

# COMPARISON OF AUDITORY, VISUAL, AND AUDIO-VISUAL NAVIGATION IN A 3D SPACE

*Matti Gröhn*

CSC - Scientific Computing Ltd  
P.O.Box 405, FIN-02101 Espoo, Finland  
matti.grohn@csc.fi

*Tapio Lokki, Tapio Takala*

Telecommunications Software and Multimedia  
Laboratory  
Helsinki University of Technology  
P.O.Box 5400, FIN-02015 HUT, Finland  
{tapio.lokki, tapio.takala}@hut.fi

## ABSTRACT

A navigation test was carried out in a spatially immersive virtual environment. The test was a game-like experience where task of subjects was to find as many gates as possible while they navigated through a track guided by auditory and/or visual cues of the gates. The results are presented as a function of the number of found gates, searching times, and normalized path lengths. Audio-visual navigation was clearly the most efficient. Visual navigation was the second, and the auditory navigation the least efficient. Further analysis of travel paths indicate that auditory cue was utilized in the beginning to locate the next gate, and visual cue was the most important in the final approach to the gate.

## 1. INTRODUCTION

Smooth and easy navigation in virtual reality applications is a desired feature. Usually navigation is based on visual information, as the major part of applications are related to visualization of 3D models. However, high quality immersive virtual environments include 3D audio equipment and therefore the use of 3D auditory cues to help navigation is motivated. For example, auditory navigation can be used as a part of immersive architectural visualization. Auditory beacons can guide the user to locations, that are not visible due to buildings or walls.

The purpose of our test was to find out how well people can navigate in virtual environments with different kind of cues. The tests were conducted in an immersive virtual environment. In order to motivate test subjects, we designed a race-game-like test environment.

The navigation control method has to be reliable and easy to use. We use a custom made wand (consists of a radio mouse and a magnetic 3D tracker). In our previous study of 2D auditory navigation [1] we applied keyboard control, which is not suitable for 3D navigation. Laakso [2] has tested different interaction devices, such as speech input, data glove, and a wand-like device. In these tests the most accurate navigation was achieved with a wand and a data glove. In addition, they were the easiest ones to use.

From previous tests [3, 4] we have preliminary knowledge on localization of moving sounds. In this article we want to apply it to a navigation task.

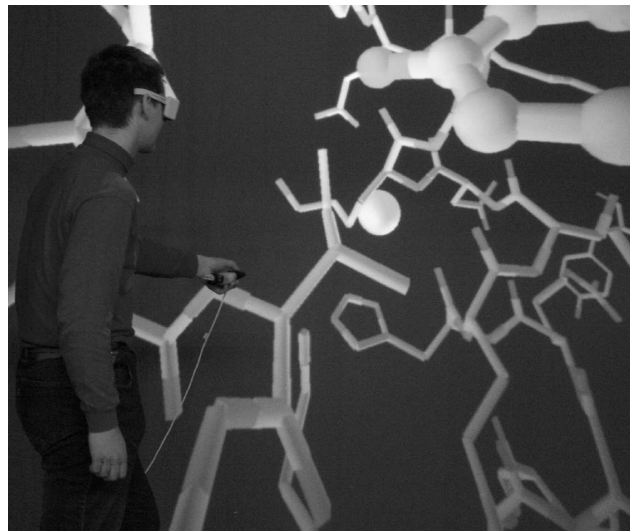


Figure 1: Protein-drug complex was used as a visual stimulus in navigation experiment. Visual cue (ball) of the gate is seen in the middle of the figure.

### 1.1. Sound reproduction system in a virtual room

The experiment was conducted in the virtual room<sup>1</sup> of the Helsinki University of Technology. For spatial audio reproduction we use 14 Genelec 1029A loudspeakers, mounted behind the screens around the virtual room. The applied spatial sound reproduction method is called vector base amplitude panning (VBAP) [5] and it enables good positional audio with almost arbitrary loudspeaker configuration. VBAP distributes a sound signal to three nearest loudspeakers of the virtual sound source. The screens attenuate high frequencies but this defect is compensated for. A more detailed description of our audio system and screen compensation can be found in another article [6].

