Some Aspects of Role of Audio in Immersive Visualization

Matti Gröhn, Tapio Lokki, Lauri Savioja, Tapio Takala
Telecommunications Software and Multimedia Laboratory
Department of Computer Science
Helsinki University of Technology
PO Box 5400, FIN-02015 HUT, FINLAND

ABSTRACT

We define four different tasks (orientation, localization, navigation and sonification), which are common in immersive visualization. Immersive visualization takes place in virtual environments, which provide an integrated system of 3D auditory and 3D visual display. The main objective of our research is to find out the best possible ways to use audio in different tasks. In the long run the goal is more efficient utilization of the spatial audio in immersive visualization application areas. Results of our first experiment have proven that navigation is possible using auditory cues.

Keywords: immersive visualization, virtual environment, spatial audio, multi-modal perception

1. BACKGROUND AND MOTIVATION

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 our research. The usage of 3D sound in virtual environments is a quite well established area,[1] and it is 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 our 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).

The main objective of our research is to find out the best possible ways to use audio in different tasks. In the long run the goal is more efficient utilization of the spatial audio in immersive visualization application areas. We have chosen the performance based approach for evaluation of usefulness of audio, and we run our experiments in our virtual room.[2] Due to a close co-operation with the Laboratory of Acoustics and Audio Signal Processing [*] our virtual room has state of the art 3D-audio system[3–5]

In this paper we first take a look on related research (section 2). In section 3 we define the four different tasks (orientation, localization, navigation and sonification), which are common in immersive visualization. After task definition we represent the variables for the analysis in section 4. In section 5 we represent our first experiment with results. Finally in section 6 we draw some conclusions and set directions for the future work.

Further author information:
Matti Gröhn: E-mail: Matti.Grohn@hut.fi
Tapio Lokki: E-mail: Tapio.Lokki@hut.fi
Lauri Savioja: E-mail: Lauri.Savioja@hut.fi
Tapio Takala: E-mail: Tapio.Takala@hut.fi

[*] http://www.acoustics.hut.fi/

Visual Data Exploration and Analysis VIII, Robert F. Erbacher, Philip C. Chen,
Jonathan C. Roberts, Craig M. Wittenbrink, Matti Gröhn, Editors,
Proceedings of SPIE Vol. 4302 (2001) © 2001 SPIE · 0277-786X/01/$15.00

