Experiments in Using Structured Musical Sound, Synthesised Speech and Environmental Stimuli to Communicate Information: Is there a case for Integration and Synergy?

Dimitrios Rigas[1], David Memery[1] and Hongnia Yu[2]

[1]School of Informatics, University of Bradford, Bradford BD7 1DP, United Kingdom
Tel.: +44 (0) 1274 235131 Fax.: +44 (0) 1274 233920 E-Mail: D.Rigas@bradford.ac.uk

[2]School of Computing and Engineering, University of Exeter, Exeter EX4 4QF,
Tel.: 01392 264685 Fax.: 01392 217965 E-Mail: H.Yu@exeter.ac.uk

ABSTRACT

This paper describes three sets of experiments in auditory information processing using structured musical sound, synthesised speech and environmental stimuli or special effects. The first experiment examines auditory information processing of small sequences of rhythmic musical tones. The second experiment examines auditory information processing of some environmental sounds. The third experiment examines auditory information processing when sound and synthesised speech are simultaneously presented. The results of this investigation aim to help multimedia and user interface developers to design auditory messages that incorporate structured musical stimuli, synthesised speech and environmental sounds or special effects. These type of auditory messages can complement visual displays or communicate information on their own in auditory interfaces. The results of these initial experiments indicate a prima facie case for integrating various types of auditory stimuli in one single message, The simultaneous presentation of sound and speech was easily recognised and interpreted. The paper concludes with some initial practical guidelines for designers and a discussion on further work.

1. INTRODUCTION

Auditory feedback can be generally divided into synthesised speech, environmental sounds (or auditory icons) and structured musical sounds (or earcons). Musical sound is a rich medium containing numerous structures introduced by musicians over many years of human evolution. Also, given that we live in an age where multimedia systems are fully capable of producing musical sounds relatively easily and effortlessly, the use of structured musical stimuli in interfaces is currently at a relatively low level. The auditory channel, as a whole, has been neglected in the development of user-interfaces, possibly because there is very little known about how humans understand and process auditory stimuli. It is not intuitively obvious how to use musical structures in interface design. Current user interfaces focus heavily on visual interaction. The consequence of this is that user interfaces have become more and more visually crowded as the user's needs to interact with the computer increase. Thus there is a need to investigate design techniques of auditory messages that combine different types of auditory stimuli.

2. AUDIO AS A METAPHOR

Earcons are short musical sounds that aim to communicate information. Earcons are classified in one element, inherited and transformed [1]. More guidelines for the design of earcons have been introduced [2, 3, 4]. Structured musical stimuli and earcons has also been used to communicate graphical information to visually impaired users [3,5,6]. Musical sound and speech has also been used to communicate information in a word processor system [7]. The usability of a graphics package has also been improved by the introduction of earcons [8]. Sound also has been used to communicate the contents of a complex software engineering database [9,10]. The execution of a sorting algorithm has also been communicated successfully.

Software engineering environments have also taken advantage from the use of sound. The Logomedia environment provided users with auditory feedback during program creation, execution and review [12]. InfoSound is another system, which enables users to create, store, and associate various types of auditory stimuli to application events [13]. The sound composition system of the InfoSound enables sound files to be created, stored and played. A sound storage system and a sound generation system were also provided. The IC* project was based on InfoSound. This project offered an environment under which sophisticated software systems could be developed. A telephone network service simulation was developed using InfoSound. In this system, speech and tone sequences communicated functions such as telephone ringing, dial tone, touch-tone dialling, busy signal, receiver hung-up, and receiver pick up. 'Audio Windows' is another example of a system that demonstrates the use of spatial sound and gesture in teleconferencing.

