# Publication VII

## Navigation with auditory cues in a virtual environment

**T. Lokki, and M. Gröhn**


Reprinted, with permission, from *IEEE Multimedia*, Vol. 12, Nr. 2, pages 80–86, Apr.–Jun. 2005

# Navigation with Auditory Cues in a Virtual Environment

**Tapio Lokki**
*Helsinki University of Technology, Finland*

**Matti Gröhn**
*CSC–Scientific Computing, Finland*

**We use 3D sound to help navigate an immersive virtual environment and report results of user tests obtained with a game-like application. The results show that auditory cues help in navigation, and auditory navigation is possible even without any visual feedback. The best performance is obtained in audiovisual navigation where auditory cues indicate the approximate direction and visual cues help in the final approach.**

Smooth and easy navigation is an essential feature in virtual reality (VR) applications. Typically navigation is based on visual information, as a great deal of VR applications are related to visualization of 3D models. Rutherford[1] presented one of the first studies on auditory navigation. He studied the effect of different auditory signals and proposed the use of auditory beacons to help navigation in building evacuation simulations. Lokki et al.[2] and Walker et al.[3] proposed navigation studies on the effect of stimulus sound signals with headphone reproduction. Loomis et al.[4] and Holland et al.[5] also proposed navigation systems that apply audio for visually impaired people in outdoor environments.

Our research applies an audio reproduction system based on loudspeakers rather than headphones. Current high-quality immersive virtual environments include multichannel 3D audio equipment, making it easier to use 3D auditory cues when navigating. For example, we can use auditory navigation as a part of immersive architectural visualization in which auditory beacons[6] can guide the user to locations that aren't visible because of buildings, walls, or other occluding objects. Auditory beacons can also emphasize an interesting part of a 3D model, such as a chemically active part of a visualized protein.

We tried two experiments in applying 3D auditory cues for navigation. In the first experiment, we compared visual, auditory, and audiovisual cues. Preliminary results of this audiovisual test are available elsewhere,[7] but here we present an in-depth analysis of the results. The second experiment involves navigation performed with different auditory cues without any visual information.

## Complementary Media

The audio and video files we used in our experiments are available at http://csdl.computer.org/comp/mags/mu/2005/02/u2toc.htm.

### Environment for navigation tests

We conducted the navigation tests in the immersive virtual room of the Helsinki University of Technology. We used 14 Genelec 1029A loudspeakers for spatial audio reproduction, mounted behind the back projection screens, as Figure 1 shows. We applied vector base amplitude panning (VBAP)[8] as a spatial sound reproduction method. This enables reasonable positional audio with almost arbitrary loudspeaker configuration.

In a virtual room the screen between the loudspeaker and listener modifies the spectrum of the perceived sound signal. We found in our case that high frequencies (higher than 2,000 Hertz) of the direct sound are attenuated at more than 10 decibels (dB). To equalize the frequency response of loudspeakers, we measured impulse responses from each loudspeaker to the middle of the room (the most common listening position). Sixth-order infinite impulse response (IIR) filters were fitted to measured responses for spectral compensation. A detailed description of the compensation of screen damping and implementation details of our audio environment can be found elsewhere.[9]

### Navigation task

To motivate test subjects, we designed a game-like test in which the task was to navigate through gates on a predefined track. We used a white ball to visually represent the aural gate and a point-like sound that acted aurally to indicate the gate. The gates were inside a complex 3D model (see Figures 2 and 3) obtained from an analysis study of protein-drug interactions.[10]

Subjects controlled the movement in the virtual world with a custom-made wand consisting of a radio mouse (Logitech Surfman) and a magnetic 3D tracker (Ascension MotionStar). The sub-



1070-986X/05/$20.00 © 2005 IEEE

Published by the IEEE Computer Society

ject navigated—controlling movement direction and velocity—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. Similarly, the user could rotate himself/herself by turning the wand. The device is intuitive to use and allows six degrees of freedom (6DOF) navigation.

The track consisted of 15 gates, located in the same space as the molecule (see Figure 3). The boundaries of the protein-drug complex at each axis were –20 at minimum and 20 at maximum. The gate coordinates were between –12 and 12 for the *x*-axis, –15 and 18 for the *y*-axis, and –15 and 10 for the *z*-axis. Therefore, the minimum distance between the gates was 9.9 and the maximum distance was 39.1.

