Publication II

Is Audio Useful in Immersive Visualization?

M. Gröhn

Reprinted from *Stereoscopic Displays and Virtual Reality Systems IX, Proceedings of SPIE Vol. 4660*, pages 411–421, San Jose, Jan. 2002.

Is Audio Useful in Immersive Visualization?

Matti T. Gröhn[a]

[a]Helsinki University of Technology, Espoo, Finland

ABSTRACT

In this article I provide results from localization experiments in virtual environment. I define common tasks (orientation, localization, and navigation) in immersive visualization. The above mentioned tasks will be examined with user tests. Two localization experiments have been accomplished. In the first localization experiment the localization accuracy was significantly better ($p \ll 0.01$ in ANOVA) with loudspeaker reproduction than with headphone reproduction (nonindividualized HRTF's). The second experiment indicated, that localization accuracy is depending on signal ($p \ll 0.01$ in ANOVA).

Although the absolute lower limit for auditory localization accuracy in front is one degree for the azimuth, the reality is much worse. For example the screens and room reverberation deteriorate loudspeaker reproduction accuracy.

Current results suggest that at least in some tasks the audio is useful addition to immersive visualization tasks.

Keywords: Spatial audio, multimodal perception, virtual environments

1. INTRODUCTION

Immersive visualization generally takes place in virtual environments, which provide an integrated system of 3D auditory and 3D visual display. Some virtual environments can provide haptical interfaces, but those are not covered in my research. The usage of 3D sound in virtual environments is a quite well established area,[1] and it is currently mainly used to emphasize the sense of presence. This is normally achieved using recorded or simulated real world sounds to create virtual audio environment. The aim of my research is to find out new efficient ways to use audio in immersive visualization.

In the immersive scientific visualization the structures and objects might not have obvious up and down directions or any other orientation or wayfinding cues. For example, large molecules (such as in figure 1) or large multidimensional datasets could be very complex and after few rotations and movements it is easy to loose orientation or location of the origin. In complex immersive visualization tasks audio can be utilized as a navigational aid or as a data representation method (sonification).

In this paper I first take a look on related research (section 2). Next in the section 3 I represent the binaural cues, and some guidelines for selecting localizable signal. In section 4 I define the different tasks, which are common in immersive visualization. After task definitions I represent two localization experiments and their results in section 5. Finally in section 6 I draw some conclusions and set directions for the future work.

Further author information:
E-mail: Matti.Grohn@hut.fi, Telephone: +358 (0)9 451 5252, Address: Helsinki University of Technology, PL 5400, FIN-02015 HUT, Finland

