SPATIAL AUDITORY DISPLAYS A STUDY ON THE USE OF VIRTUAL AUDIO ENVIRONMENTS AS INTERFACES FOR USERS WITH VISUAL DISABILITIES

Christopher Frauenberger, Veronika Putz, Robert Höldrich

Institute of Electronic Music and Acoustics
University of Music and Dramatic Arts, Graz, Austria
Inffeldgasse 10/3, A-8010 Graz
{frauenberger|putz|hoeldrich}@iem.at

ABSTRACT

This paper presents the work on a prototype spatial auditory display. Using high-definition audio rendering a sample application was presented to a mixed group of users with visual disabilities and normal sighted users. The evaluation of the prototype provided insights into how effective spatial presentation of sound can be in terms of human-computer interaction (HCI). It showed that typical applications with the most common interaction tasks like menus, text input and dialogs can be presented very effectively using spatial audio. It also revealed that there is no significant difference in effectiveness between normal sighted and visually impaired users. We believe that spatial auditory displays are capable to provide the visually impaired and blind access to modern information technologies in a more efficient way than common technologies and that they will be inevitable for multimodal displays in future applications.

1. INTRODUCTION

The most common mode for human-computer interaction, the visual mode, is dominating user interfaces, but exploiting other modes of interaction becomes more important. This has various reasons, firstly, the content of interfaces is growing and becoming more complex. Secondly, in more and more applications the visual mode is restricted by form factors, the mobility of the user or simply by being occupied for other tasks. Thirdly, the fact that computers play a more central role in our society nowadays, builds up the awareness that they must be available for all parts of society. People with visual disabilities have major disadvantages in accessing computers because of the lack of efficient non-visual user interfaces.

The paper presented shows the research work conducted on the use of spatialised audio in human-computer interaction. It includes the building of a prototype presenting a sample application and its evaluation with a mixed group of visually impaired and normal sighted participants.

The following Sections will briefly describe the motivation for this work and provide information about some previous work. Section 2 describes in detail the built prototype including the sample application and its implementation. Subsequently, Section 3 describes the conducted evaluation process covering the tasks to be completed and the whole test setup. In Section 4 the results of the test are presented and the most important findings are pointed out. Finally, Section 5 draws some conclusions and looks on future research.

1.1. Motivation

To develop audio interfaces can improve the usability of human-computer interaction for a wide range of users and may be applicable on very different systems.

Visually impaired and blind people are not capable of using computers as they come out of the box. Every customary PC comes with a screen device which is of no use for this user group. The state of the art access for visually impaired people are screen-readers and braille-lines, both available only for significant additional costs to raise for the target group to have only inferior access to modern information technology. It is our social responsibility not to exclude such a big user group (approx. 190 million people with visual disabilities in the world [1]) from the developments in modern information technology by providing them adequate access to it.

Mobile computing is being considered as one of the key applications in present and future information technology. Devices like cellular phones, personal digital assistants (PDAs) and notebooks with wireless connection capabilities for information transfer are already widely used. Wearable computing or devices providing ambient intelligence are the very next logical step, connecting even more compact devices with the networks of the future. More complex user interfaces need to be presented with devices unsuitable for the use of conventional graphical displays.

Due to these facts new modalities must be introduced to HCI and for several compelling reasons audio is a good candidate. The most obvious advantage is that *"...audio display space is not wed to the disappearing resource of screen space"* [2]. Audio interfaces can be potentially very large without the need of large interface devices.

Another scientific field which may benefit from high performance auditory displays is the sonification of complex data. Spatialisation techniques have the potential to widen the possible display area for sonification and may therefore be very useful in depicting big multi-dimensional data sets.

1.2. Previous Work

Three dimensional, surrounding interfaces have the advantage of a much bigger display area than two dimensional screens or even one dimensional screenreaders. But they need to overcome the problem of concurring information and its interference [3, 4].

The segregation of sound sources in a virtual 3D environment is a necessary prerequisite of presenting information parallel to the user without confusion. Research has shown that sound source discrimination is a complex process. From the physiological point of view segregation depends on localisation cues like monaural spectrum changes or binaural sound wave differences [5]. But as stated in [6] sound source segregation is not only a question of localising the sources, it is heavily dependent on psychoacoustic effects. The content of the presented audio streams is crucial for the user to be able to segregate them. This has been shown for non-speech sounds [7], earcons [8] and speech sources [6]. Other psychoacoustic effects like informational masking [9] can additionally influence the presentation.