The gate was a point in a space. We considered the gate found when a subject reached the gate's capture radius. A visual cue's radius (see the white ball in Figure 2) in the audiovisual experiment was smaller than the gate's capture radius.

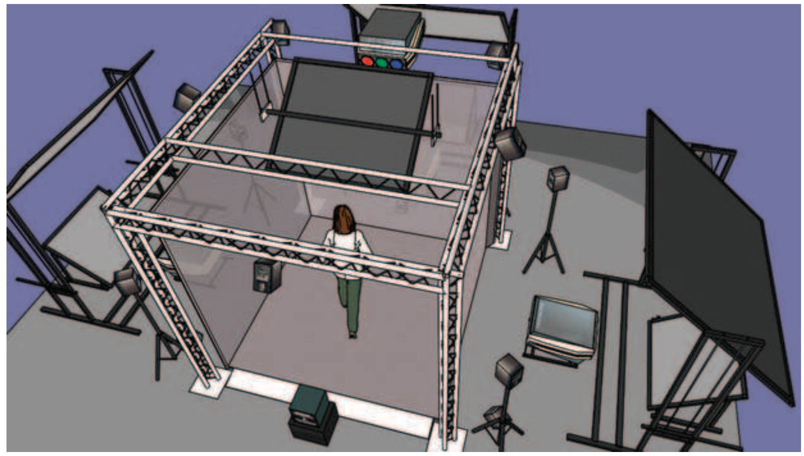


*Figure 1. Schematic drawing of the virtual room at the Helsinki University of Technology.*

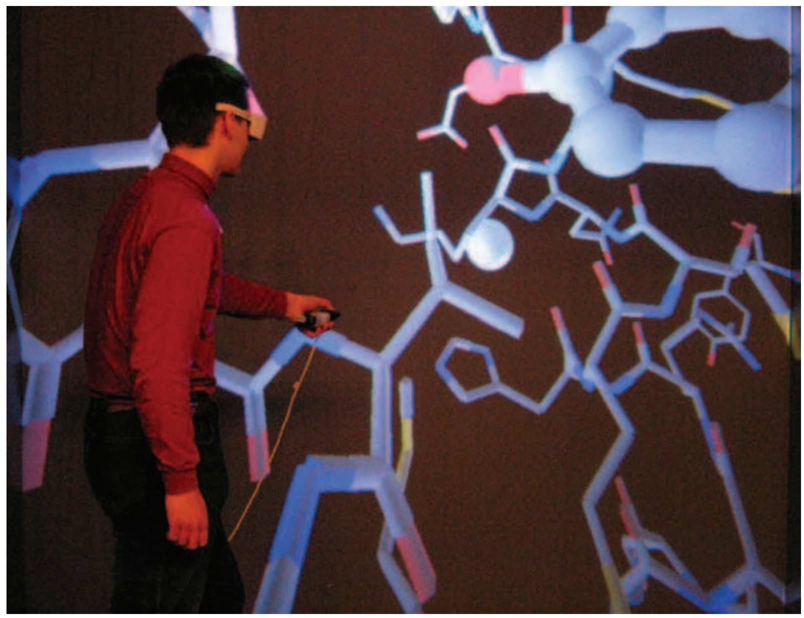


*Figure 2. Protein-drug complex used as a virtual world in navigation experiments. The white ball in the middle served as the gate's visual cue.*

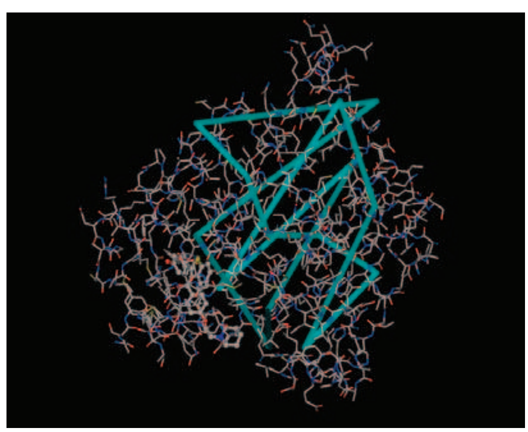


*Figure 3. Track configuration of the navigation experiment, drawn with a blue tube. Gates are in the corners of the track.*

### Test procedure

Each task was started in the middle of the molecule and the index of the first gate was randomized. After the first gate a subject followed the predefined track gate by gate in either direction. A short bell-like auditory signal indicated a found gate, which was hidden and/or muted. At the same time the next gate was displayed according to the task's cue type.