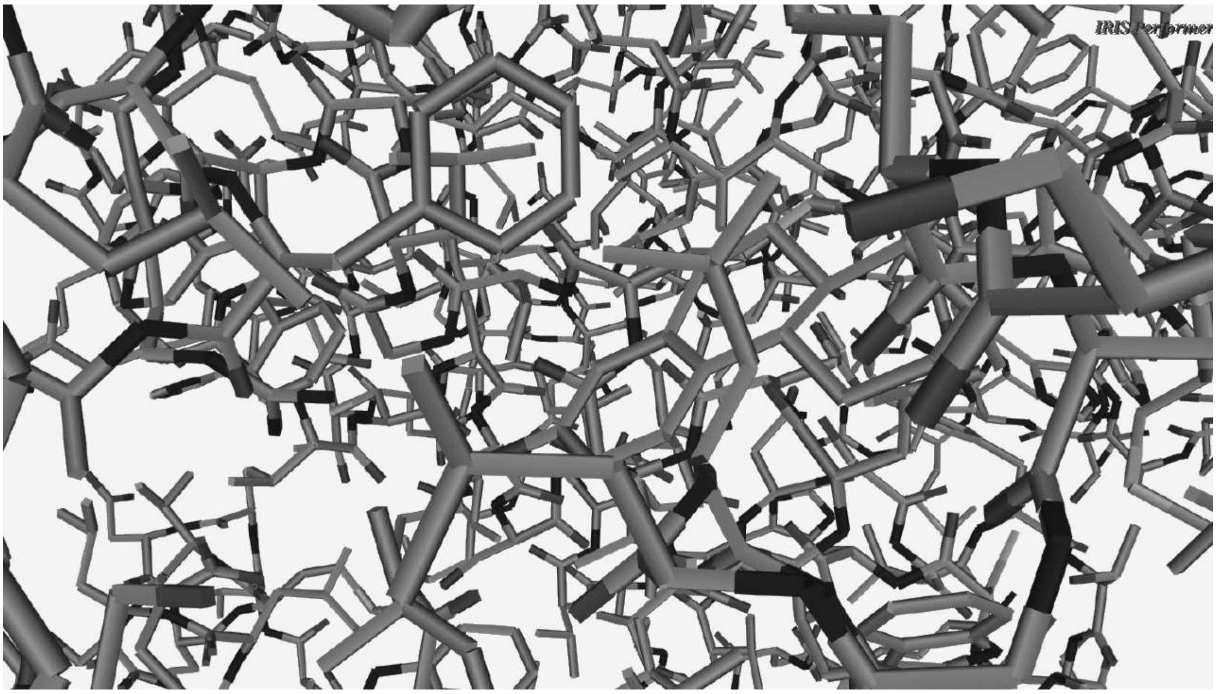

Figure 1: Typical complexity of a molecule model representing a chemical structure of a protein.

2. RELATED RESEARCH

2.1. Spatial sound

Spatial sound is a wide research area. There are many subareas like virtual acoustics[2] and spatial sound reproduction,[3] which are quite well explored. These and some other areas are well covered also by Begault.[1] There is a recent study on perceptual issues on spatial reproduction systems,[4] though it is concentrated on virtual home theater systems.

Auditory localization of 3D sound sources have been tested in several experiments.[5–8] Unfortunately most of them have been done using static sound sources. There has been few experiment covering effects of movement in perception[9–11] of auditory signal.

The cross-modal perception of auditory and visual stimuli is explored mostly with animals.[12] The inter-sensory interaction of visual and auditory stimuli have also been explored.[13] Typically most of these tests have been done in static test situations. So far, little research have been done in the area of the cognitive aspects of simultaneous visual and auditory stimuli in dynamic environments.[14] My research will in the future concentrate on this complex area of cross-modal interaction of dynamic auditory and visual stimuli in virtual environment.

2.2. Implementation of spatial audio in virtual environments

The are two main methods for reproduction of spatial audio: headphones and loudspeakers. For the headphones the most generally used method is head-related transfers functions (HRTF),[3,8] also cardioidic ears[15] has been used.

For the loudspeaker reproduction there are several alternatives like Ambisonics,[16,17] wave field synthesis,[18] and vector based amplitude panning (VBAP)[19]). In our own virtual environment called EVE[20] at Helsinki University of Technology we use VBAP. We have fourteen Genelec 1029A loudspeakers for audio reproduction. Alternatively it is possible to use headphones (Sennheiser 590A). Our audio reproduction system is more in detail described in articles.[21,22]

2.3. Other areas

One additional way to use audio in a virtual room is speech input. We have made some preliminary experiment with speech input.[23] Although it is interesting research area, it is not covered in this paper.

Presence (published by the MIT Press) had a special issue (Vol. 8 Issue 6 Dec. 1999) on Spatial Orientation and Wayfinding in Large-Scale Virtual Spaces. Most of the papers in this issue were concentrated on wayfinding in a realistic virtual worlds. None of them explored the possibility to use audio as a tool for orientation and wayfinding.

3. PSYCHOACOUSTICAL FACTORS

3.1. Binaural cues

There are two main binaural cues. They are derived from temporal and spectral differences of signals at ear canal. Temporal differences are called the interaural time differences (ITD) and spectral differences are called the interaural level differences (ILD). ITD is the primary cue in frequencies below 1.5 kHz and ILD is the primary cue above that threshold. The spatial hearing is exhaustively covered in[8]

The ILD is small at low frequencies, regardless of source position, because the dimensions of the head and pinna are small compared to the wavelengths of sound at frequencies below about 1500 Hz.[24]

3.2. How to choose an adequate signal

The signal is an adequate when then user can localize it, and the user is also able to differentiate it from other simultaneous signals. To utilize both main binaural cues (ITD and ILD), the signal should have enough energy in low frequency area (below 1.5 kHz) and in high frequencies. There also other factors in stimulus affecting the localization accuracy like spectral shape and temporal structure (see for example[24–26]).

It has been found[27] that frequencies near 6 kHz are important for elevation perception. (At least with VBAP). In addition auditory signals should not be annoying for users.

4. COMMON TASKS IN IMMERSIVE VISUALIZATION

In our article,[28] we defined four common tasks in immersive visualization: orientation, localization, navigation and sonification. I have focused my research to first three of them (defined in table 1). Sonification is in itself a large research area, and for example Sonification report[29] covers quite well this research field. According this report the sonification can shortly defined as use of nonspeech audio to convey information.

