other hand, the actual types of feedback sounds (from part 4 of the study) did not play any significant role. Only one of six subjects reliably noticed the changes in the panning of forward and backward feedback sounds. The interviews also revealed that all the participants found feedback sounds necessary, but that the types of sounds were not important. This disparaging of the qualities of feedback sounds is most probably related to three things. Firstly, there was only a small number of functions so it was quite easy to make the distinction between them without multiple cues. The types of sounds may become important when the number of functions is increased to that of the full Media Player. Secondly, the main role of the feedback sounds in this case was to confirm that the touch was registered by the application (that the touch was strong enough), not that the gesture was the right one. Still it is possible that the sounds were not illustrative enough to support the directional metaphor. The third possible reason for the ignorance of audio details arises from the metaphors. The gestures for the next and previous functions were directed from back to front and front to back of the

user, but the feedback sounds illustrated a CD player control panel and were therefore located on left-right axis.

The drawing task (when each participant sketched the visual counterpart of the UI of this application) showed that the control panel metaphor was very strong. Five of six drawings had the previous track button on the left, symbol pointing to the left and next track button on the right, pointing

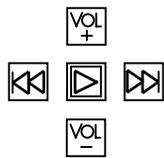


Figure 11:
Participants'
suggestion
for the GUI.

to the right (in one case, the corresponding push-buttons were up and down). Figure 11 shows the most common style (3 drawings were like this). Despite the strength of a left-right metaphor only one participant even recognised the change in panning, so it seems that there was confusion between the two contradictory spatial metaphors (see [18]). As a result, the metaphor that was related to a user's own actions (sweeps forward and backward) dominated and reduced the effectiveness of the panning of the sounds.

Subjective satisfaction toward the application varied a lot amongst the participants. This could be seen in the test sessions and the interviews. Even if the basic principles of the implemented user-interface were supposed to be quite novel and intuitive, the most satisfied participants were young, male students of computing – the ones who could have been supposed to be most oriented toward a more computer-like interaction. In addition, these participants' gestures were stronger and faster than those of the others, indicating less computer anxiety. This was particularly important in the use of the iPaq touch-screen, because it is designed to be used with a stylus but was used in our application with a finger. Inadequate force when pressing on the screen was one of the most common problems in the use of the application.

Concerning the development of the method it can be concluded that digital video provides a compact, high quality and accurate recording and analysis tool. Using separate close-up and overall views proved a suitable combination when trying to make both general observations and detailed analysis about the movements fingers. To developing this evaluation method further an additional close-up camera could be used to improve the reliability of the observations. This additional camera could focus on the fingers on the screen as well as the first, but from a different angle to avoid problems of the hand occluding the display.

Comparison of the results

Both studies collected detailed information about usability problems, but, interestingly, the results overlapped very little. They both uncovered different types of problems. For the usability experiment the data collected were accurate and easy to handle with statistical methods. For the qualitative experiment (as usual in qualitative research) the accuracy was not as high but the scope was wider. The usability experiment showed overall effects of the interface on the user in the environment and compared the new design to an existing one. It did not give details about the usability of

the individual parts of the interface. The video study elicited much more information about the gestures and the use of the sounds but gave little information about the impact of the device on the user on the move.

One major difference in the results was caused by the differences in the data collection methods. Some erroneous user-actions were not registered in Study 1 as the users' gestures were done too softly on the screen and were therefore not picked-up by the device (as far as the device was concerned no gesture had been made so no error was recorded). In Study 2 these could be observed independently though the video. The percentage of errors classified in this way was 53 in Study 2.

Each factor of usability should be investigated with the appropriate method [13]. These two studies strengthened the widely held belief (e.g., [6]), that qualitative and quantitative methods should not be seen contradictory, but that these methods usually complement each other. Here, the quantitative experiment provided accurate information how the user succeeded in effecting the application. In turn, the qualitative experiment helped to see the indicated user actions from the perspective of user's intentions.

The core of the differences between these two studies was in the object of comparison. In the formal usability study, the TouchPlayer was exposed to direct comparison with an alternative design (Media Player). In the other study, the comparison was of the participants' mental representations [19]. These representations are formed on the basis of previous experiences with different kinds of music players and on the given instructions, i.e., evoked expectations. Study 1 benefits from having a clear physical implementation with which to compare the TouchPlayer. As both of these designs (TouchPlayer and MediaPlayer) could be tested with the same method, straight comparisons were possible. On the contrary, Study 2 lacked this opportunity, as there is no straight access to the mental representations. Therefore, Study 2 had a somewhat weaker basis for arguing for and against the design of TouchPlayer, but on the other hand, has a very relevant object of comparison - the participants "ideal" player. In other words, Study 1 shows that the TouchPlayer is in many ways more usable than MediaPlayer, but does not tell whether even that meets participants' requirements. Therefore, both of these studies were needed to get a reliable overall view of usability.

DISCUSSION AND CONCLUSIONS

The results from the studies show that it is possible to create an effective mobile music player without a visual interface. This is important as mobile users must focus their visual attention on the world around them and not on the device they are using. Most existing dedicated mobile music players do this with physical buttons. As more functions and applications become incorporated into PDAs this is no longer possible and it is easier to add more gestures than it is to add more buttons.