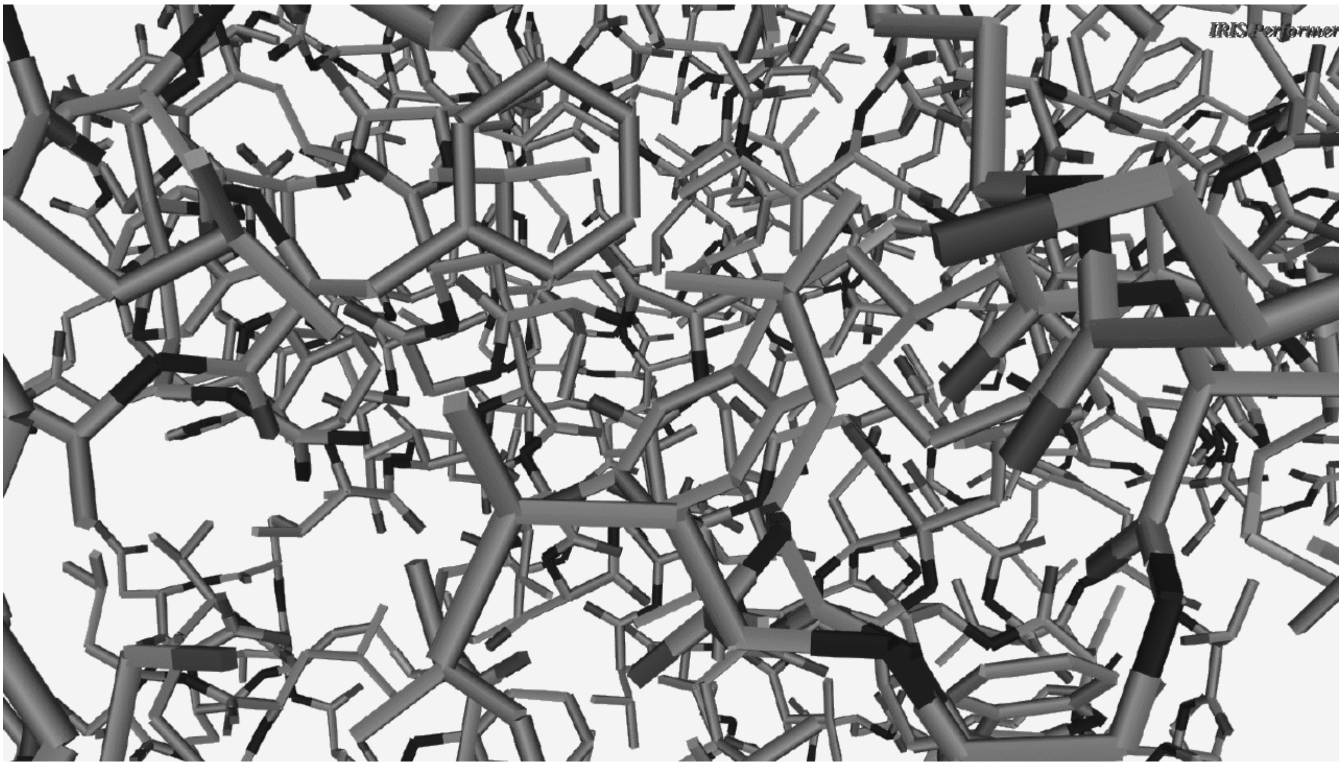

Figure 1. Typical complexity of a molecule model representing a chemical structure of a protein.

2. RELATED RESEARCH

2.1. Sonification

Sound has been used in data analysis for several years. This is called sonification. Sonification brings out different aspects of data than visualization. It can reduce the visual overload. Ears are better than eyes for finding time dependent changes and identifying periodic patterns. With multiple data streams some of the streams can be shown in visual display and some can be presented on auditory display.

Two basic features of auditory perception have been discovered that suggest sound can be effective for representing data in a variety of settings.

First, auditory perception is particularly sensitive to temporal characteristics, or changes in sounds over time. This points to a distinct advantage of auditory over visual displays. Fast-changing or transient data that might be blurred or completely missed by visual displays may be easily detectable in even a primitive, but well-designed auditory display.

Second, unlike visual perception, sound does not require the listener to be oriented in a particular direction. Auditory displays can therefore be used in situations where the eyes are already busy with another task.

First experiments in sonification were made as early as 1954 by Pollack and Ficks[6] and 1961 by Speeth.[7] Pollack and Ficks used eight variables and they reported that sound is a possible way to classify variables but that “extreme subdivision of each stimulus dimension does not appear warranted”. Speeth reported that auditory inspection can be carried out rapidly and auditory methods seem to be a promising addition to other means for interpreting seismograms. However, for two decades there was almost no reported research on this field.

Since the beginning of the 80’s the research has gone forward, and after the first International Conference of Auditory Display conference 1992,[8] the researchers have regularly done new experiments and shared the information. The ICAD organization[†] has published ’Sonification Report’.[9] This report was prepared at the request of National

---

[†] http://www.icad.org

Science Foundation and the purpose is to provide an overview of sonification research, including the current status of the field and a proposed research agenda.

2.2. Spatial sound

Spatial sound is a wide research area. There are many subareas like virtual acoustics[10] and spatial sound reproduction,[11] 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,[12] though it is concentrated on virtual home theater systems.

Auditory localization of 3D sound sources have been tested in several experiments.[13–16] Unfortunately most of them have been done using static sound sources. The cross-modal perception of auditory and visual stimuli is explored mostly with animals.[17] The intersensory interaction of visual and auditory stimuli have also been explored.[18] 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.[19] Our research is concentrated on this complex area of cross-modal interaction of dynamic auditory and visual stimuli in virtual environment.

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.[20] 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. DEFINITION OF THE TASKS

In our research we concentrate on role of the audio in four typical tasks (defined in Table 1) in immersive visualization: orientation, localization, navigation, and data analysis (sonification). We have done our first experiments with auditory navigation (described in section 5). They have confirmed us, that at least in a limited test environment the navigation is possible using auditory cues alone.