By training the subjects before the experiment we ensured that they could concentrate on finding the gates instead of using the navigation system. The criterion to start the main experiments was that a subject felt comfortable with the navigation system. Therefore, the number of training tasks varied according to the subjects' familiarity with the navigation system. During the training session each subject tried each cue type at least once.

In the main experiments each subject had each cue type twice in random order. The subjects had 3 minutes (for audiovisual tests) or 2.5 minutes (for auditory tests) to find as many gates as possible. The collected data was the number of found gates. In addition, we recorded the subject's location with a 10-Hz sampling rate.

### First experiment: Auditory, visual, and audiovisual navigation

In the first experiment the track's gates were presented using visual, auditory, or audiovisual

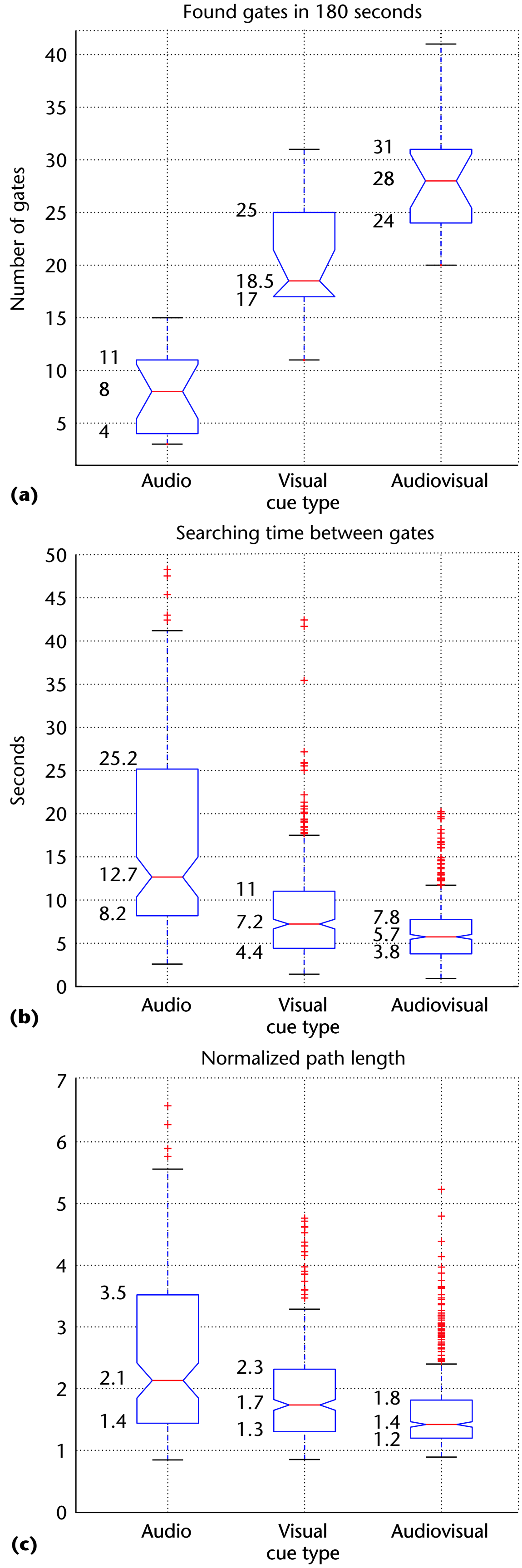



*Figure 4. Results of the audiovisual navigation test. (a) Number of found gates. (b) Search times between a gate pair. (c) Normalized path lengths of traveled distance.*

cues in a 3D space. The visual cue was a white ball. The auditory cues were pink noise bursts. We computed the sound pressure level with the $1/r$ law distance attenuation. A pink noise burst has a wide spectrum and temporal variation, both of which are important cues for localization. The third cue type consisted of both auditory and visual cues. An example video of the navigation with the third cue type is available at http://csdl.computer.org/comp/mags/mu/2005/02/u2toc.htm.

