An Evaluation of the Communicative Ability of Auditory Icons and Earcons

Paul A. Lucas
School of Computing and Mathematics
The University of Huddersfield, Queensgate
Huddersfield, HD1 3DH, UK
Email: p.lucas@hud.ac.uk

Abstract

Audio cues are increasingly used to present information in the human-computer interface. To be of use, these cues must effectively communicate the intended information to the user. An evaluation of two types of nonspeech audio cue, auditory icons and earcons, was performed. The audio cues were designed to communicate information about actions and objects in the human-computer interface. The importance of audio cue design knowledge in the use of the cues was investigated. The results indicate that knowledge of the auditory cue design can significantly improve the accuracy of cue recognition.

1 Introduction

In recent years interface designers have been turning to the auditory channel as a way of presenting useful information to the user. Two methods of extending the direct manipulation interface into the auditory domain are auditory icons [1] and earcons [2].

1.1 Auditory Icons

Auditory icons are caricatures of sounds occurring as a result of our everyday interactions with the world. These real-world sounds are mapped onto events and objects in the interface about which they provide auditory feedback. The basis of the auditory icon is the construction of an analogy between the computer-world event and its real-world counterpart. In the SonicFinder [3], an interface which makes use of auditory icons, everyday sounds are used to reinforce and supplement the visual feedback. Dragging a file across the desktop elicits a scraping sound, while dropping a file into the trash is accompanied by a crashing sound. The power of the auditory icon lies in the directness of the mapping between the interface event (selecting a file) and the sound producing event (tapping an object). The ability to perform these mappings is limited, however, by the number of analogies that exist between the interface world and the real world. As long as acoustic parallels can be found between the real and computer worlds, this directness of mapping is supported. It is recognised, however, that there are limits to the number of direct parallels that can be found. Despite the limited number of direct analogies that exist between the real and computer worlds, auditory icons can still be constructed. The result, however, is that in certain cases the mapping between the interface feature and its auditory representation may be less direct and therefore less effective.

1.2 Earcons

Earcons are used to represent objects or actions in the interface. It is the ability to combine these simple representations into more complex messages that gives earcons their communicative power. A compound earcon may, for example, indicate the creation of a file by combining the audio representations for 'create' and 'file' [2]. Due to the abstract nature of the *musical* sounds being used in their construction, earcons have no inherent meaning. Brewster [4] identifies this when noting that "Earcons are completely abstract: the sound has no relation to the object it represents." This results in arbitrary mappings between the sounds and the interface features they represent. Thus, there would appear to be less capacity for the first time or novice user to correctly associate the earcon with its meaning.

1.3 Communicating with Auditory Cues

The success of auditory cues in the interface is largely dependent upon the development of good analogies between the sounds and the events or objects they represent. The closer the metaphorical link between the sound and its intended meaning, the easier that link will be recognised. It is the ability of the user to recognise this association that determines the usefulness of the cue. Thus, if the mappings are not known or hard to determine, the meanings will be ambiguous. The sounds proposed for use in the interface must, therefore, be tested for their ability to successfully convey their meanings.

The work of Brewster, Wright and Edwards [5] indicates that earcons can be an effective means of communication in the interface. Earcons were found to be better at presenting information than symbolically mapped, unstructured bursts of sound for users with minimal exposure to the cues. Thus, the theory underlying the construction of the earcons appears to provide a structure whereby the abstract sounds can be useful and informative. This structure is not always enough, however, to aid users in their interactions. Barfield, Rosenberg and Levasseur [6] found that in menu navigation tasks, subjects failed to improve their performance with the introduction of earcons.

Auditory icons have been used in a number of systems, such as SharedARK [7], the SonicFinder [3], and ARKola [8]. The results of the ARKola simulation [8], which used auditory icons to aid in the running of a soft drinks factory, suggests that auditory icons can be used successfully in multiprocessing and collaborative systems. It was found that subjects were quick to learn the meanings and functions associated with the auditory icons. This lends support to the idea that users will readily make the connections between real world sounds and their counterpart representations in the interface.