Task	**Definition**
Orientation	User awareness about the front-back, up-down, and left-right directions.
Localization	User ability to define direction and distance of the target
Navigation	User ability to move from starting point to target
Sonification	Use of nonspeech audio to convey information

Table 1. Definition of tasks

3.1. 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.

3.2. Localization

In this paper 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[8] 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.

3.3. Navigation

Navigation (also known as wayfinding) 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 (see section 5) showed that navigation is possible with auditory cues.

3.4. Sonification

Sonification is a large and complex research area. In the 'Sonification Report'[9] the sonification is defined as the use of nonspeech audio to convey information. More specifically, sonification is transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation.

In our research we concentrate on spatial sonification.[8,21] In spatial sonification 3D audio techniques are used to place the sound sources in their locations. Spatial sonification enables separation and localization of the most interesting or critical values during the data exploration. For example, the spatial sonification of distance information is applicable as an aid in a molecule docking task. The listening point could be put inside the molecule and the user will hear the critical distance information from accurate direction.

4. THE TEST SETTING

We have chosen the performance based approach for evaluation of usefulness of audio. The tasks described in section 3 are evaluated with several user tests. In these tests the subjects accomplish a well defined subset of these tasks. During the test we measure user performance data like speed and accuracy. In the navigation task the trajectory (or path) from starting point to target reveal important information about the differences in various test settings as seen in figure 5.

Additionally we collect subjective evaluations, which will make it possible to find the least annoying variable combinations.

4.1. Test variables

Many different variables should be taken into account, while defining the most useful way to use auditory information. In our research we have considered to explore the effects of 3D-panning method, visual display, interaction, visual cues, viewing and listening position, and timbre. It is not reasonable to test all these variables and all different tasks in one large test (If we just test the orientation, navigation and localization and use three different audio stimuli, the amount of different test combinations is still over four hundred different subtasks for each subject). It is much more convenient to start (as we have already done) with smaller subset of the variables. After evaluation the best variables will be used as fixed variables in next tests.

4.1.1. 3D-panning method

Spatial sound can be reproduced using either headphones or speakers. Headphone playback is considered optimal for reproducing spatial sound because it allows the greatest degree of control over the location of the spatial source.[19] Head related transfer functions (HRTF)[11] are the most common method for headphone reproduction.

Multichannel sound reproduction using multiple loudspeakers is a more convenient solution for spatial sound in virtual rooms. With multichannel reproduction we avoid the need for individualized HRTFs and head tracking. It is possible to reproduce sound signals naturally from correct directions. When the direction of the virtual sound source coincides with the direction of a real loudspeaker, the source direction is exact. When these directions do not coincide, different panning methods can be used. In our system we apply vector base amplitude panning (VBAP), which is a simple mathematical way to calculate the gain coefficients for the loudspeakers.[22] VBAP also allows an arbitrary loudspeaker placement which is good feature in virtual rooms where mirrors and projectors hinder the fixed loudspeakers positioning.

Another used multichannel spatial sound reproduction method is Ambisonics[23,24] which is suitable to create background soundscape because a recording and coding methods are available for Ambisonics. Due to it's known limitations (small "sweet spot") it is not suitable for our purposes.

4.1.2. Visual display

The stereo images can be displayed either on normal desktop monitor or on immersive virtual reality display. The latter can be divided into three distinct classes[25]: 1. Head-Mounted Displays (HMD), 2. Virtual Model Displays (VMD), and 3. Spatially Immersive Display (SID)

In our research we will accomplish most of the tests in spatially immersive display. For the reference purposes we will do some tests (mostly navigation) using just normal desktop monitor.

4.1.3. Interaction methods and devices

In our first experiment we used keyboard as an navigation device. In the immersive visualization the keyboard and mouse combination is not convenient solution. There are many alternative solutions for the interaction in virtual environments like data gloves, wands, and trackers.

In our virtual environment we will start the tests with our prototype wand (actually it is a tracked radiomouse seen in figure 2). We will compare these with results we have achieved in our first tests with keyboard control.

Figure 2. Prototype of custom wand.

4.1.4. Visual cues

We are interested in the real usage scenarios, where user has always visual and auditory information available. To get the reference values we start our tests using only auditory information (no visual cues).