<sup>1</sup><http://eve.hut.fi>

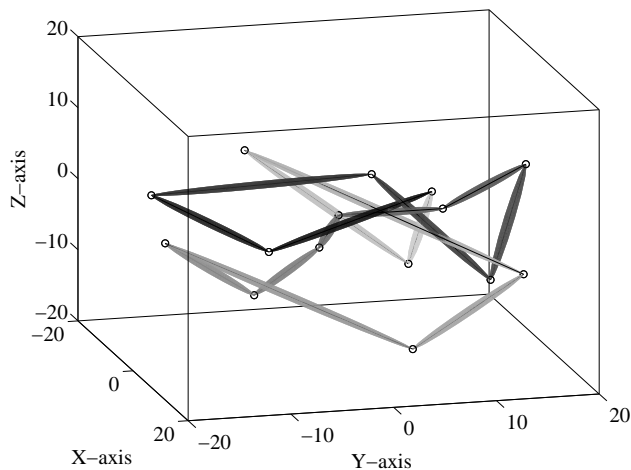


Figure 2: Track configuration of the main experiment. Circles indicate the gate positions. Gate order is indicated with gray-scale value.

## 2. NAVIGATION EXPERIMENT

### 2.1. Task

The task of a subject was to find as many gates as possible in three minutes time. They navigated through a predefined track controlling direction and velocity of movements by pointing with the wand. The gesture of pushing a wand button and moving the wand in space defines a vector, the length and direction of which are translated into motion speed and direction in a virtual space. In addition, rotations can be handled correspondingly by turning the wand.

Subjects were flying in the middle of the protein-drug complex (Figure 1). This complex is a real model from analysis of protein-drug interactions [7, 8, 9].

The minimum and maximum values of the coordinates of the protein-drug complex were -20 and 20 for each axis. The track consists of fifteen gates, which were located in the same space as the molecule. The minimum and maximum values for gate coordinates were -12 and 12 for x-axis, -15 and 18 for y-axis, and -15 and 10 for z-axis. Track configuration is seen in Figure 2. Distances between gates range between 9.9 and 39.1.

The shape of a gate was a sphere. The gate has been considered as found, when subject hit the border of the sphere. The radius of a visual cue (see Fig. 1) was smaller than the radius of the gate.

### 2.2. Subjects

We had eight (six male and two female) non-paid volunteers for this experiment. Each of them reported to have normal hearing, although this was not verified with audiometric tests.

### 2.3. Stimuli

The gates of the track were presented using visual, auditory, or audio-visual cues. The visual cue was a white ball as seen in Figure 1. The auditory cue was pink noise bursts and the sound pressure

| Cues         | Median | Min | Max |
|--------------|--------|-----|-----|
| Auditory     | 8.5    | 3   | 15  |
| Visual       | 18     | 11  | 31  |
| Audio-visual | 27.5   | 20  | 41  |

Table 1: Number of gates found in three minutes for each cue type.

| Cues         | Median | Min | Max  |
|--------------|--------|-----|------|
| Auditory     | 12.8   | 2.6 | 99.2 |
| Visual       | 7.5    | 1.4 | 52.7 |
| Audio-visual | 5.7    | 1.2 | 20.2 |

Table 2: Searching times for each cue type

level was computed with the  $1/r$ -law distance attenuation. The audio-visual cue included both auditory and visual cues.

The choice for auditory cue was based on the fact that to utilize both main binaural cues (interaural time and level differences), the auditory cue should have enough energy at low (below 1.5 kHz) and high frequencies (above 1.5 kHz). In addition, other factors affect localization accuracy such as spectral and temporal structure of stimulus (see for example [10, 11, 12]). The auditory cue was in a free field without reverberation.