Nine subjects (seven males and two females) completed the tests so that each cue type was presented twice to each subject. Thus, in total each subject had six of the 3-minute-long runs in randomized order. All subjects reported having normal vision and hearing.

We analyzed three factors from the measured data:

- number of found gates,
- search times between gates, and
- normalized path lengths.

The statistical analysis was made with the Kruskal–Wallis test, which is a nonparametric version of one-way analysis of variance (ANOVA). The traditional ANOVA model couldn't be applied in this case since group variances weren't equal. If the Kruskal–Wallis test indicated significant differences between groups, post-hoc comparisons were performed by applying critical mean rank values from the $t$ distribution, after a Bonferroni adjustment to compensate for multiple comparisons.[11]

The results show that navigation with audiovisual cues is remarkably easier and faster than with auditory or visual cues alone. Figure 4a shows that the median values of found gates in 3 minutes were 8 (auditory cue), 18.5 (visual cue), and 28 (audiovisual cue) gates. The difference is statistically significant ($p$ = 0.0000) and post-hoc testing indicates that the three group means differ. Figures 4b and 4c present the search times and normalized path lengths between gates. Because the task was to find as many gates as possible in 3 minutes, the search times compare cue types similarly as the number of found gates. The normalized path lengths tell how much distance subjects travel between gate pairs. For both search times and normalized path lengths, the differences between groups are significant (both $p$ = 0.0000) and all three group means differ.

To further study the subjects' behavior we analyzed the navigation paths with each cue type. As an example, Figure 5 depicts all travel paths between one gate pair. The plots suggest that the subjects used the auditory cue (if available) to define the approximate direction to the target 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. When both cues were available navigation from one gate to the next was straightforward.

## Second experiment: Auditory navigation with different stimuli

The results of the first experiment showed that navigation based on auditory cues only is possible and all subjects could find gates without any visual cues. This result encouraged us to further develop auditory navigation.

Based on the analysis of travel paths of the audiovisual experiment, we thought that the main problem in auditory navigation is the inaccuracy in elevation perception. In many cases subjects came close to the sound source, but went above or below it. Therefore, we designed a new stimulus in which the height of the sound source was