After these we will add visual information. We use term supporting visual information for such visual stimuli, which is related to auditory stimuli. Non supporting visual cues are not related to auditory stimuli. Examples of no visual, supporting, and non-supporting situations are seen in figure 3.

Comparison of supporting and non-supporting visual is essential. It is interesting to see, if the user can locate the sound sources correctly if there is only non-supporting visual cues represented simultaneously.

The hypothesis is that the non-supporting visual cues will decrease the accuracy, and restrain the speed. This is due to a dominating nature of the sight.

Typically it is considered, that when visual and auditory cues conflict, sounds are localized to the position of the visual stimuli. This is known as the "ventriloquism effect".[17] However, at least one study[26] suggests that visual dominance is not unilateral across azimuth positions, and there exist positions where auditory information provides more accurate localization information

4.1.5. Viewing and listening position

It is natural to think that the viewing and listening points are the same. That is the normal situation in everyday life. We have a hypothesis, that in some situations, it will be useful to separate viewing and listening points (figure 4). That might provide additional information for example in molecule docking task described earlier in this article.

Separate viewing and listening position is analogous with the situation where sound recorder is using a remote microphone. We have preliminarily studied these in.[21]

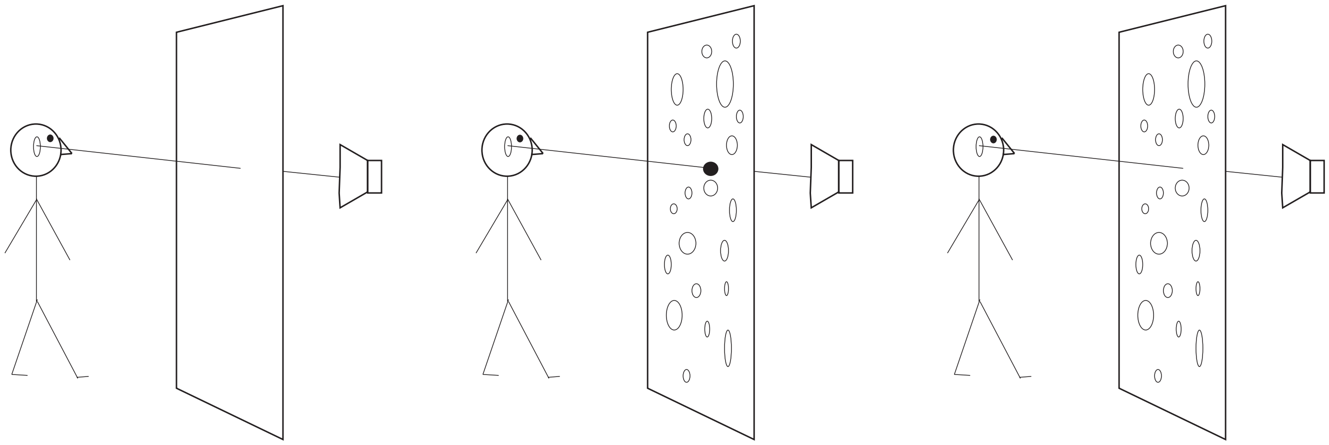

Figure 3. To the left there are no visual cues, in the middle there are visual and auditory stimuli in the same location (supporting), and to the right visual, and auditory stimuli are not related to each other (non-supporting).