Spatial audio was used for auditory displays as soon as the simulation of sound fields became possible in real-time [10]. The interactive nature of user interfaced required “online” authoring of acoustic scenes so that pre-recorded audio was not satisfactory. With the improvement of audio rendering methods and the increasing knowledge about psychoacoustic effects virtual environments were able to present more complex information [11, 3].

However, the use of high definition audio rendering in human-computer interaction is still rare because little is known about how to design scenes effectively. The major problems identified were the users mental load [3] in combination with disorientation [11] and insufficient accuracy and naturalism in acoustic simulation. The presented work tries to address these issues with combining usability engineering methods with sophisticated real-time sound-field recreation techniques. Key idea is to consider semantic information of the user interface during the depiction process to find working acoustic representation for interaction tasks [12]. The presented work should provide the basis for further investigation on this field.

2. THE SYSTEM

The following Sections describe in detail the prototype system and the sample application to present.

2.1. The Application

The application to be presented in the auditory display was chosen to be a simple food market. It includes selecting food articles for a shopping list and an ordering process with payment and delivery information. The application was chosen because it combines important interaction tasks occurring in HCI, but is on the other hand abstract so that participants are most likely not very familiar with similar programs.

The application is compiled of the following interaction tasks: navigation in a structured menu, text input, exclusive and non-exclusive selection out of a list, confirmation of selections and confirmation of notifications or alerts. In a typical graphical user interfaces these tasks would be represented by menus, lists, text fields, buttons and check boxes. However, it is important to state that interaction tasks are mode independent. We were not looking for acoustic mappings for graphical user interface elements, but presenting the interaction task in the auditory domain.

Figure 1 illustrates the menu structure of the application. The menu is structured in 3 separate title menus with different sub-menus. Each sub menu leads to a corresponding dialog, where

Figure 1: *Menu structure of the shopping application.*

different actions can be performed by the user. In the dialog *New list* , personal data (name, street, village, date of delivery) can be entered. The *Search* menu naturally hosts a search function. With *Close* the user is asked for confirmation before leaving the program. Each of the *Goods* dialogs includes a list with 4 different products (e.g. milk products: milk, cheese, yoghurt, cream cheese) and each may be selected for the shopping list. In the *Delivery* dialog users can choose between picking up the items from the shop or if they should be delivered. In the *Payment* dialog, the user can pay cash or with credit card. Additionally, a few alerts and notifications were implemented, e.g. if one of the text fields was left empty, an error message appears. If the search function finds an item, there is a corresponding notification.

2.2. Auditory Representation

The auditory display is set in a virtual room with the dimensions 20m x 20m x 7m (width, length, height). The items presented are arranged in a semicircle in front of the user and their number is restricted to 6. Interaction with the application is performed through the participants head position and a conventional keyboard. The virtual room is divided into several active areas (a region of +/- 10° around each Icons position). If the user turns its head towards one of the icons, it becomes active as soon as he looks into one of the active areas. This means that any input via keyboard will apply to the active icon only. The total range of the display is adjustable

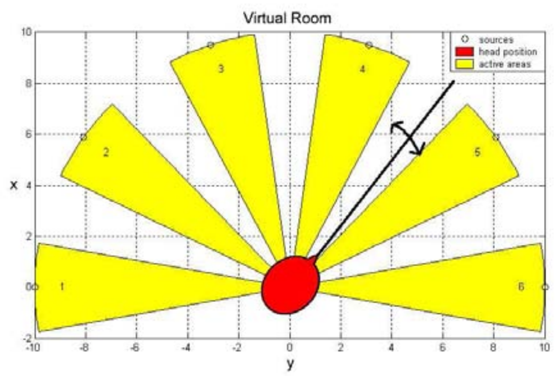

Figure 2: *Active regions in the presented auditory display.*

and after pretesting the system the overall region was scaled to +/- 60°. The spacebar is used for clicking (as substitute for the mouse which cannot be utilised by blind users). The F1 and F2 keys provide context sensitive help: F1 further information about the active area, F2 the context of the menu or dialog the user is in. The Escbutton leads back to the main menu. Dialogs have to be left with *OK* or *Cancel*.

To mimic the visual ability of focusing on a certain area of the monitor, an acoustic zoom function is implemented. This is done by weighting the input gain of the sources according to the relative position to the user. As weighting function a Gaussian window with variable width is used so that users may widen or narrowing their focus. However, the weighting function does not affect the room model used, the perception of room size does not shrink with the use of the zoom function.