### 2.4. Procedure

There were three types of conditions: auditory, visual, and audio-visual cues.

Each task was started in the middle of the molecule. The index of the first gate was randomized. After the first gate the subject followed the track gate by gate. Both directions were used. Finding of the gate was indicated with a short bell-like auditory signal. The next gate was displayed/made audible according to the cue type of the task and the found gate was hidden/muted.

Before the experiment each subject had training tasks. The purpose of the training tasks was to ensure, that subjects were able to use the navigation system. The number of the training tasks varied according to subjects familiarity with the navigation system. Length of each training task was two minutes.

Track configuration with five gates was used in training. For each user the training was started with audio-visual cues. Each subject had each cue type condition at least once. The criterium to start the main experiment was, that the subject felt comfortable with the navigation system. In other words, the subject could concentrate on finding the gates instead of using the navigation system.

In the main experiment each subject had each cue type twice and the subjects had three minutes time to find as many gates as they could. Number of the found gates was recorded. In addition, location of the subject, and time were recorded with 10 Hz sampling rate.

## 3. RESULTS

Based on our previous 2D experiment [1] and preliminary studies we believed that auditory navigation is feasible also in 3D environments. According to this experiment each subject was able to find at least a few gates by using auditory cues only.

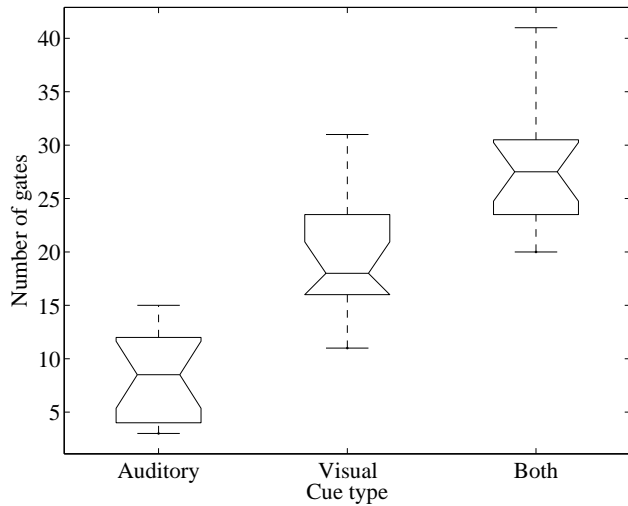


Figure 3: Boxplot of number of found gates for each cue type. The box indicates the lower quartile, median, and upper quartile values.

In Table 1 the median, minimum, and maximum number of found gates for each cue type are displayed. With a visual cue the subjects found twice as much gates as with an auditory cue. If audio-visual cue was used the number of found gates increased even more. The differences are clearly seen in Figure 3.

In Table 2 the median, minimum, and maximum of the searching times (in seconds) for each gate are displayed. The differences between the median times are not as large as could be expected according to differences in number of found gates. If we compare the maximum searching times the differences are more remarkable. The differences are seen also in Figure 4.

In Figure 5 the boxplot of normalized path lengths are displayed. As was expected based on searching times, the subjects travel the longest paths with an auditory cue, and with an audio-visual cue the paths were the shortest.

For further analysis we inspected one of the longest gate pairs. In Figure 6 the navigation paths for each cue type are displayed for the long distance situation. With the auditory cue we see, that subjects had some problems to find the accurate location of the gate. They started their approach well, but they had problems to finish the approach. With the visual cue the situation is almost opposite. The subjects had some difficulties to localize the next gate. After they localized it, the final approach was straightforward. If both cues were used, the subjects started their approach well and they also finished it without difficulties.

#### 4. DISCUSSION

Applied six degrees of freedom navigation method enabled totally free moving in the test space. Our subjects did not recognize, that only one track was used in the main experiment. They did find out, that there were short and long gate distances, but they did not recognize the shape of the track.