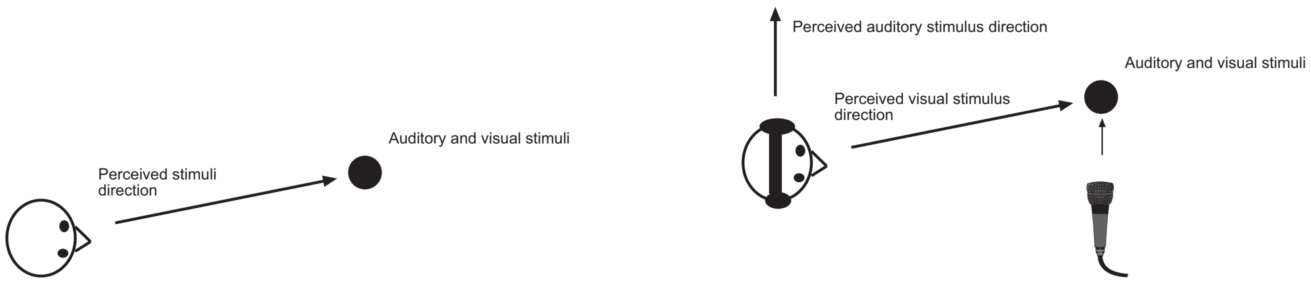

Figure 4. To the left is a normal situation where auditory and visual stimuli from the same source are perceived in the same direction. To the right the listening position is separated from the viewing position and the perceived directions of auditory and visual stimuli differ from each other.

Stimulus	**Panning method**	**Acoustic environments**
pink noise	ITD alone	direct sound
artificial flute	ITD + simple amplitude panning	direct sound + 6 early reflections
recorded anechoic guitar	ITD + minimum-phase HRTF (FIR 30 taps)	direct sound + 6 early reflections + reverberation (length about 1 second)

Table 2. The three tested factors.

5. FIRST EXPERIMENT

In this section we review results of an auditory navigation experiment, reported in more detail earlier in.[27] Auditory navigation tests have been done earlier e.g. by Loomis et al.[28] and Rutherford.[29] Our aim was to do the experiment in a dynamic system, in which perceived acoustics changes according to the movements of the subject.

In this experiment the task of the subjects was to find a sound source by moving and turning in a virtual space. Our purpose was to analyze the effect of various factors in the test setup. These factors were influence of the sound stimulus, the directional cues, and acoustics of the environment. We carried out a complete test set with three variables each having three different choices.

In this experiment we tested three different factors: stimulus, panning method, and influence of acoustic environment. Each factor contained three choices summarized in Table 2.

The interaural time difference (ITD), was included as an auditory cue to all tests. The second panning method included also a simple model for frequency independent interaural level difference (ILD). This method, also called cardioid method, was presented by Takala and Hahn.[30] The third panning method used minimum-phase head-related transfer function (HRTF) filters instead of simple ILD. Original HRTFs were measured from an artificial head.[31] They were approximated with 30 tap FIR filters designed by Huopaniemi.[11]

5.1. Results

Stimulus: Pink noise was clearly the best stimulus. Pink noise gave the minimum number of errors and was fastest, and it has also found to be easiest in subjective judgments. Guitar sound gave worst results, which was also the subjective opinion of the subjects.

Panning Methods: It is quite clearly shown that ITD alone is inferior for auditory navigation, because almost 30% of these cases failed. The best panning method was cardioid panning which gave clearly fastest results. Surprisingly, in terms of median times ITD and ITD+HRTFs were not statistically very different, although the error rate is much smaller with ITD+HRTFs.

Acoustic environment: Reverberation increased both the spent times and the error rate, which is an expected result. Direct and direct+reflections gave almost equal results both in the time spent and in and the error rate, which is an expected result. Direct and direct+reflections gave almost equal results both in the time spent and in the error rate.

The results of our experiment showed that navigation is possible with the auditory cues. The 27 subjects completed the 27 navigation tasks (all variable combinations). The results, which were also statistically validated, proved that noise is the best stimulus, reverberation complicates the navigation and simple models of spatial hearing give enough cues for auditory navigation.

Figure 5 shows all paths (27 subjects) for six different navigation tasks. The upper row displays the test cases with most errors (11 to 13 errors). In all these the stimulus was flute and the panning method ITD only. Due to the sine-wave like nature of the flute sound the ITD can be very confusing panning method. The subjects had problems to find correct direction to target area.

The three lower figures display three navigation tasks with no errors. In these cases the right direction to target area is found very well. (It is easy to see, that there have been few front-back confusions and some subjects have first headed away from the target area.) These tasks have also been completed much faster than three tasks with most errors (mean of median times 37 s. vs. 64 s.).