The user interface items were modelled with auditory icons. Table 2.2 provides an overview over the used sounds. For some of the general sounds an "iconic language" was created. A vocabulary of sounds was created which were combined according to a syntax. For example, *OK* was always represented by a gong, but in dialogs was braided with the auditory icon of the dialog it was intended to confirm.

All Auditory Icons were played in a loop, but as all icons were natural sounds their gain envelope is fluctuating making localisation more difficult. As an additional help, to each source position a continuous musical tone was added. The pitch of the 6 musical tones was increasing from the left to the right (fground = 392, 523, 659, 784, 932, 1046 [Hz] or g1, c2, e2, g2, b2, and c3. Together, the musical tones form a C5 7- chord). The gain was kept lower than the gain of the icons. Each musical tone was a combination of 4 harmonic partial tones. Additional, a slight amplitude vibrato was added to make the tones easier to distinguish.

Menu	**Sub menu**	**Auditory Icon**
File		Crumpling of paper
	New List	new-born baby crying
	Search	dog panting
	Close	door closing
Goods		footsteps entering shop
	Fruit	twittering birds
	Bread	bread being eaten
	Milk	products cow mooing
	Beverages	canned drink being opened
Handling		cash register + money
	Delivery	truck starting
	Payment	sound of cash register
General Sounds	OK	gong sound
	Cancel	glass breaking
	Alert	police siren
	Main menu	trumpet fanfare
	Checkbox	piano sound with different pitches
	Radiobutton	violin sound with different pitches
	Selection	clicking of a photo camera
	Click	clicking sound
	Textfield	typewriter sound
	New menu	Swooshie

2.3. Implementation

The system was implemented on a customary personal computer with standard audio capabilities. Additionally, electrostatic headphones[1] and a headtracking system[2] were employed. All signal processing and application logic was implemented using the Pure Data[3] real-time music program.

[1]STAX SR-007

[2]Ascension Tech. Corp., Flock of Birds

3. EVALUATION

For evaluating the developed prototype a hearing test was set up. The subsequent Sections describe the setup of the test, the tasks to fulfil and which data has bee collected for analysis.

3.1. Test Setup

The test was performed by a group of 10 test participants. Among them were 4 participants who were blind or visually impaired. All participants got the same introduction before the test: The functionality of the program and the menu structure was presented, and the participants heard the auditory icons one time via loudspeakers. Afterwards, the participants had 15 min of free training with the application. In this time, all questions were answered and the participants were encouraged to try the help and the zooming functions. For equal conditions between participants with normal seeing ability and the visually impaired ones, no optical cue was presented on the monitor.

3.2. Tasks

The following tasks were asked to be completed from the participants:

Task 1: Opening the menu *File→New List* and entering the personal data in 4 text fields (name/street/village/date). Correct result: 4 text fields filled in.

Task 2: In *Order*, the terms of delivery and payment had to be chosen among a list of two possibilities for each. The meaning of the list elements head to be found via F1- function. For payment, a fictitious credit card number had to be entered. Correct result: 2 selections and 1 correct number entered in a text field.

Task 3: From the *Goods* sub menu participants had to choose 8 correct items out of 4 lists (fruit / milk products / beverages /bread). Correct result: 8 selections.

Task 4: Participants had to enter 2 specific items into the search function and, if found, add them to their shopping list Correct result: 2 search results added to list.

Task 5: Finally, participants were asked to quit the application. Correct result: 1 decision, leaving the application.

3.3. Collected Data

In a background questionnaire all participants were asked about their personal data, their education and about their background in using computers. After conducting the test the participants were interviewed to collect subjective data about the users' view of the system.

During the test bottom-line data was recorded. Firstly, regarding the task results meaning the filled in text in most cases. Secondly, all navigation and events were recorded. Head rotation was recorded every 100ms, any keyboard events including time stamps as they occurred.

Additionally, all sessions were taped on video and the test administrator attended the test with a pair of extra headphones and

[3]Pure Data by Miller Pukette, `http://www-crca.ucsd.edu/msp/`

took notes on anything remarkable uncovered by any logging facilities.