In our comparison audio-visual cues were remarkably better than auditory or visual cue alone. Although the visual cue was occasionally occluded and auditory cue was presented in a free field, the visual navigation was more efficient than auditory navigation.

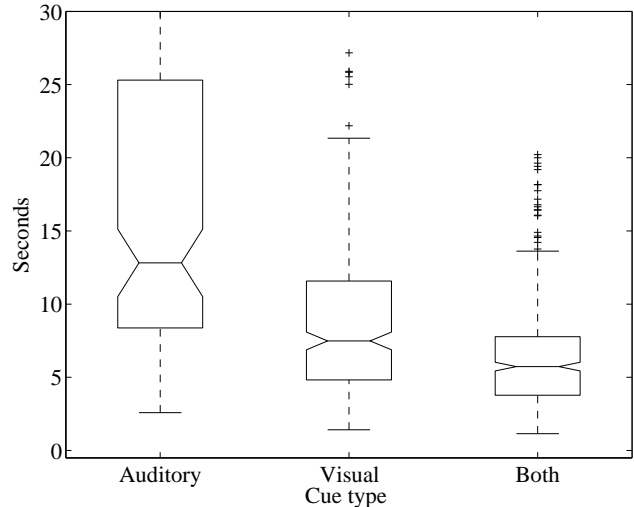


Figure 4: Boxplot of searching times for each cue type. Searching time = time spent between the gates.

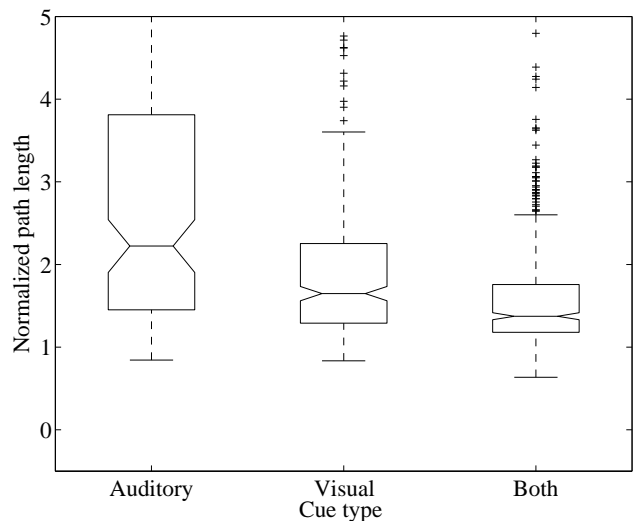


Figure 5: Boxplot of normalized path lengths for each cue type. Normalized path length = Path length / Distance between gates. Normalized path length could be a little less than one, because the subject could find the gate without going to the center of gate.

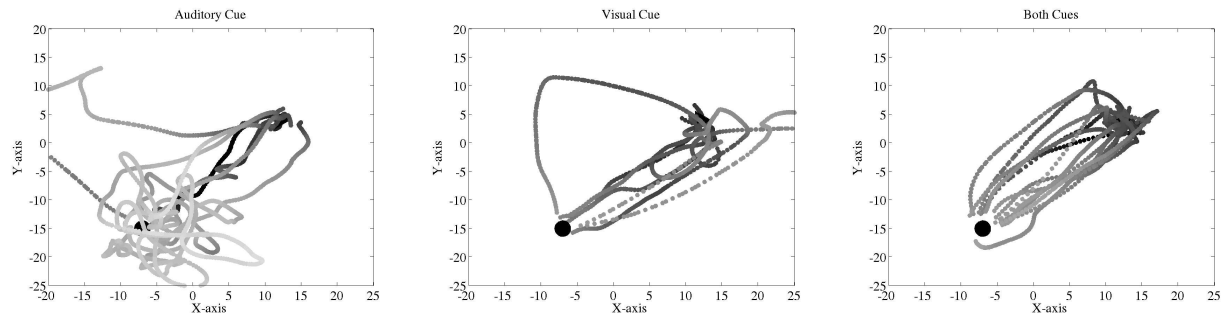


Figure 6: 2D projection of the navigation paths for long distance from left to right: auditory, visual, and audio-visual cues. Coordinates of the starting point (triangle): 12, 4, and end point (dot): -7, -15.