Environmental sounds (or auditory icons) is another form of auditory stimuli that can be used to communicate information. SonicFinder was the first user interface to introduce environmental sounds. These sounds communicated interface events in a way that each sound implied the sound being performed. SonicFinder also provided visual stimuli [15,16]. Recognition of environmental sounds (e.g., tearing paper, hammering, or walking) was also investigated in other experiments [17]. Subjects successfully matched sounds with their sources when the sound derived from one source. It was observed that subjects interpreted the sources of sounds instead of the pitch, timbre and other qualities [18]. Sound sources have also been used to communicate environmental information [19]. For instance, these types of sounds could communicate physical events (e.g., a bottle breaks and smashes or bounces when dropped on the floor), events in space (e.g., an ambulance's siren approaching), dynamic changes (e.g., overflow can be detected when liquid is poured into a glass), abnormal structures (e.g., the sound of a faulty car engine is usually different to the sound of a non-faulty engine), and invisible changes (e.g., a hollow space in a wall can be detected by tapping blocks).

3. EMPIRICAL INVESTIGATION

This investigation involves experiments in which short rhythmic sequences of notes and rhythmic tunes, some environmental and electronic sounds, and the simultaneous use of speech and musical stimuli were investigated. The goal of these initial experiments was to obtain an initial view of some of the capabilities and limitations of using musical stimuli, environmental stimuli or a combination of those forms of auditory stimuli with speech.

The general method of the experiments involved pre-experimental interviews, training, the actual experiment, and post experimental interviews. Subjects were requested to answer a questionnaire on musical knowledge. On completion of the questionnaire and introductory remarks, subjects were given 5 minutes training with the particular auditory design that the experiment aimed to test. During the experiment, auditory messages were presented in a random order. Experiments with musical and environmental stimuli were performed with the same group of subjects. A different group of subjects was used for the final experiment, which speech and musical stimuli were communicated simultaneously.

4. MUSICAL STIMULI

The first experiment involved the introduction of a musical metaphor that incorporated rising pitch notes. All short sequences of notes started rising at the middle C. These rhythmic sequences were rising in pitch in a way that a single note communicated value 1, two rising notes communicated value 2, three rising notes communicated value 3 and so on. These short sequences of notes were communicated to 25 subjects using the musical voices of piano and horn as they were produced from the multiple timbre synthesiser of the sound card. The concept of this rising pitch metaphor is illustrated below:

N1	(communicated value 1)
N1 N2	(communicated value 2)
N1 N2 N3	(communicated value 3)
N1 N2 N3 N4	(communicated value 4)

On overall, successful recognition rates were 99%. There were 198 correct recognitions of the short sequences out of 200 trials. These results demonstrate that the rising pitch metaphor, with the help of design consistency and to a certain extent counting the rhythmic notes, was a suitable auditory design for communicating small values.

In a further experiment with 32 subjects, recognition of four distinctive rhythms was tested. Subjects were presented with the auditory stimuli once and 1 minute was given to all subjects in order to answer the relevant question. Each answering sheet had 16 questions in total. On overall, 489 out of 512 questions about auditory messages presented were answered correctly. More specifically, 21 subjects (or 66% of the sample) answered correctly all 16 questions in the answering sheet, 7 subjects answered correctly 15 questions, 2 subjects answered correctly 14 questions, 1 subject answered correctly 12 questions, and 1 subject answered correctly 8 questions. The first rhythm was recognised correctly by 96.8% of the sample (or 124 answers out of 128 questions). The second rhythm was recognised correctly by 97.6% of the sample (or 125 correct answers out of 128 questions). The third rhythm was recognised correctly by 91.4% of the sample (or 117 correct answers out of 128 questions). The fourth rhythm was recognised correctly by 94.5% of the sample (or 121 correct answers out of 128 questions).