4. ANALYSIS

For analysis, the participants are divided into two groups: six participants with normal seeing ability (participant 5 to 10: group S for seeing) and four visually impaired or blind participants (participant 1 to 4: group B for blind). All members of group B need additional tools for working with a PC. Three of them use screen-readers (JAWS[4]) , two of them in combination with a Braille-lines. Only one member of this group is able to work with spectacles and at a very high resolution without additional tools. All members work in an office job with the aid of a PC and mainly use the MS Office package. Three of them have a lot of working experience and have absolved the ECDL (European Computer Driving License). The 4th participant has little working experience and is currently absolving the ECDL. Group S consists of five students of different fields and one employee of the institute. All of them have a lot of computer experience and spend several hours each day with a PC. They use different operating systems (Windows, Linux, Apple) and different programmes.

4.1. Bottom-line Data

As stated above all head movements and user events were recorded during the sessions. Figure 3 shows a typical result of this data. The azimuth angle on the x-axis cuts into active and inactive areas like illustrated in Figure 2. The y-axis shows the time elapsed during a task. In this case participant number 2 was completing task 2 (from minute 6 to 9 in his test session). The chart illustrates the navigation through the menus including all selection events and provides additional information of the zoom factor active at each time. Remarkable are the straight head movements from one menu to the other and the comparatively little focusing effort of the user when selecting an item.

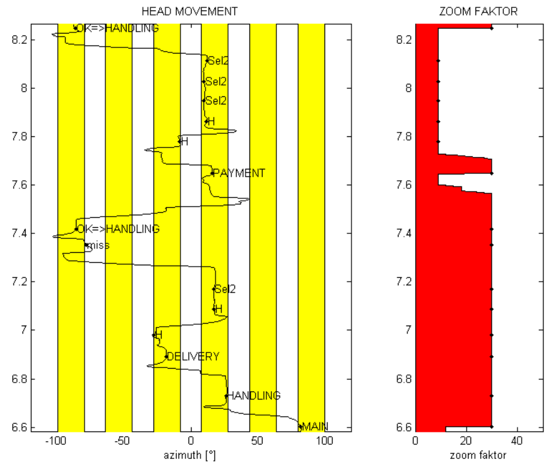

Figure 3: *Head movements and zoom factor over time of participant 2 in task 2.*

[4]©Freedom Scientific Inc.: `http://www.freedomscientific.com`

Participant 2 was at the time of conducting task 2 already very advanced in using the system. However, Figure 4 shows that in average the focusing effort of users resulted in a high ratio of hits. Four out of five attempts to select an item were successful. This shows that the quality of the directional sensation and the chosen size of the active areas were sufficient to present six items in the auditory display. Another remarkable result is that there was no

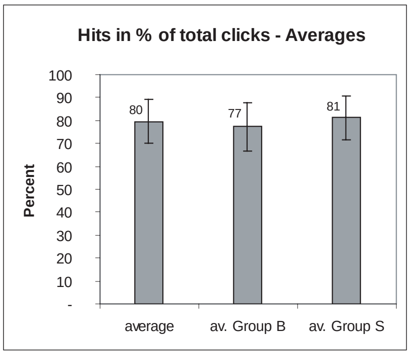

Figure 4: *Hits of active regions in percent.*

significant difference between group S and group B in terms of successful hits. This means that there could not be any difference determined in the accuracy of sound source localisation between blind and sighted users.

The equality in performance was also shown from the fact that the selection events per time showed no difference between the groups. Figure 5 shows that the hits and misses per minute are equally distributed among the groups. With an average of one successful selection every 15 seconds the performance in general is surprisingly near to graphical applications.

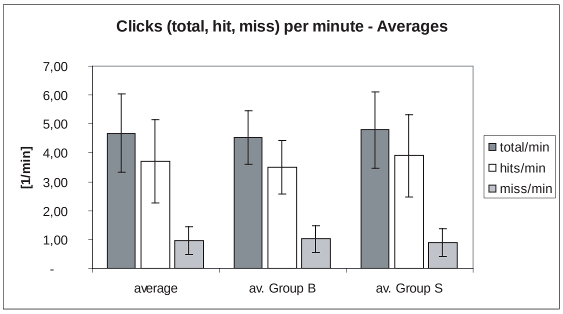

Figure 5: *Selection events per time.*

Another interesting result is the relation of the angel covered by head rotation and the number of events. Figure 6 shows the average angle covered for one event including help, zoom and selection events. The angle represents the average head movement of the user to reach a certain goal. Figure 7 illustrates the same relation as above, but considering only successful selection events. Both charts show that the needed head movement is in the range of the presented area of auditory icons. However, it is remarkable that half the way is wasted with either help events or misses. As in the charts above there is no significant difference identifiable between the group S and B. Although it seems that sighted users are

