

Audio Minimization: Applying 3D Audio Techniques to Multi-Stream Audio Interfaces

Yolanda Vazquez-Alvarez, Stephen Brewster
Department of Computing Science
University of Glasgow, Glasgow G12 8QQ, UK
{yolanda, stephen}@dcs.gla.ac.uk

ABSTRACT

Audio-driven eyes-free interactions in which simultaneous audio streams are employed can overload the user. We propose the use of an audio minimization technique in a 3D audio interface that will allow the user to minimize streams while focusing on a different interaction. We first introduce the concept of audio minimization and then describe the design and evaluation of a study set out to investigate the requirements for audio minimization.

Author Keywords

3D audio interfaces, audio minimization, divided attention.

ACM Classification Keywords

H.5.1. Multimedia Information Systems: Audio input/output; H.5.2. User Interfaces: Interaction styles.

INTRODUCTION

Current cable TV interfaces deal with the issue of presenting concurrent visual streams by minimizing the TV image when the user interacts with the television menu to change channels or just browse what is available in the different channels. In the same way, in a rich auditory interface we need to be able to minimize audio streams when we are busy and need to focus on something else (e.g. talking to someone, crossing the road). We also need to minimize the current sounds (using spatial audio and distance attenuation) to be able to interact with the auditory menus controlling our user interface. We believe audio minimization, as with minimization in visual systems, could act as an important component in any audio interface.

Audio minimization can be achieved by, 1) reducing the loudness of an audio stream (similar to reducing the size of the TV image on the TV screen), and 2) moving the perceived location of the audio stream from a central position (similar to moving the reduced TV image to the side of the TV screen). Once such minimization has occurred, a second audio stream can be played at normal loudness with a perceived location directly in front of the listener (for example, the menu to control the interface).

However, using minimization as a strategy for alternating focus between audio streams will affect the attention demand required from the users. Several studies have looked into how well listeners are able to understand the content of multiple, simultaneous sources (a situation requiring

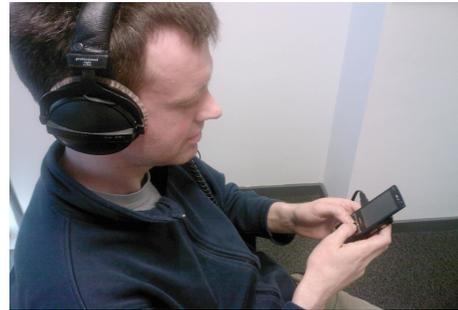


Figure 1. Spatial audio minimization study setup.

divided attention, see [1,2]). However, it is unclear what the most useful implementation of audio minimization might be: What loudness differential is most appropriate? Is audio spatialization effective enough to support audio minimization? How can such a technique be successfully used during a realistic application setting?

Previous research has shown that in divided-attention tasks, listeners are better at reporting the content of the louder source [3]. Marentakis and Brewster [4] also showed that a loudness cue was useful in assisting divided attention in an audio display when audio sources were played simultaneously, especially in mobile settings. However, using minimization as a strategy for alternating focus between the audio streams will affect the attention demand required from the users. Ihlefeld and Shinn-Cunningham [5] have recently looked at performance in divided-attention tasks in which subjects had to report content from two sources. Results from this study showed that in cases where two independent sources of information are presented simultaneously to a listener, spatial acoustic cues can help a listener identify the content of multiple competing sources. Also, listeners appeared to use attention to spatial location and other features to modulate the salience of competing sound sources. However, the speech stimuli used in this work, i.e. two set phrases with varying contents, was far from the kind of audio sources faced by real users in everyday applications.

SPATIAL AUDIO MINIMIZATION STUDY

We are conducting a study that will act as a baseline to investigate the requirements for audio minimization, namely limits of cognitive load, user acceptability and the extent to

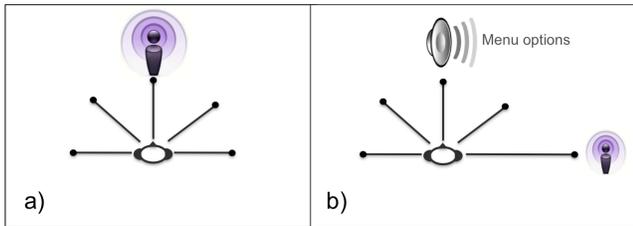


Figure 2. (a) Shows a single continuous audio stream. The black filled circles represent the different azimuth locations at which the audio streams can be located, placed 1 m away from the listener. (b) Shows the stream moved to the background - from the front to the right - and the perceived distance of the source increased, as a user interacts with a foreground spatialized audio menu.

which simple spatialization will support a simultaneous streaming strategy as part of a 3D audio interface on a mobile device. The spatial audio interface will run on a Nokia N95 8GB¹ using the HRTFs and the JAVA JSR-234 Advanced Multimedia Supplements API² to position the audio sources. The audio will be streamed over a pair of DT770 PRO – 250 OHM Beyerdynamic headphones (see Figure 1).