5. ENVIRONMENTAL STIMULI

Another experiment testing electronic sounds and sound effects recorded from environmental sounds were tested with 25 subjects. These sounds were:

1. Closing and shutting an iron door.

2. A pulse tone.
3. A sound similar to a conventional doorbell.
4. Electronic tones (peep beep peep beep).

The sounds of 'closing and shutting an iron door' and the 'electronic tones' was successfully recalled and remembered by 96% of the sample. The sounds of the 'pulse tone' and the 'sound similar to a conventional doorbell' were successfully recalled and remembered by 100% of the sampled. Post-experimental interviews with all subjects indicated that subjects easily recognised and remembered all environmental sounds because of the simplicity of the design and metaphorical nature.

6. SPEECH AND MUSICAL STIMULI

This experiments aims to test the simultaneous use of synthesised speech and structured musical stimuli. The auditory messages presented consisted of 640 synthesised speech messages communicating numerical values, 640 synthesised speech message communicating types of stock in words (e.g., radio), and either rhythm 1 or rhythm 2. The rhythms were taken from the rhythms tested in the experiment described in section 4. These rhythmic tunes were communicated simultaneously with the synthesised speech messages. On overall, 73% of the speech messages communicating the numerical values were correctly recognised, 89% of the speech messages communicating type of stock in words were also recognised successfully, and 93% of the rhythms were recognised successfully. These results indicate that the simultaneous use of synthesised speech and musical stimuli did not confuse the recognition rates of subjects.

7. DISCUSSION OF RESULTS

This work indicates that there is a prima facie case for utilising sound and for integrating different types of auditory stimuli. Subjects, with no special musical knowledge or ability, were in a position to extract information and meaningfully relate short sequences of rising pitch to numerical values. This could potentially prove to be a useful interface mechanism, particularly when is combined with synthesised speech. The experiment with the simultaneous use of speech and musical stimuli demonstrated that subjects were easily in a position to recognise the integrated auditory messages. This integrated type of communication metaphor could potentially help in many different interface circumstances or multimedia systems. For example, multimedia designers could potentially take advantage by an audio-visual feedback in their designs. Users with special needs, such as the visually impaired, could also benefit by integrating different types of auditory stimuli. For instance, auditory user interfaces could offer a promising alternative for blind computer users who currently have considerable difficulty in using computer systems with their emphasis on visual interfaces, which could also involve speech and environmental sounds.

A combination of musical parameters and structures also appears to be useful, as indicated in the use of a tunes and synthesised speech to communicate more information. This kind of integration was found to be useful as a communication metaphor.

8. GUIDELINES FOR DESIGNERS

These results indicate that speech complemented with sound can communicate successfully information in multimedia systems. A part of the overall message (information) is communicated using speech and another part using sound. This is particularly useful for user interface or multimedia designers who wish to use this part of media. The experimental results and observations suggest the following guidelines for auditory user interface or multimedia designers:

- Structured musical stimuli should be low base (avoid low frequency notes).
- Speech messages should be short, sharp, and clear.
- Provide mechanisms for user feedback or confirmation before the user interface proceeds to any further actions.
- Utilise different types of voices according to the musical sounds.
- Design the message in a way that the user's attention is attracted.
- Keep the auditory design simple.
- Provide mechanism for creating a user perceptual context by producing a user interaction model or by providing visual stimuli that creates a perceptual context for the auditory stimuli.
- Encourage semantic coding by allowing users to reason with the communicated auditory stimuli.
- Provide or repetition mechanism of the auditory message.

9. CONCLUSSION AND FUTURE WORK

Results of this study indicated that speech and sound could be successfully combined to communicate information. The simultaneous use of speech and sound did not confuse subjects who on overall successfully perceived both types of stimuli.

It is believed that auditory messages could utilise different types of auditory stimuli to communicate information. For example, these types of stimuli could include the use of musical sound and speech but also environmental sounds and special sound effects. A further set of experiments is under way to investigate how successfully these different types of auditory stimuli could be better integrated in one communication metaphor. This approach of integration could have particular application to a wide range of applications, which range from audio-visual or multimedia systems to telephone or mobile telephony interfaces.