Task	Definition
Localization[30]	User ability to define direction and distance of the target
Orientation	User awareness about the front-back, up-down, and left-right directions.
Navigation	User ability to move from starting point to target

Table 1: Definition of tasks

4.1. Localization

In this article localization is defined as a task, in which user defines the direction and distance of the source (could be auditory, visual or combined).

In data analysis auditory beacons[31] or some other auditory stimuli are applied to localize the most interesting features of the data. For example, while a researcher is exploring a large protein, he can 'highlight' the most important amino acids with auditory beacons. In a dynamic representation it is important that the user is able to follow the location of the moving sound source.

4.2. Orientation

In this research orientation is defined as a task, in which user is aware about the front-back, up-down, and left-right directions.

The orientation can be represented in such a way, that each direction has its own characteristic timbre and the sound source is located in that direction. While user rotates the global geometry the sound sources indicating orientation move as well. Applying this method the user hears all the time which way at the moment is for example the original front-back direction. In informal tests (done in horizontal plane) the method has been successful.

4.3. Navigation

Navigation is a task which utilizes both localization and orientation information. In this research navigation is defined as a task, in which user goes from one specified position (starting point) to another specified position (target).

Typically in a complex visualization the visualized objects may occlude the target. If the target is presented using sound, it can be located even when it is not visible. For example while exploring large protein the user has a awareness of locations of the most important amino acids, and could easily move near them even when he doesn't at first see them behind the other chemical structures. Our first experiment[32] showed that navigation is possible with auditory cues.

5. LOCALIZATION EXPERIMENTS

5.1. First experiment

In the first experiment[30] we compared head-tracked binaural headphone reproduction (using HRTFs), direct loudspeaker reproduction, and vector based amplitude panning (VBAP) using loudspeakers.[19] Inside a real virtual room are many factors, that decrease the localization accuracy. For example, we didn't use individualized HRTF's, because they typically are not available. Instead, HRTF's were measured a from dummy head. On the other hand, the screens and room reverberation will deteriorate loudspeaker reproduction accuracy.

According to Djelani et al.[33] pointing is an appropriate method for localization experiments. In the first test the user could freely move inside the virtual environment (as they typically do while they are using some application).

To find out the localization accuracy, a listening test was conducted. The auditory stimulus was pink noise. The task of the subjects was to point to the direction of the perceived sound source using a wand (see picture 3). We had eight male test subjects. Each subject accomplished four different tasks, and each task had 17 different sound source locations. Locations have been played to subjects in randomized order to avoid learning effect. Each subject had 34 locations with HRTF reproduction, 15 locations where sound was reproduced using one loudspeaker (LS), and 19 locations with VBAP reproduction.

We measured the pointing accuracy using Ascension magnetic tracking device. We measured the azimuth, and elevation angle separately. The duration of finding each location was also stored.

According to Zahorik et al[34] there was no significant difference in localization performance between anechoic virtual environment and echoic virtual environments. In the experiments the anechoic environment was used because in our navigation experiment,[32] we found out, that the anechoic situation was most accurate in navigation tasks.

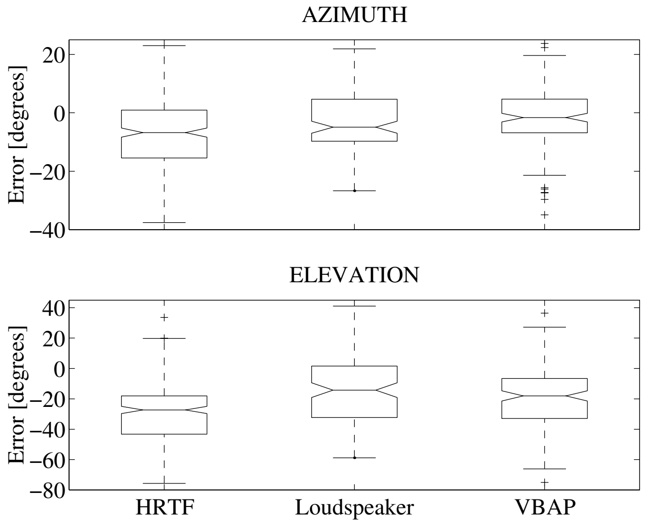

Figure 2. The azimuth and elevation errors for the first experiment for each reproduction system. The boxplot depicts the median and the 25%/75% percentiles.

	Average Azimuth	Average Elevation	Median Azimuth	Median Elevation
HRTF	10.6	31.7	8.3	27.4
LS	8.7	22.5	7.5	20.1
VBAP	8.9	23.0	5.6	18.9

Table 2. Average and median values of azimuth and elevation errors for each reproduction type in the first experiment.