Jones and Furner [9] investigated the ability of auditory icons and earcons to convey their meanings. The results showed that after recorded natural speech, the cue most accurately associated with its meaning was the auditory icon. The most preferred cue after speech, however, was the earcon. The comparatively low association accuracy scores recorded by Jones and Furner indicate that the information carrying potential of the cues is not being fully realised. The audio cues being tested, may lack the intuitiveness required for novice users to determine their meanings. It is suggested that providing the users with knowledge of the theory underlying the construction of the auditory cues can overcome this lack of intuitiveness and improve their use of the cues. Knowledge of cue design should help the users determine the intended mappings between the sounds they hear and the interface features represented. This knowledge should increase the users understanding of the cues, compensating for a lack of directness in the mappings between the sounds and the interface features they represent.

2 Method

2.1 Subjects

Sixteen adult undergraduate students of computing, both males and females, participated in this experiment. The ages of the subjects ranged from 20 years to 33 years. The level of computing experience possessed by the subjects ranged from 2.5 years to 10 years. None of the subjects had any formal musical training nor had they heard the audio cues prior to this experiment. Each of the subjects was a volunteer and received no remuneration for their participation.

2.2 Materials

A HyperCard program presented the auditory stimuli and recorded the subjects responses. The three screens contained twelve buttons each; six sound buttons and six choice buttons. The layout of each screen appeared identical. The labels on the choice buttons consisted of descriptions of the objects and actions represented (e.g., delete file).

Six sounds from each audio cue design type were chosen for use in the experiment. The same objects and actions were represented in each type and were designed following the guidelines [1, 2]. The speech cue consisted of the name of the command function. Speech was included since it was seen as the audio cue most likely to invoke the best possible response times and accuracy from the subjects and thus provide a good contrast for the other sound types.

3 Procedure

After listening to the familiarisation presentations, the subjects were allowed to proceed with the presentation/selection tasks at their own speed. The tasks involved the subjects choosing a sound button, listening to the cue, then choosing the choice button which they felt most closely represented the sound. This was repeated for all sounds in each type. The subjects were not restricted in their order of selection of the sound buttons. The response time began after the presentation of the cue and ended when a choice button was clicked. Thus, the duration of the presentation of the cue did not affect the response times. The order of presentation of the sound types was altered randomly between subjects. This random order of presentation was engineered in order to reduce the possibility of the presentation of one sound type affecting the subjects performance in the subsequent sound type tests.

After testing, half the subjects, chosen at random, were given an explanation of the different design methods and the metaphors used in the construction of the sound cues. One week after the first testing, all subjects were tested again. The second testing consisted of the same presentation/selection procedure as in the first test, including the familiarisation presentation of the audio cues for each type. Those subjects who received an explanation after the first test were presented with the explanation again prior to the second test.

3.1 Pilot Study

A pilot study, using two subjects, was conducted, prior to the main study. It was noted that the subjects performance in the association tasks was affected by the shock of initial presentation of the unfamiliar sounds. To counteract this effect, each of the sounds was presented to the subjects once prior to performing the associations in the main study. These familiarisation presentations increased the likelihood that the subjects performance would not be affected by the shock of initial presentation of the sounds.

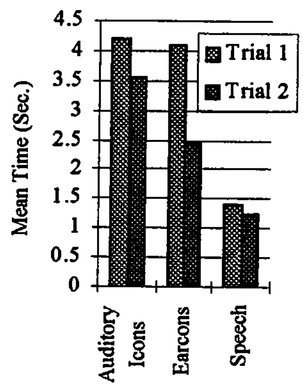

Figure 1: Mean association times by sound type over both trials.

	Trial 1	Trial 2
Mean time (sec.)	3.20	2.43
Mean Errors	2.52	2.00

Table 1: Mean association times and errors for all sound types in each trial.

4 Results

A balanced analysis of variance (ANOVA) was carried out to analyse errors and response times for the three sound types. Figure 1 shows the mean time taken by subjects to associate the audio cue with its meaning. The subjects improved their association times on the second trial, for each sound type. As expected, the earcons and auditory icons showed more improvement than speech. The improvement on the second trial was much greater for the earcons than for the auditory icons. The differences in improvement, however, were not significant ($F2, 28 = 1.21, p > 0.05$).

It can be seen in Figure 2 that speech showed the fastest overall association time, with auditory icons showing the slowest. The results indicate that the difference between the association times for each sound type was significant ($F2, 28 = 19.94, p < 0.001$). This was shown to result from the much lower times recorded for speech, since an ANOVA performed on the auditory icons and earcons, excluding speech, showed no significant difference between sound types ($F1, 14 = 3.25, p > 0.05$).