Our results showed it was very important to have explicit and immediate audio feedback on the gestures, as without it

users did not feel confident using them (even if implicit feedback was given). In addition, the gestures must be reliable to do on the move. Simple taps were too easy to do by mistake with all of the movement around the device when the user was in motion. This would not have been found if the gestures had been tested with the users performing them standing still (and was not seen in our initial think aloud).

The gestures used in the TouchPlayer have now been extended to a larger set and we have included synthesised speech so that we can now provide more of the track management features not available in the experimental prototype. The multimodal solution we designed could be applied to a range of other applications as the forward/backward, up/down and select operations are very common, e.g. skipping through emails, voicemails or news articles being read out, or in the type of eyes-free music player from Pauws *et al.* [16], or a TV remote control.

Developing successful evaluation methods for mobile devices is an active area of research. New devices are being developed all the time and new ways of evaluating them in realistic use are needed to ensure their effectiveness. The work presented here has shown some new evaluation techniques and demonstrated that they can elicit useful information about usability. The combination of two very different techniques was shown to bring out a wide range of different usability problems. The work presented here gives designers of mobile devices new multimodal interaction techniques to allow their devices to be used more effectively on the move and some new tools to allow them to assess the effectiveness their future designs.

ACKNOWLEDGEMENTS

This work was funded by EPSRC Grant GR/L79212 (UK) and the National Technology Agency (Finland).

REFERENCES

1. Blattner, M., Sumikawa, D. and Greenberg, R. Earcons and icons: Their structure and common design principles. *Human Computer Interaction* 4, 1 (1989), 11-44.
2. Brewster, S.A. and Cryer, P.G. Maximising Screen-Space on Mobile Computing Devices. In Summary Proceedings of ACM CHI'99 (Pittsburgh, PA) ACM Press, 1999, pp. 224-225.
3. Brewster, S.A. and Walker, V.A. Non-Visual Interfaces for Wearable Computers. In Proceedings of IEE workshop on Wearable Computing (IEE, London) IEE, 00/145, 2000.
4. 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.
5. Clark-Carter, D.D., Heyes, A.D. and Howarth, C. The efficiency and walking speed of visually impaired people. *Ergonomics* 29, 6 (1986), 779-789.
6. Garcia, L. and Quek, F. Qualitative research in information systems: Time to be subjective? In Proceedings of Information systems and qualitative research (Philadelphia, USA), Chapman and Hall, 1997, pp. 444-465.
7. Gentner, D. and Nielsen, J. The anti-mac interface. *Communications of the ACM* 39, 8 (1996), 70-82.
8. Harrison, B.L., Fishkin, K.P., Gujar, A., Mochon, C. and Want, R. Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. In Proceedings of ACM CHI'98 (Los Angeles, CA) ACM Press Addison-Wesley, 1998, pp. 17-24.
9. Hart, S.G. and Wickens, C. Workload assessment and prediction. In MANPRINT, an approach to systems integration, Boohar, H.R. (Ed.), Van Nostrand Reinhold, New York, 1990, 257-296.
10. Johnson, P. Usability and Mobility; Interactions on the move. In First Workshop on HCI with Mobile Devices (Glasgow, UK) Glasgow University, 1998.
11. Kvale, S. Issues of validity in qualitative research. Chartwell Bratt, London, 1989.
12. Lakoff, G. and Johnson, M. Metaphors we live by. The University of Chicago Press, Chicago, 1980.
13. Lee, S.H. Usability testing for developing effective interactive multimedia software: concepts, dimensions, and procedures. *Educational Technology & Society* 2, 2 (1999).
14. Long, A., Landay, J. and Rowe, L. Implications for a gesture design tool. In Proceedings of ACM CHI'99 (Pittsburgh, PA) ACM Press Addison-Wesley, 1999, pp. 40-47.
15. Nardi, B.A. and Zарner, C.L. Beyond models and metaphors: Visual formalism in user interface design. *Journal of Visual Languages and Computing* 4 (1993), 5-33.
16. Pauws, S., Bouwhuis, D. and Eggen, B. Programming and enjoying music with your eyes closed. In Proceedings of ACM CHI2000 (The Hague, The Netherlands) ACM Press, Addison-Wesley, 2000, pp. 376-383.
17. Petrie, H., Furner, S. and Strothotte, T. Design Lifecycles and Wearable Computers for Users with Disabilities. In First workshop on HCI mobile devices (Glasgow, UK) Glasgow University, 1998.
18. Pirhonen, A. Analysis of the concept of redundancy concerning the design of multimodal combinations of output-elements. In Proceedings of APCHI98 (Los Alamitos, CA) IEEE, 1998, pp. 273-278.
19. Pirhonen, A. Creating virtual objects. In Proceedings of IRIS 22 Technical Report TR-21, Vol. 3. University of Jyväskylä, Department of Computer Science and Information Systems, 1999, pp. 57-69.
20. Sawhney, N. and Schmandt, C. Nomadic Radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Human-Computer Interaction* 7, 3 (2000), 353-383.