5.2. Results of the first experiment

The azimuth localization accuracy was in average 9.7 degrees. With current reproduction methods, the perceived elevation accuracy was not so good, average error was 27.0 degrees. Especially with HRTF's the perceived sound source location was in average much higher, than given location. The average and median values of azimuth and elevations errors are shown in Table 2.

We use an analysis of variance (ANOVA) model for the analysis. HRTF reproduction was significantly worse both in azimuth ($p \ll 0.01$), and elevation ($p \ll 0.01$) accuracy compared with other reproduction methods. On the other hand there was no significant difference in accuracy between the direct loudspeaker reproduction, and VBAP. Boxplot of the azimuth and elevation localization accuracy is seen in figure 2*.

There was significant difference in localization time between the HRTF reproduction, and the direct loudspeaker reproduction ($p = 0.004$). The localization last in average longer with HRTF reproduction.

5.3. Second Experiment

In the second experiment I used only direct loudspeaker reproduction (LS) and vector based amplitude panning (VBAP), because emphasis in our environment is in loudspeaker reproduction.

Also in this second experiment I had eight male test subjects (six of them were the same as in the first experiment). Each subject accomplished four different tasks, and each task had this time 28 different sound

*In this first experiment equation for error is: error = real location - perceived location. In the second experiment it is: error = perceived location - real location

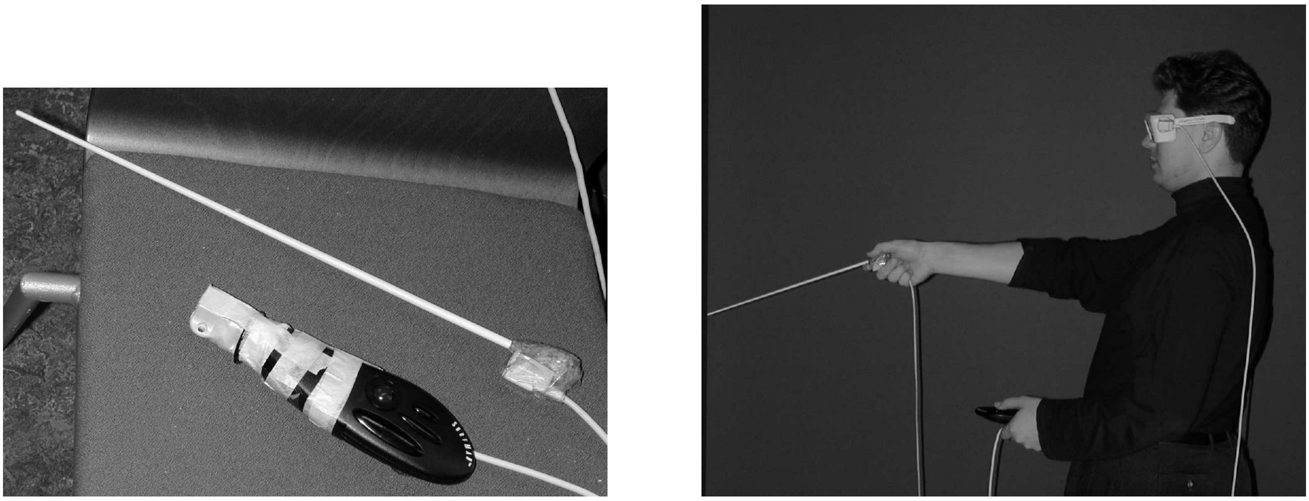

Figure 3: On the left tracked baton, and our wandlike device. On the right test subject is using these devices.

Reproduction	**Median Azimuth Error Both**	**Median Elevation Error Both**	**Median Azimuth Error LS**	**Median Elevation Error LS**	**Median Azimuth Error VBAP**	**Median Elevation Error VBAP**
Signal						
Pink Noise	6.1	16.2	5.0	11.2	7.3	25.5
Pink Floyd	7.4	13.9	5.8	11.0	11.4	18.3
Frog	7.8	17.8	4.0	13.1	14.3	28.1
Phone	6.9	18.0	5.8	15.0	13.6	23.9
All signals	6.9	15.9	5.3	12.1	11.9	23.4

Table 3: Median values of azimuth and elevation errors for reproduction types and signals in the second experiment.

source locations. Each subject had 56 LS locations and 56 VBAP locations. Also in this experiment locations have been played to subjects in randomized order to avoid learning effect.