It can be seen from Figure 3 that both the explanation and no-explanation groups showed a decrease in association times. Although the decrease in association times was greater for those subjects receiving the explanation, this decrease was not significant ($F1, 14 = 0.00, p > 0.05$). Table 1 shows the mean association times, for all sound types, in each trial. The analysis showed that the decrease in association time on the second trial was significant ($F1, 14 = 9.98, p = 0.007$). The reduction in errors on trial two was also shown to be significant ($F1, 14 = 5.09, p = 0.041$).

Figure 4 shows that those subjects receiving an explanation about the design of the cues, reduced their association errors, while those subjects without an explanation, showed no decrease in errors. This result is suggestive of an interaction between explanation and trial. It was shown

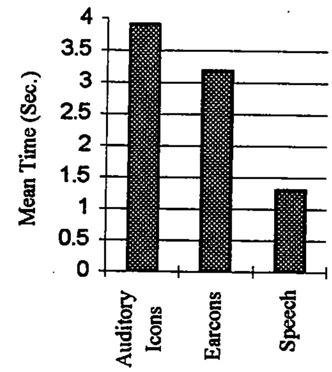

Figure 2: Mean association times by sound type.

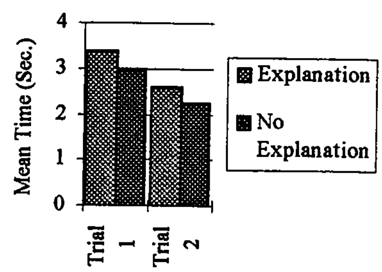

Figure 3: Mean association times, for all sounds, by explanation group.

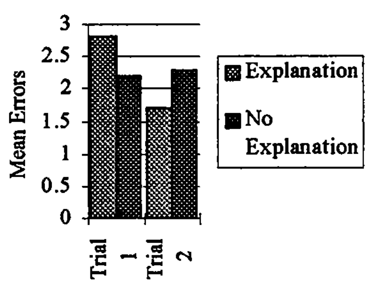

Figure 4: Mean association errors by explanation group.

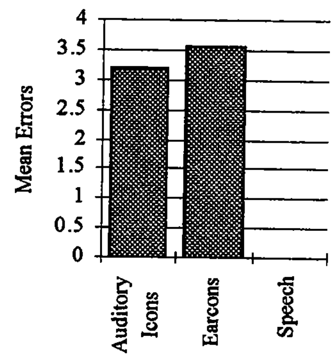

Figure 5: Mean association errors for the three sound types.

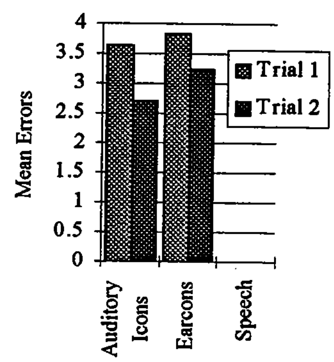

Figure 6: Mean errors by sound type over both trials.

that this interaction between explanation and trial was significant ($F1, 14 = 6.85, p = 0.020$).

Figure 5 shows a noticeable difference in mean errors (errors for both trials) between the three sound types. Subjects made more errors associating the earcons than the auditory icons, while speech showed a complete absence of errors. This difference in errors between the three sound types was significant ($F2, 28 = 154.93, p < 0.001$). An analysis of the earcons and auditory icons only, indicated that the lack of errors for speech was responsible for the significant difference in errors. The difference in mean error scores between the earcons and auditory icons was not significant ($F1, 14 = 1.85, p > 0.05$).

The difference in errors for all sound types between the two trials can be seen in Figure 6. As expected, speech showed a consistent absence of errors in the two trials. The earcons and auditory icons both showed a decrease in errors on Trial 2. Although the decrease in errors was greater for the auditory icons on trial 2, this difference was not found to be significant ($F2, 28 = 2.33, p > 0.05$).