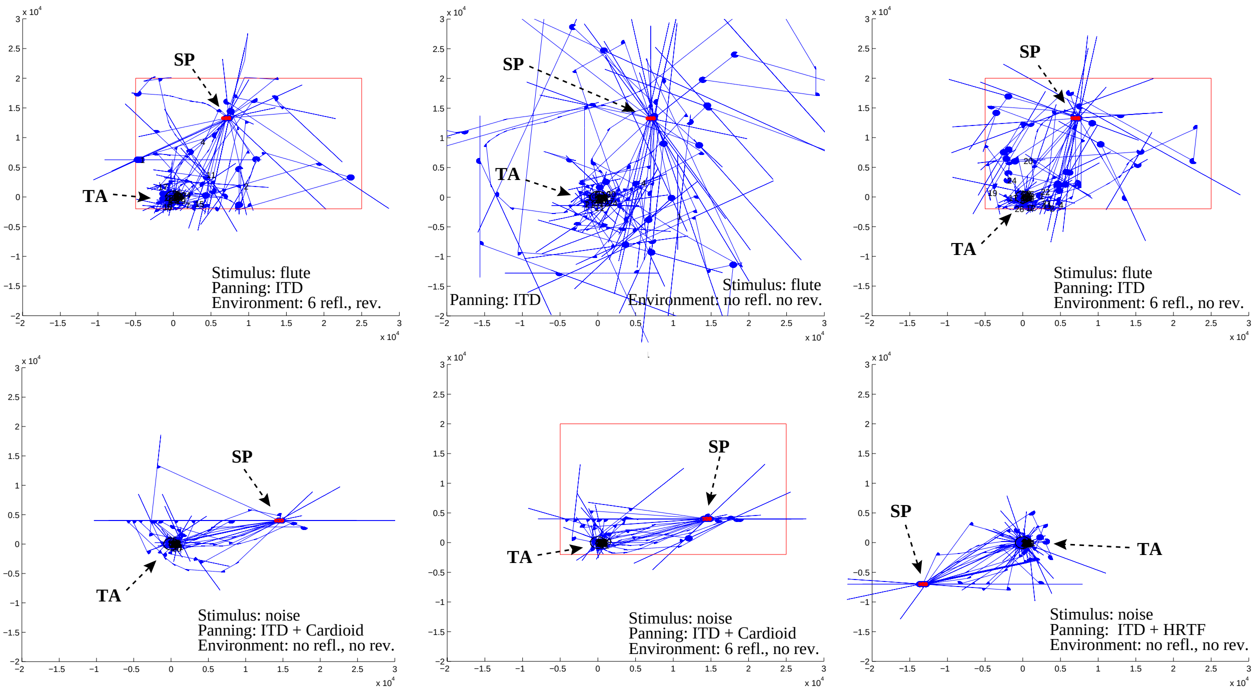

Figure 5. All paths (27 subjects) of six different navigation tasks. Boxes indicates cases where at least the early reflections were rendered. Abbreviation SP marks the starting point and TA the target area.

5.2. Discussion

In this first experiment the user interface was quite limited. The subjects could only turn their head or move forward and backward. These restricted movements enforced subjects to behave in same manner. First they panorated sound source in middle of the head and then moved forward or backward. That limited movement control might have affected the results of panning methods. The cardioid panning method gives the best front-back separation although the externalization is not as good as with HRTFs. A possible explanation is that the employed artificial head HRTFs were not suitable for all the test subjects.

6. CONCLUSIONS AND FUTURE WORK

In this paper we defined four different task in which the spatial sound can be utilized in immersive visualization (orientation, localization, navigation, and sonification). The test setting with multiple variables was presented.

Results of our first experiment have shown that proper auditory cues alone are enough for successful navigation (at least in limited test environment). Next step is to test cross-modal situations with simultaneous visual and auditory stimuli. In addition, there is lot of work to be done in testing the other tasks (orientation, localization, and sonification), and different combinations of test variables.

Our first step in this complex area has been successful, but there is lot of work ahead of us before we will reach our long term goal, which is more efficient utilization of the spatial audio in immersive visualization application areas.