In addition I compared four different auditory stimuli: pink noise, music (excerpt from The Wall by Pink Floyd), frog and phone ring. These signals have different kind of spectral continent as seen in figure 4. Pink noise covers the whole frequency area. Music has clear temporal and harmonic structure. Frog sound has silence (black part of the image) and most of its energy is under 2 kHz. Phone stimulus has lot of energy in high frequencies.

We use the same tracking device for the measurements than in the first experiment. In this second experiment the task of the subjects was to point to the direction of the perceived sound source using a baton (see picture 3). It was expected, that this method will provide more accurate results, than pointing with wand.

5.4. Results of the second experiment

The median azimuth localization error was 6.9 degrees. This approximately the same accuracy as achieved in the first experiment. The median elevation localization error was 15.9 degrees, which is less than in the first experiment. The median values of azimuth and elevations errors for reproduction types and signals are shown in Table 3.

The analysis of the variance (ANOVA) is again used for analysis. There were statistically significant difference ($p \ll 0.01$) between the reproduction methods. The VBAP is more inaccurate and produces more variance, which can bee seen in figure 5 and in table 3. The median value for direct loudspeaker reproduction azimuth

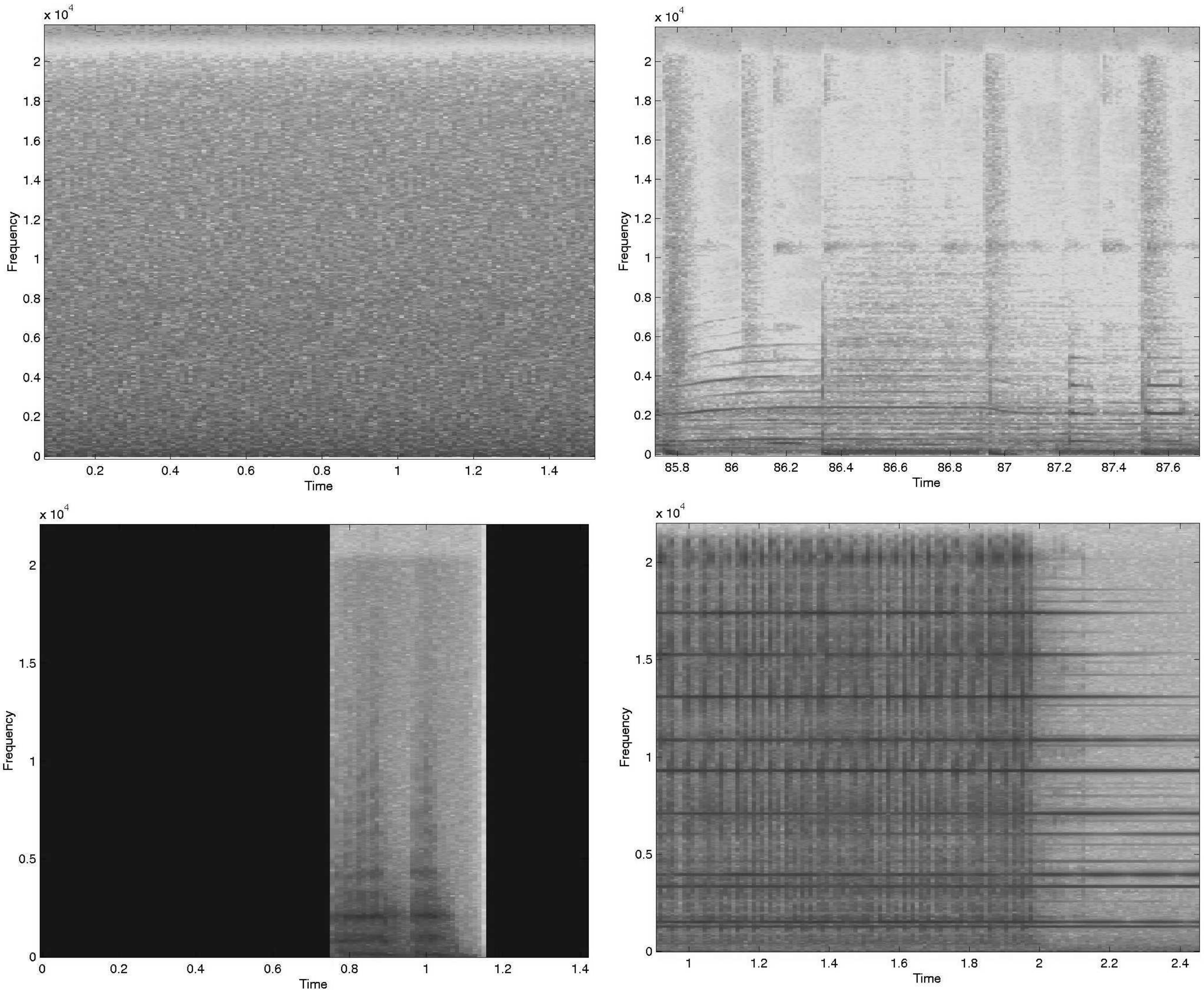

Figure 4. Spectrograms of all the four auditory stimuli. Upper left is pink noise, upper right is part of the Pink Floyd. Lower left is the frog and lower right is the phone.