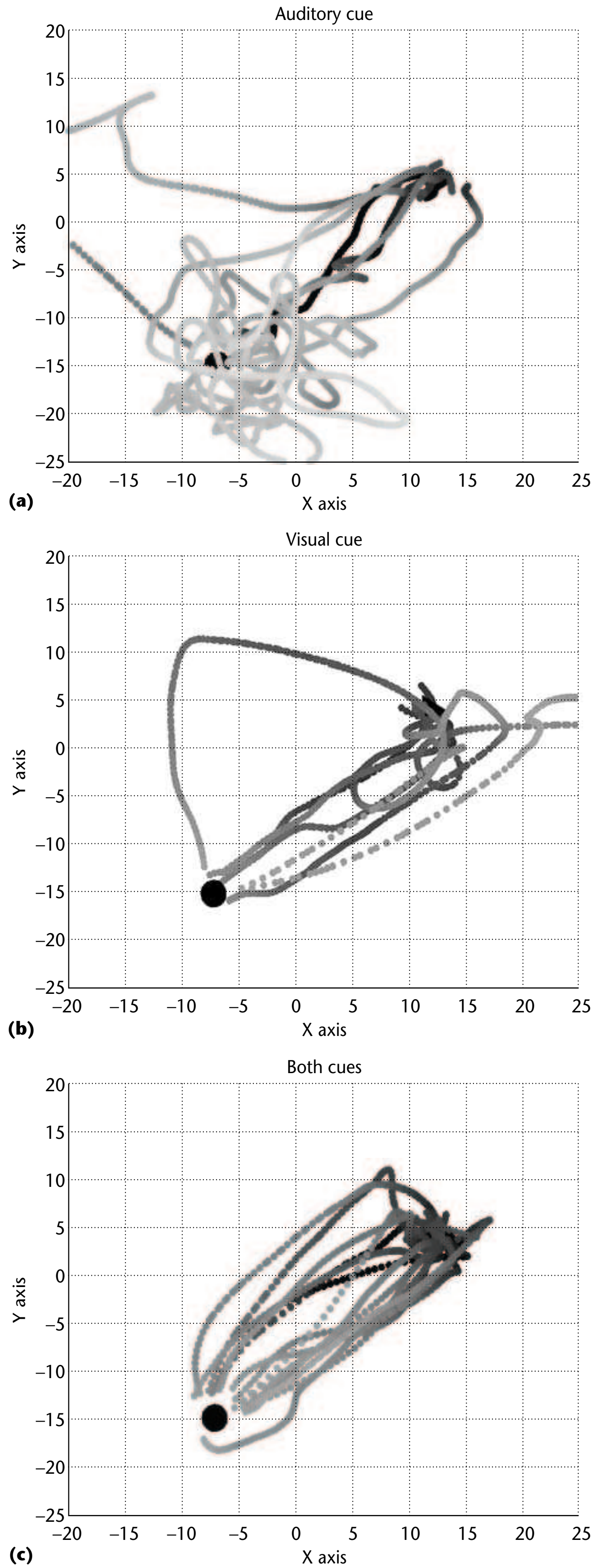
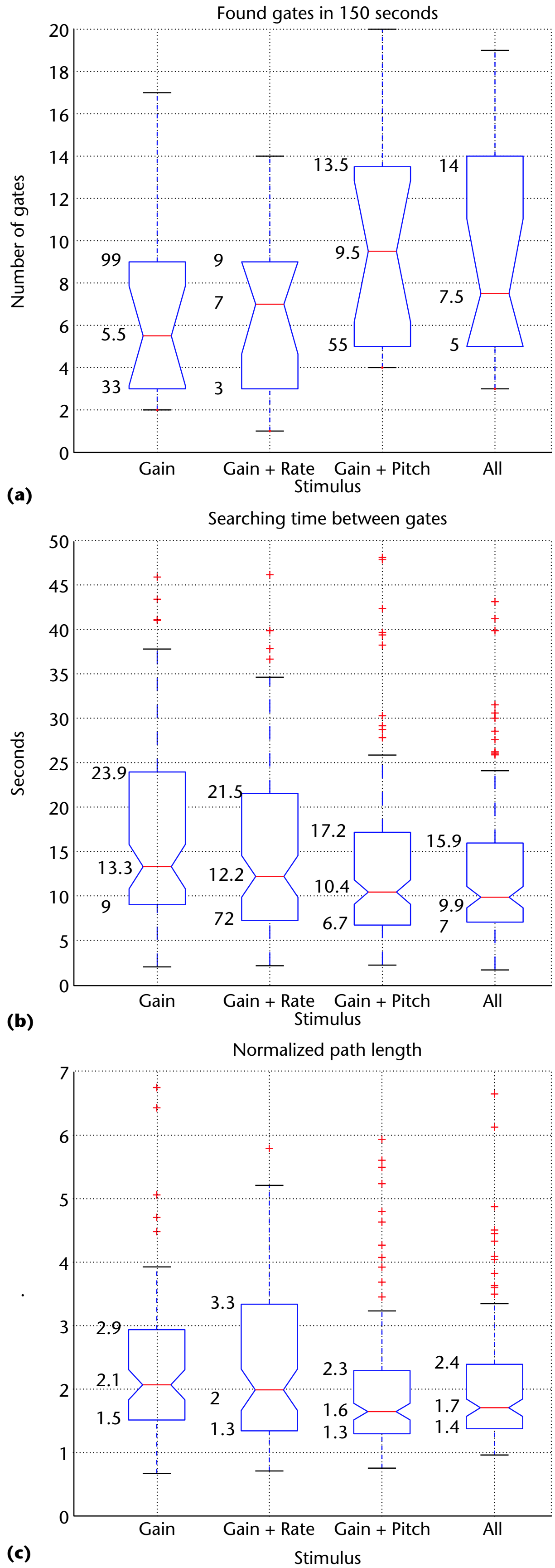



***Figure 5. Each curve is a 2D projection of the traveled path from the beginning (12, 4, –15) to the end (–7, –15, –5; marked with a dot). The gray scale indicates the height (z) of the paths. (a) Navigation paths with auditory cues. (b) Navigation paths with visual cues. (c) Navigation paths with audiovisual cues.***

*Figure 6. Results of the second experiment, with different auditory stimulus signals. (a) Number of found gates. (b) Search times between a gate pair. (c) Normalized path lengths of traveled distance.*




also encoded to the sound signal itself. In addition, we made another new sound stimulus with better distance cues.