7. ACKNOWLEDGMENTS

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

REFERENCES

1. D. Begault, *3D Sound for Virtual Reality and Multimedia*, Academic Press, Cambridge, MA,, 1994.
2. J. Jalkanen, *Building a spatially immersive display - HUTCAVE. Licenciate Thesis*, Helsinki University of Technology, Espoo, Finland, 2000.
3. L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *Journal of the Audio Engineering Society* **47**, pp. 675–705, Sept. 1999.
4. J. Hiipakka, T. Ilmonen, T. Lokki, and L. Savioja, "Sound signal processing for a virtual room," *Proc. X European Signal Processing Conference (EUSIPCO 2000)* , (Tampere, Finland), Sep 2000.
5. J. Hiipakka, T. Ilmonen, T. Lokki, M. Gröhn, and L. Savioja, "Implementation issues of 3d audio in a virtual room," *Proc. SPIE* **4297B**, (San Jose, California), Jan 2001.
6. I. Pollack and L. Fickst, "Information of elemantary multidimensional auditory displays," *The Journal of Acoustic Society of America.* **Vol 26 n:o 2**, Mar 1954.
7. S. Speeth, "Seismometer sound," *The Journal of Acoustic Society of America.* **Vol 33 n:o 7**, Jul 1961.
8. G. Kramer, *Auditory Display: Sonification, audification and auditory interfaces.*, Addison-Wesley, Reading, MA., 1994.
9. 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.
10. L. Savioja, *Modeling Techniques for Virtual Acoustics.* PhD thesis, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, report TML-A3, 1999.
11. 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.
12. 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.
13. 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.
14. E. Wenzel, "Localization in virtual acoustic displays," *Presence: Teleoperators and Virtual Environments* **1**(1), pp. 80–107, 1992.
15. E. Wenzel, M. Arruda, D. Kistler, and S. Foster, "Localization using non-individualized head-related transfer functions," **94**, pp. 111–123, 1993.
16. J. Blauert, *Spatial Hearing, The psychophysics of human sound localization.*, The MIT Press, Cambridge, MA,, 1997.
17. B. Stein and M. Meredith, *Merging the Senses,*, The MIT Press, Cambridge, MA, 1993.
18. R. Welch and D. Warren, *Intersensory interactions*, Wiley, New York, 1986.
19. 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.
20. M. Gröhn, M. Laakso, M. Mantere, and T. Takala, "3d visualization of building services in virtual environment," *Proc. SPIE* **4297B**, (San Jose, California), Jan 2001.
21. M. Gröhn and T. Takala, "Magicmikes: method for spatial sonification.," *Proc. SPIE* **2410**, (San Jose, California), Jan 1995.
22. V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society* **45**(6), pp. 456–466, 1997.
23. M. Gerzon, "Periphony: With-height sound reproduction," *Journal of the Audio Engineering Society* **21**(1/2), pp. 2–10, 1973.
24. D. Malham and A. Myatt, "3-d sound spatialization using ambisonics techniques," *Computer Music Journal* **19**(4), pp. 58–70, 1995.
25. E. Lantz, S. Bryson, D. Zeltzer, M. Folas, B. de la Chapelle, and D. Bennet, "The future of virtual reality: Head mounted displays versus spatially immersive displays," *Proc. SIGGRAPH'96* , pp. 485 – 486, (New Orleans, USA), 1996.

26. D. Perrot, "Auditory and visual localization: two modalities, one world," *Audio Engineering Society 12th International Conference: The perception of reproduced sound* , pp. 221–231, (Copenhagen), 1993.
27. T. Lokki, M. Gröhn, L. Savioja, and T. Takala, "A case study of auditory navigation in virtual acoustic environments," *Proc. ICAD 2000* , (Atlanta GA,), Apr 2000.
28. J. Loomis, R. Golledge, and R. Klatzky, "Navigation system for the blind: Auditory display modes and guidance," *Presence: Teleoperators and Virtual Environments* **7**, pp. 193–203, April 1998.
29. P. Rutherford, "Virtual acoustic technology: Its role in the development of an auditory navigation beacon for building evacuation," in *Proc. 4th UK Virtual Reality SIG Conference*, R. Bowden, ed., Brunel University, (London, UK), 1997.
30. T. Takala and J. Hahn, "Sound rendering," *Computer Graphics* **SIGGRAPH'92**(26), pp. 211–220, 1992.
31. K. Riederer, "Repeatability analysis of hrtf measurements," in *the 105th Audio Engineering Society (AES) Convention, preprint no. 4846*, (San Francisco, USA), Sept. 26-29 1998.