To investigate the limits of cognitive load when using minimization, participants will be asked to carry out tasks while listening to a podcast, i.e. streamed source. These tasks will include: (a) finding the next track title, (b) checking appointments for the day and (c) finding the current time. To perform these tasks, participants will use a baseline non-spatialized audio menu system located at the origin (0° azimuth) and at a distance of 1 m in the frontal horizontal plane so we can control for minimization effects alone.

The experiment will consist of four different conditions each consisting of two single point sources, one streamed and one user-activated, controlled by button presses on the phone: (1) *Concurrent*: The podcast is playing while the subject carries out the tasks. (2) *Interrupted*: The podcast is paused while the subject carries out the three tasks and then resumed after the task is complete. (3) *Fixed spatial minimization*: The podcast is fixed to the minimization location (90° azimuth) from the start and for the whole duration of the condition. (4) *User-activated Spatial minimization*: The podcast is located at the origin (0° azimuth) and is minimized only when the subject starts each task.

The streamed source will be a story selected from a BBC Radio 4 podcast from the programme ‘From Our Own Correspondent’. These stories are approximately 3 minutes in length. Based on our previous evaluation results of the 3D audio localisation capabilities of the Nokia N95 8GB [6], the minimization effect will be created by moving the streamed audio source to the right hand-side (90° azimuth),

as shown on Figure 2. This specific location showed less variation in the location perception by listeners. In addition, the level of the streamed audio source will be attenuated by approx. -10 dB by moving the source to the right hand-side (-3 dB drop in intensity) and doubling the perceived distance of the source. Shinn-Cunningham *et al.* showed that -10 dB intensity improved information recall out of a range from 0 dB to -40 dB, and avoids energetic masking when performing a divided attention task in which the audio sources are spatially separated. In order to measure intelligibility from the divided task, we will ask subjects to report on information from the tasks described earlier and the streamed podcast as in Ihlefeld and Shinn-Cunningham’s study. Effectiveness and usability will be computed from the total time taken to complete the tasks, and mental workload will be calculated based on results from the NASA-TLX forms completed by subjects after each condition.

CONCLUSION

3D audio allows us to segregate between multiple and simultaneous audio sources. However, strategies to interact with the audio sources to avoid cognitive overload and maintain intelligibility are needed. We propose the minimization of sound sources in a 3D auditory interface, in a simple and coherent way, which aims at improving the user attention on the interaction they are focused on while still allowing background monitoring of the original source. We believe the results of this work will allow better usability and effectiveness results when minimization strategies are used in the interactions compared to non-minimized implementations. Audio minimization will provide users with an interface that enables them to switch their attention in complex soundscapes.

REFERENCES

1. Yost, W. A., Dye, J., Raymond H., and Sheft, S. A simulated “cocktail party” with up to three sound sources. *Percept. Psychophysics*, vol. 58 (1996), 1026-1036.
2. D. S. Brungart, Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust. Soc. Am.*, 109, (2001) 1101-9.
3. Shinn-Cunningham, B.G. and Ihlefeld, A. Selective and divided attention: extracting information from simultaneous sound sources. *In Proc. ICAD 04*, (2004).
4. Marentakis, G. and Brewster, S. A. A Comparison of Feedback Cues for Enhancing Pointing Efficiency in Interaction with Spatial Audio Displays. *In Proc. MobileHCI 2005*, ACM Press (2005), 55-62.
5. Ihlefeld, A., and Shinn-Cunningham, B. G. Spatial release from energetic and informational masking in a divided speech identification task. *J. Acoust. Soc. Am.* 123, (2008) 4380-4392.
6. Vazquez-Alvarez, Y., Brewster, S. Investigating Background & Foreground Interactions Using Spatial Audio Cues. *Ext. Abstracts CHI 2000*, ACM Press (2009), 3823-3828.

¹ www.nseries.com/index.html#l=products,n95_8gb

² <http://theoreticlabs.com/dev/api/jsr-234/javax/microedition/amms/package-summary.html>