5 Discussion

The results reveal that association time and errors both decreased after the explanation of cue design. Only the improvement in association accuracy, however, was found to be statistically significant. This indicates that knowledge of the auditory cue design can improve the accuracy of cue recognition while not necessarily improving the speed of that recognition. The lack of a significant reduction in the speed of association time may result from the subjects conscious use of the explanation in an attempt to map the sound cue onto its meaning. It is possible, therefore, that knowledge of cue design coupled with repeated exposures could result in a reduction of processing time and therefore an improvement in the association times of those cues. This may present possibilities for further testing.

The cue most accurately associated with its meaning, after speech, were the auditory icons. The earcons showed the most errors in association. This result supports the findings of Jones and Furner [9] who also found that auditory icons were more accurately associated than earcons. These results may result from the more direct metaphorical mapping of the auditory icons. The less intuitive, symbolically mapped, earcons proved to be more difficult for subjects to associate with their meanings. In both experiments relatively small numbers of cues were tested. Thus, the results may be affected by the particular sounds chosen, rather than the audio cue design method. A larger sample of sounds may provide a more accurate picture of the communicative ability of the design types. A significant improvement was recorded for both time and accuracy of association for the sound types tested. This improvement, after only two exposures, emphasises the ease of learning of both sound types. This ease of learning has obvious implications for the use of these audio cues by novice users. It can be seen that in such an application, the accuracy of use of the cues can be improved through providing the users with knowledge of the cue design.

As expected, the errors for speech remained consistently low for both trials. This can be attributed to the subjects' knowledge of and familiarity with this method of communication. Thus, if disregarding the more complex aspects of speech and looking at it as simply a process of associating meanings with abstract sounds, it appears that experienced users of the auditory icons and earcons could achieve similar performance levels. The significant decrease in both time and errors for auditory icons and earcons after only two exposures, would seem to support this possibility. This interpretation of the results may, therefore, have implications for the use of audio cues in the computer interface. This may imply a possible increase in the level of performance, in both speed and accuracy, after training and experience. Applications requiring speed and accuracy of audio stimulus recognition may possibly benefit from these or similar types of audio cues.

Conclusions

The results indicate that knowledge of the design of audio cues can improve the effectiveness of those cues. Providing subjects with knowledge of the auditory cue design has been shown to improve the accuracy of cue recognition for both auditory icons and earcons. Further study may indicate whether this knowledge is beneficial for audio cues in general and whether providing design knowledge is beneficial as part of the formal training of cue users.

Acknowledgments

This work was undertaken at London Guildhall University. Thanks to Preeti Patel for her advice and to Janet Finlay, Alan Dix, and Mark Kirby for help in improving the readability of the report. I would like to thank Robin Iwanek for help with the statistical analysis of the data.

References

[1] Gaver, W. W. "Auditory Icons: Using Sound in Computer Interfaces." *Human-Comp. Interaction* **2(2)** (1986): 167–177.

[2] Blattner, M. M., D. A. Sumikawa, and R. M. Greenberg. "Earcons and Icons: Their Structure and Common Design Principles." *Human-Computer Interaction* **4(1)** (1989): 11–44.

[3] Gaver, W. W. "The SonicFinder: An Interface That Uses Auditory Icons." *Human-Com. Interaction* **4(1)** (1989): 67–94.

[4] Brewster, S. A. "Providing a Model for the Use of Sound in User Interfaces." University of York Technical report YCS 169, 1992.

[5] Brewster, S. A., P. C. Wright, and A. D. N. Edwards. "An Evaluation of Earcons for Use in Auditory Human Computer Interfaces." *Proce. of INTERCHI '93* (1993): 24–29.

[6] Barfield, W., C. Rosenberg, and G. Levasseur. "The Use of Icons, Earcons, and Commands in the Design of an Online Hierarchical Menu." *IEEE Trans. on Prof. Comm.* **34(2)** (1991): 101–109.

[7] Gaver, W. W., and R. B. Smith. "Auditory Icons in Large Scale Collaborative Environments." *Proc. of INTERACT '90* (1990): 735–740.

[8] Gaver, W. W., R. B. Smith, and T. O'Shea. "Effective Sounds in Complex Systems: The ARKola Simulation." *Proc. of CHI '91* (1991): 85–90.

[9] Jones, S. D., and S. M. Furner. "The Construction of Audio Icons and Information Cues for Human Computer Dialogues." *Proc. of Contemporary Ergonomics* (1989): 436–444.