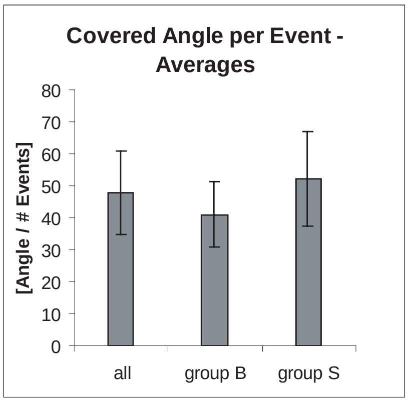

Figure 6: *Average angle covered for events.*

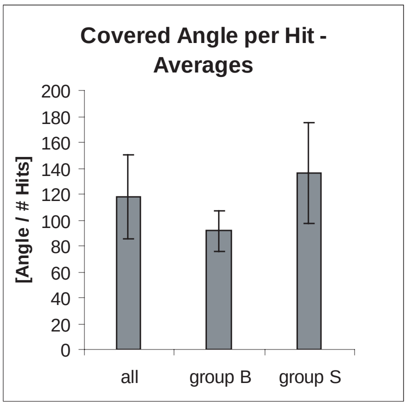

Figure 7: *Average angle covered for successful selection.*

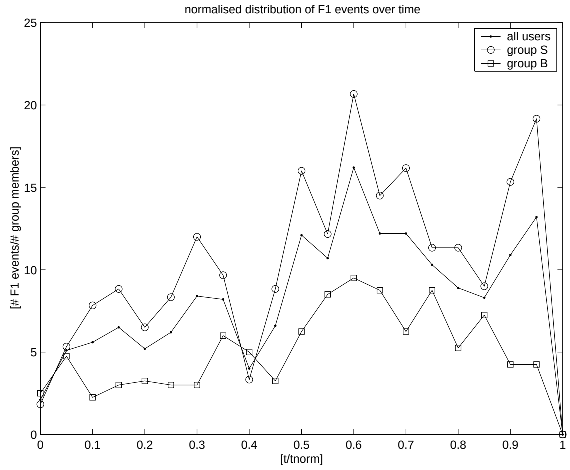

Figure 8: *Usage of F1 (help) function over the time.*

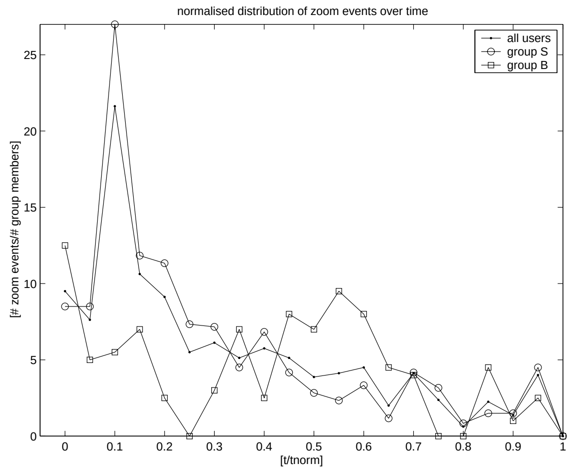

Figure 9: *Usage of zoom function over the time.*

slightly more effective in this figure, the difference lies within the variance of the data and has therefore no clear outcome.

The use of the F1 help function over the time of a user's session is illustrated in Figure 8. It does not show that people used it less to the end of the test as it was expected. However, it is remarkable that it seems that blind people used it less than their sighted colleagues.

This also is true for the use of the zoom function. It seems that sighted participants used the zoom function more often than the blind. In contrast to the usage of the F1 help function the zoom function was used less frequently the longer the test lasted. This points out that the zoom function may not be a very useful interaction or that at least the handling of it was not very valuable for the participants.

4.2. Subjective Evaluation

From the questionnaire after the test held in form of an informal interview the following key points could have been extracted:

From ten participants, two stated that they had problems with the spatial segregation in the beginning but improved with practise. Two of them explicitly said that the zoom-function was helpful, and two of them found the active areas around the sources to small, so that they often missed the target.

Seven participants said that they had a good orientation within the menu structure, and two of them explicitly reported a good mental image of the menus. Two participants found the F2-function a good idea, only one participant wanted more help-functions with regard to the menu structure. Three participants found the menu structure rather confusing, one had problems when entering a new room or menu.

The participant's description of the virtual room were very different. Five participants had a clear impression of a big room and could describe it afterwards, four participants did not have such an impression and one participant did not concentrate on the room at all.