Navigation paths indicate, that subjects utilized the auditory cue (if available) to define the approximate location of the gate, and the visual cue (if available) in the final approach. In other words, rough location was based on auditory perception while the final approach was dominated by vision.

According to these results the auditory 'highlighting' of important features of the objects can be expected to work well in virtual reality applications. Especially in such applications, where the information is surrounding the user, and many of the important features are located outside the main field of view.

## 5. ACKNOWLEDGEMENTS

We would like to thank Mr. Tommi Ilmonen for his work for our audio software and hardware. We also would like to thank Mr. Iikka Olli for providing the navigation system for our experiment.

## 6. REFERENCES

- [1] T. Lokki, M. Gröhn, L. Savioja, and T. Takala, "A case study of auditory navigation in virtual acoustic environments," in *Proc. ICAD 2000*, Atlanta GA, Apr 2000, pp. 145–150.
- [2] M. Laakso, *Practical Navigation in Virtual Architectural Environments*, Master's Thesis, Helsinki University of Technology, Espoo, Finland, 2001.
- [3] M. Gröhn, T. Lokki, and T. Takala, "Static and dynamic sound source localization in a virtual room," in *Proc. AES 22nd Int. Conf. on Virtual, Synthetic and Entertainment Audio*, Espoo, Finland, June 15-17 2002, pp. 337–344.
- [4] M. Gröhn, "Localization of moving virtual sound source in a virtual room, the effect of a distracting stimulus," in *Proc. of ICAD 2002*, Kyoto, Japan, July 2-5 2002, pp. 394–402.
- [5] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, June 1997.
- [6] J. Hiiipakka, T. Ilmonen, T. Lokki, M. Gröhn, and L. Savioja, "Implementation issues of 3D audio in a virtual room," in *Proc. SPIE*, San Jose, California, Jan 2001, vol. 4297B.
- [7] T. Nyrönen, M. Pihlavisto, J.M. Peltonen, A.M. Hoffrén, M. Varis, T. Salminen, S. Wurster, A. Marjamäki, L. Kanerva, E. Katainen, L. Laaksonen, J.M. Savola, M. Scheinin, and M.S. Johnson, "Molecular mechanism for agonist-promoted alpha(2a)-adrenoceptor activation by norepinephrine and epinephrine," *Molecular Pharmacology*, , no. 59, pp. 1343–1354, 2001.
- [8] T. Salminen, M. Varis, T. Nyrönen, M. Pihlavisto, A.M. Hoffrén, T. Lönnberg, A. Marjamäki, H. Frang, J.M. Savola, M. Scheinin, and M.S. Johnson, "Three-dimensional models of alpha(2a)-adrenergic receptor complexes provide a structural explanation for ligand binding," *Journal Of Biological Chemistry*, , no. 274, pp. 23405–23413, 1999.
- [9] M. Gröhn and T. Nyrönen, "Integration and utilization of different visualization methods and devices in a structure-based drug design process," in *Proc. SPIE*, Santa Clara, California, Jan 2003, vol. 5009.
- [10] F.L. Wightman and D.J. Kistler, "Factors affecting the relative salience of sound localization cues," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R.H. Gilkey and T.R. Anderson, Eds., pp. 1–23. Lawrence Erlbaum Associates Inc., 1997.
- [11] R.O. Duda, "Elevation dependence of the interaural transfer function," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R.H. Gilkey and T.R. Anderson, Eds., pp. 49–75. Lawrence Erlbaum Associates Inc., 1997.
- [12] J.C. Middlebrooks, "Spectral shape cues for sound localization," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R.H. Gilkey and T.R. Anderson, Eds., pp. 77–97. Lawrence Erlbaum Associates Inc., 1997.