To verify our hypothesis, we did a second experiment in which we tested the auditory navigation with four different sound signals without any visual cues. The track configuration and the test procedure were the same as in the audiovisual experiment, except that each run lasted only 2.5 minutes.

All four new stimuli were based on pink noise bursts with 1/*r* law distance attenuation. The first one (Gain) was pure pink noise bursts—that is, the same signal as the auditory cue (Audio) in the audiovisual experiment. In the second signal (Gain + Rate) the distance to the next gate was also indicated with the density of bursts—that is, the pulse rate. The closer the subject was the denser the bursts. The third stimulus (Gain + Pitch) consisted of noise bursts and a narrow band-pass noise, the center frequency of which represented the height of the sound source. A clear step in pitch indicated the exact height of the sound source. Finally, the fourth signal type (All) contained all cues. Short example videos of all cue types are available at http://csdl.computer.org/comp/mags/mu/2005/02/u2toc.htm.

Eight male subjects completed the second experiment (three were also in the audiovisual tests). In this second experiment each signal type was presented twice to each subject; thus in total each subject did eight of the 2.5-minute-long runs in randomized order. We performed the statistical analysis with the same methods as in the audiovisual test.

Figure 6 shows the results of the second experiment. Here we can see that the results with the Gain stimulus are similar to the results of the Audio in the audiovisual experiment (from Figure 4). This fact confirms that our test environment