The musical tones in the background were found helpful by four participants, two participants did not even notice them, only one found them too loud and one participant suggested that they should decrease in loudness when zooming out.

From ten participants, only one had problems remembering the meaning of the icons, because the time for learning them was too short. The other participants found them easy to remember, and three of them said that the F1 function was very helpful, if they had forgotten one of the icons.

Working with headphones was no problem for seven participants, two participants did not like working with headphones, and one said to prefer speakers.

5. CONCLUSIONS

The paper showed the building of a prototype auditory display using high-definition audio rendering. The evaluation revealed some remarkable findings above all that there is no evidence that there is any difference in performance between sighted users and users with visual disabilities.

Furthermore, it showed that relatively complex applications can be interfaced with spatial auditory displays easily and reach remarkable performance. Methodologies and approaches in this first prototype are nevertheless in their fledgling stages. For example, for practical work the selection with head rotation is definitely not a convenient way.

However, the development of "iconic languages", the use of high definition audio rendering and complex room models, zoom functionality and additional orientation cues are promising approaches towards a usable spatial auditory display.

In our opinion future research in this scientific field should focus on the development of a generic depiction process for user interfaces. This ranges from mode independent descriptions of interaction tasks within user interfaces to their acoustical representation in virtual environments [12]. Auditory displays must not be simple mappings of graphical user interfaces. Furthermore, a lot more must be known about the psychology of human hearing and information retrieving to be able to exploit the major advantage of the virtually huge display area of spatial auditory displays.

6. ACKNOWLEDGEMENTS

Many thanks to all who volunteered to participate in the hearing test. We also would like to thank the ISIS (Integration, Service, Information and Schooling for the blind) group with which we collaborate since two years. They helped us to acquire test participants and helped us understanding the needs of the people concerned. ISIS is a project of the Berufsförderungsinstitut Styria, Austria.

7. REFERENCES

[1] United Nations World Health Organization, "Prevention of blindness," `http://www.who.int/pbd/pbl/pblhome.htm`, 2002.

[2] A. Walker and S. Brewster, "Spatial audio in small screen device displays," *Personal Technologies*, vol. 4, no. 2, pp. 144–154, 2000.

[3] A. Walker, S. Brewster, and S. A. McGookin, "Diary in the sky: A spatial audio display for a mobile calendar," in *Proc. Interaction Homme-Machine – Human Computer Interaction (IHM-HCI)*, Lille, France, September 2001, pp. 531–540.

[4] T. Lokki et al., "Creating interactive virtual auditory environments," *IEEE Computer Graphics and Applications, special issue "Virtual Worlds, Real Sounds"*, vol. 22, no. 4, pp. 49–57, July/August 2002, Electronic publication: `http://www.computer.org/cga/`.

[5] V. Best, A. van Schaik, and S. Charlile, "Two point discrimination in auditory displays," in *Proc. Int. Conf. on Auditory Display (ICAD)*, Boston, MA, USA, July 2003, pp. 17–20.

[6] B. Arons, "A review of the cocktail party effect," *Journal of the American Voice I/O Society*, vol. 12, pp. 35–50, 1992.

[7] G. Lorho, J. Marila, and J. Hiipakka, "Feasability of multiple non-speech sounds presentation using headphones," in *Proc. Int. Conf. on Auditory Display (ICAD)*, July 2001, Espoo, Finland.

[8] D. K. McGookin and S. Brewster, "An investigation into the identification of concurrently presented earcons," in *Proc. Int. Conf. on Auditory Display (ICAD)*, Boston, MA, USA, July 2003, pp. 42–46.

[9] E. L. Oh and R. A. Lufti, "Informational masking by everyday sounds," *J. Acoust. Soc. of Am.*, vol. 6, no. 106, pp. 3521–3528, 1999.

[10] J. Mulder and E. Dooijes, "Spatial audio in graphical applications," Technical report CS-R9434, CWI-Amsterdam, 1991.

[11] C. Schmandt, "Audio hallway: A virtual acoustic environment for browsing," in *Proc. ACM Conf. Computer Human Interactions*, Los Angeles, California, USA, April 1998, pp. 163–170, ACM Press.

[12] C. Frauenberger, R. Höldrich, and A. de Campo, "A generic, semantically based design approach for spatial auditory computer displays," in *Proc. Int. Conf. on Auditory Display (ICAD)*, Sydney, Australia, July 2004.