error is 5.3 degrees and for VBAP 11.9 degrees. Also the difference between different signal was statistically significant ($p \ll 0.01$) in azimuth.

In the first experiment there were no difference between reproduction methods in elevation, but in this second experiment the difference was statistically significant ($p \ll 0.01$). The VBAP was more inaccurate, which is seen in figure 6. The median value for elevation error for LS is 12.1 degrees and for VBAP 23.4. In signal level analysis the phone was only signal, which no statistically significant difference between reproduction types in elevation accuracy.

Also the difference between the signals in elevation errors was statistically significant ($p \ll 0.01$). The median elevation error for music signal is 13.9, which is more than two degrees less than the second best median value for pink noise (16.2 degrees).

6. CONCLUSIONS AND FUTURE RESEARCH

According to Blauert[8] in optimal conditions the localization blur in azimuth is approximately one degree. The localization blur for elevation is more signal dependent and according to Blauert[8] it can variate from four

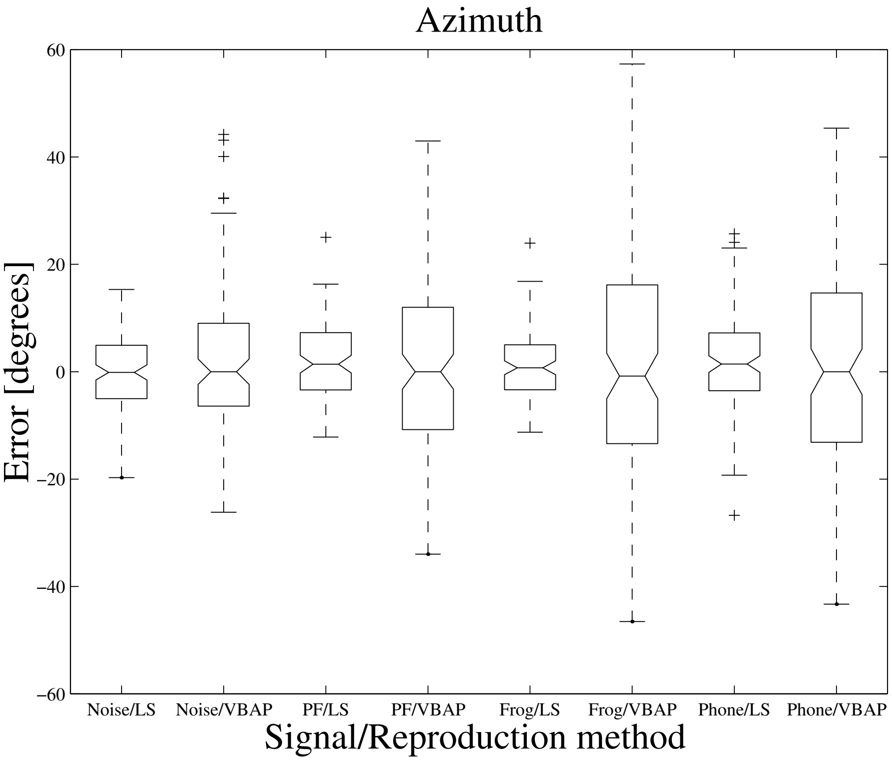

Figure 5. The azimuth error for each signal reproduction combination. The boxplot depicts the median and the 25%/75% percentiles.

degrees (for white noise) to seventeen degrees (continuous speech by unfamiliar person). Familiarity with the signal also plays a role in elevation perception.

In our environment there is more localization blur in azimuth, than in optimal conditions. The measured elevation accuracy is also blurred. Two main reasons for increased blurring are low-pass filtering of screens and room reverberation. Achieved localization accuracy is in the level, that it is reasonable to continue experiments with dynamic multimodal cases (see table 4). In addition I have planned to test situations where there are many simultaneous sound sources with the same or different movement directions.

Audio	**Visual**
dynamic	static
static	dynamic
dynamic	dynamic (same movement direction)
dynamic	dynamic (different movement direction)

Table 4: Different combinations of localization experiments

During the spring I will start the orientation experiments and continue navigation experiments. Our first navigation experiment[32] proved, that navigation is possible just using the auditory cues.