10. REFERENCES

[1] Blattner, M., Sumikawa, D. A., and Greenberg, R.M.(1989). Earcons and icons: Their structure and common design principles. Human-Computer Interaction, 4:11-44. Lawrence Erlbaum Associates, Inc.A.B. Author, Title of Book, New York: IEEE Press, 1986, ch. 6, pp. 23-35.

[2] Brewster, S. A.(1994). Providing a structured method for integrating non-speech audio into human-computer interfaces. PhD thesis, University of York, England, UK.

[3] Rigas, D. I. (1996). Guidelines for Auditory Interface Design: An Empirical Investigation. PhD thesis, Loughborough University, Leicestershire, UK.

[4] Rigas, D. I. and Alty, J. L. (1998). How can multimedia designers utilise timbre? In H. Johnson, L. Nigay, and C. Roast, editors, Proceedings of HCI'98: People and computers XIII, pages 273-286, Sheffield, UK. Springer and British Computer Society.

[5] Rigas, D. I. and Alty, J. L. (1997). The use of music in a graphical interface for the visually impaired. In INTERACT-97, International conference on Human-Computer Interaction, pages 228-235, Sydney, Australia. Chapman and Hall.

[6] Alty, J. L. and Rigas, D. I. (1998). Communicating graphical information to blind users using music: The role of context. In CHI-98, Human Factors in Computing Systems, pages 574-581, Los Angeles, USA. ACM Press.

[7] Edwards, A. D. N. (1989). Soundtrack: An auditory interface for blind users. Human Computer Interaction, 4(1):45-66.

[8] Brewster, S. A. (1998). Using earcons to improve the usability of a graphics package. In H. Johnson, L. Nigay, and C. Roast, editors, Proceedings of HCI'98: People and computers XIII, pages 287-302, Sheffield, UK. Springer and British Computer Society.

[9] Rigas, D. I. (1993). A graphical browsing tool for the PCTE OMS. Master's thesis, University of Wales, Aberystwyth. MPhil.

[10] Rigas, D. I. and Alty, J. L. and Long, F. W. (1997). Can music support interfaces to complex databases? In EUROMICRO-97, New Frontiers of Information Technology, pages 78-84, Budapest, Hungary. IEEE, Computer Society.

[11] Rigas, D. I. and Alty, J. L. (1998). Using sound to communicate program execution. In Proceedings of the 24th EUROMICRO Conference, volume 2, pages 625-632, Vasteras, Sweden. IEEE, Computer Society.

[12] Digiano, C. J., Baecker, R. M. and Owen R. N. (1992). LogoMedia: A sound-enhanced programming environment for monitoring program behaviour. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel, and T. White, editors, INTERCHI'93, pages 301-302, Amsterdam, ACM Press, Addison-Wesley.

[13] Sonnenwald, D. H., Gopinaht B., Haberman G. O., Keese W. M., and Myers, J. S. (1990). InfoSound: An audio aid to program comprehension. In Proceedings of the twenty-third Hawaii International Conference on System Sciences, volume 11, pages 541-546.

[14] M. Cohen and L. F. Ludwig. Multidimensional audio window management. Man-Machine Studies, 34:319-336, 1991.

[15] Gaver, W. (1986). Auditory Icons: Using sound in computer interfaces. Human-Computer Interaction, 2(2):167-177.

[16] Gaver, W. (1989). The SonicFinder: An interface that uses auditory icons. Human-Computer Interaction, 4(1):67-94. Lawrence Erlbaum Associates, Inc.

[17] N. J. Vanderveer. Ecological acoustics: Human perception of environmental sounds. PhD thesis, Dissertation Abstracts International, 40/09B, 4543, 1979.

[18] E. A. Bjork. The perceived quality of natural sounds. Acoustica, 57(3):185-188, 1985.

[19] S. J. Mountford and W. Gaver. Talking and listening to computers. The art of human-computer interface design, pages 319-314, 1990.