All the accomplished experiments supports my hypothesis, that audio is useful addition to immersive visualization.

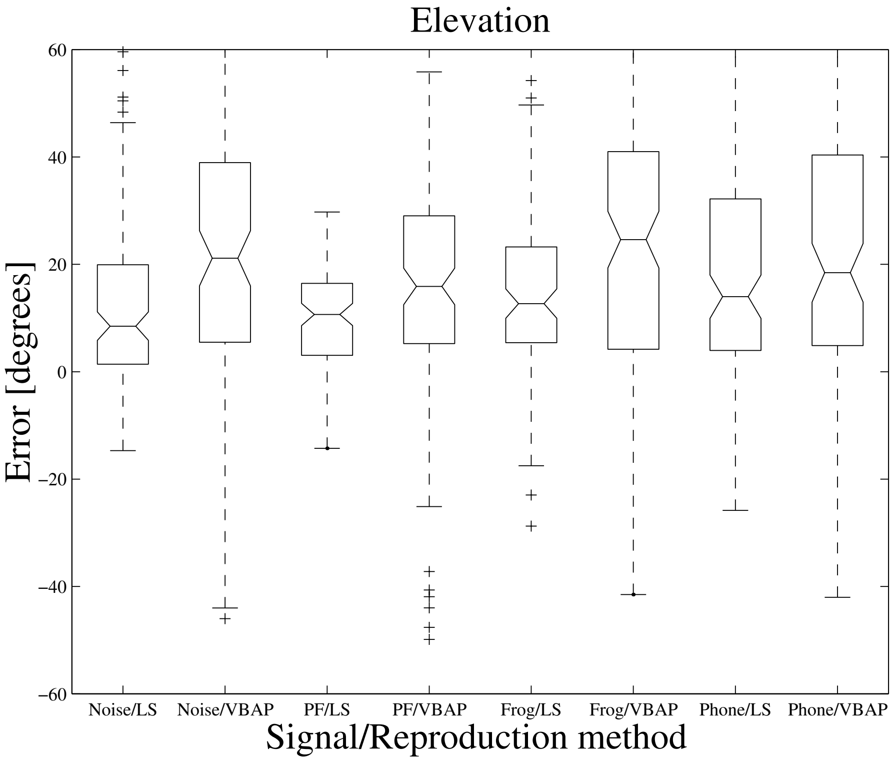

Figure 6. The elevation error for each signal reproduction combination. The boxplot depicts the median and the 25%/75% percentiles.

ACKNOWLEDGMENTS

I would like to thank Tapio Lokki and Lauri Savioja for their valuable advice during the test design and analysis. In addition I would like to thank Tommi Ilmonen for his work with our audio software and hardware

This work has been partly financed by the Helsinki Graduate School in Computer Science and Engineering (HeCSE, http://www.cs.helsinki.fi/hecse), HPY Research Foundation, and KAUTE Foundation.

REFERENCES

1. D. Begault, *3D Sound for Virtual Reality and Multimedia*, Academic Press, Cambridge, MA,, 1994.
2. L. Savioja, *Modeling Techniques for Virtual Acoustics.* PhD thesis, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, report TML-A3, 1999.
3. J. Huopaniemi, *Virtual acoustics and 3-D sound in multimedia signal processing.* PhD thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, report 53, 1999.
4. N. Zacharov, *Perceptual Studies on Spatial Sound Reprodution Systems.* PhD thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, report 57, 2000.
5. F. Wightman and D. Kistler, “Localization of virtual sound sources synthesized from model HRTFs,” in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA’91)*, (New Paltz, NY), 1991.
6. E. Wenzel, “Localization in virtual acoustic displays,” *Presence: Teleoperators and Virtual Environments* **1**(1), pp. 80–107, 1992.
7. E. Wenzel, M. Arruda, D. Kistler, and S. Foster, “Localization using non-individualized head-related transfer functions,” **94**, pp. 111–123, 1993.

8. J. Blauert, *Spatial Hearing, The psychophysics of human sound localization.*, The MIT Press, Cambridge, MA,, 1997.
9. D. Perrot and T. Strybel, "Some observations regarding motion without direction," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. Anderson, eds., pp. 275–294, Lawrence Erlbaum Associates Inc., 1997.
10. D. Grantham, "Auditory motion perception: Snapshots revisited," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. Anderson, eds., pp. 295–313, Lawrence Erlbaum Associates Inc., 1997.
11. K. Saberi and E. Hafter, "Experiments on auditory motion discrimination," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. Anderson, eds., pp. 315–327, Lawrence Erlbaum Associates Inc., 1997.
12. B. Stein and M. Meredith, *Merging the Senses,*, The MIT Press, Cambridge, MA, 1993.
13. R. Welch and D. Warren, *Intersensory interactions*, Wiley, New York, 1986.
14. D. Begault, "Auditory and non-auditory factors that potentially influence virtual acoustic imagery," in *Proc. AES 16th Int. Conf. on Spatial Sound Reproduction*, pp. 13–26, (Rovaniemi, Finland), April 10-12 1999.
15. T. Takala and J. Hahn, "Sound rendering," *Computer Graphics* **SIGGRAPH'92**(26), pp. 211–220, 1992.
16. M. Gerzon, "Periphony: Width-height sound reproduction," *Journal of the Audio Engineering Society* **21**(1/2), pp. 2–10, 1973.
17. D. Malham and A. Myatt, "3-d sound spatialization using ambisonics techniques," *Computer Music Journal* **19**(4), pp. 58–70, 1995.
18. A. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *Journal of the Acoustic Society of America* **93**, May 1993.
19. V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society* **45**, pp. 456–466, June 1997.
20. J. Jalkanen, *Building a spatially immersive display - HUTCAVE. Licenciate Thesis*, Helsinki University of Technology, Espoo, Finland, 2000.
21. J. Hiipakka, T. Ilmonen, T. Lokki, and L. Savioja, "Sound signal processing for a virtual room," in *Proc. X European Signal Processing Conference (EUSIPCO 2000)*, (Tampere, Finland), Sep 2000.
22. J. Hiipakka, T. Ilmonen, T. Lokki, M. Gröhn, and L. Savioja, "Implementation issues of 3D audio in a virtual room," in *Proc. SPIE*, **4297B**, (San Jose, California), Jan 2001.
23. M. Gröhn, M. Laakso, M. Mantere, and T. Takala, "3D visualization of building services in virtual environment," in *Proc. SPIE*, **4297B**, (San Jose, California), Jan 2001.
24. F. Wightman and D. Kistler, "Factors affecting the relative salience of sound localization cues," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. Anderson, eds., pp. 1–23, Lawrence Erlbaum Associates Inc., 1997.
25. R. Duda, "Elevation dependence of the interaural transfer function," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. Anderson, eds., pp. 49–75, Lawrence Erlbaum Associates Inc., 1997.
26. J. Middlebrooks, "Spectral shape cues for sound localization," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. Anderson, eds., pp. 77–97, Lawrence Erlbaum Associates Inc., 1997.
27. V. Pulkki, *Spatial Sound Generation and Perception by Amplitude Panning Techniques.* PhD thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, report 62, 2001.
28. M. Gröhn, T. Lokki, L. Savioja, and T. Takala, "Some aspects of role of audio in immersive visualization," in *Proc. SPIE*, **4302**, (San Jose, California), Jan 2001.
29. G. Kramer, B. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner, J. Neuhoff, R. Bargar, S. Barrass, J. Berger, G. Evreinov, M. Gröhn, S. Handel, H. Kaper, H. Levkowitz, S. Lodha, B. Shinn-Cunningham, M. Simoni, W. Tecumseh Fitch, and S. Tipei, *Sonification Report: Status of the Field and Research Agenda.*, ICAD, 1999.

30. M. Gröhn, T. Lokki, and L. Savioja, "Using binaural hearing for localization in multimodal virtual environments," in *Proc. 17th Int. Congr. Acoust. (ICA 2001)*, **IV**, (Rome, Italy), September 2001.

31. G. Kramer, *Auditory Display: Sonification, audification and auditory interfaces.*, Addison-Wesley, Reading, MA., 1994.

32. T. Lokki, M. Gröhn, L. Savioja, and T. Takala, "A case study of auditory navigation in virtual acoustic environments," in *Proc. ICAD 2000*, pp. 145–150, (Atlanta GA), Apr 2000.

33. T. Djelani, C. Pörschmann, J. Sahrhage, and J. Blauert, "An interactive virtual-environment generator for psychoacoustic research II: Collection of head-related impulse responses and evaluation of auditory localization," *ACUSTICA acta acustica* **86**, pp. 1046–1053, 2000.

34. P. Zahorik, D. Kistler, and F. Wightman, "Sound localization in varying virtual acoustic environments," in *Proc. ICAD'94*, pp. 179–186, (Santa Fe, NM), Nov 1994.