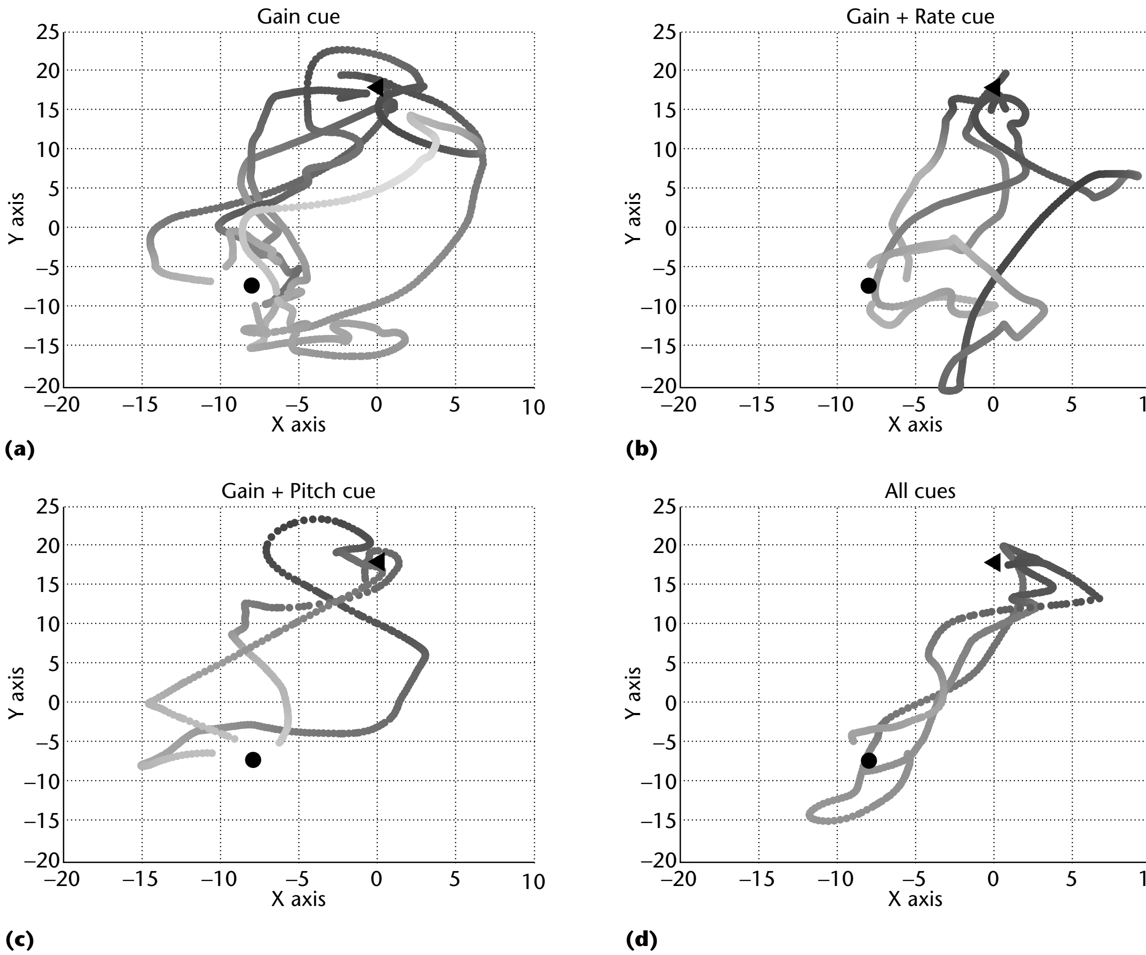



*Figure 7. Each curve is a 2D projection of the traveled path from the starting point (0, 18, –10; marked with a triangle) to the end point (–8, –7, 7; marked with a dot). The gray scale indicates the height (z) of the paths. Note that fewer paths are depicted than in Figure 5, since in this test fewer gates were found than in the first experiment. (a) Navigation paths with Gain cue. (b) Navigation paths with Gain + Rate cue. (c) Navigation paths with Gain + Pitch cue. (d) Navigation paths with All auditory cues.*

hasn't changed between the tests.

Although Figure 6a indicates that with the Gain + Pitch stimulus the number of found gates is larger, the difference between stimuli isn't significant ($p$ = 0.1000). However, difference in search times (Figure 6b) and normalized path lengths (Figure 6c) are statistically significant ($p$ = 0.0083 and $p$ = 0.0042, respectively). Post-hoc analysis shows that with search times the mean ranks of Gain + Pitch and All differ significantly from the mean rank of Gain. Correspondingly, the mean rank of Gain + Pitch differs significantly from the mean ranks of Gain and Gain + Rate with normalized path lengths.

Similarly with the audiovisual test, we analyzed the travel paths between gate pairs, and Figure 7 depicts one example gate pair. The paths strengthen the results. Travel paths with Gain + Pitch information (Figures 7c and 7d) are less complex than in Figures 7a and 7b. In particular, the elevation of the subjects rises smoothly when approaching the target gate.

## Discussion

The results of the second experiment indicate that our assumption about the difficulty of hearing the height of the sound source was correct. When we encoded elevation information to the navigation cue signal, subjects found gates faster and the travel distance was shorter.

When comparing the data of the first and second experiment, we find an interesting result. The normalized travel path lengths with the Visual cue are close to Gain + Pitch and All stimuli results (no significant difference, $p$ = 0.5001). However, searching times are longer with auditory navigation (significant difference, $p = 0.0000$). In any case, our results suggest that auditory navigation is almost as easy as visual navigation, when sound signals are carefully designed.

## Conclusion

What we've learned in our experiments is that 3D navigation in a virtual environment is possible with auditory cues alone. However, the fastest and most accurate navigation is obtained when both auditory and visual cues are available.

It's helpful for us to see that we can enhance a user's performance with reasonable sound signal design. For example, navigation is easier if the height information of a sound source is encoded in the signal itself. This finding is reasonable also from the perspective of human spatial hearing, since the azimuth perception is more accurate than the elevation perception.

In addition to the benefits of using sound in navigation cues, we can also apply 3D sound to "highlight" the important features of objects in VR applications. This is especially helpful in situations where the information is surrounding the user, and many of the important features are located outside the main field of view.

In the future, auditory navigation should be tested in dffierent virtual environments with real applications. In the long run, we'd like to apply spatial auditory display in data representations, object selections, and object manipulations in VR applications. **MM**

## References


1. P. Rutherford, "Virtual Acoustic Technology: Its Role in the Development of an Auditory Navigation Beacon for Building Evacuation," *Proc. 4th UK Virtual Reality SIG Conf.*, R. Bowden, ed., United Kingdom Virtual Reality Special Interest Group, 1997.
2. T. Lokki et al., "A Case Study of Auditory Navigation in Virtual Acoustic Environments," *Proc. 6th Int'l Conf. Auditory Display*, 2000, Int'l Community of Auditory Display, pp. 145-150.
3. B. Walker and J. Lindsay, "Effect of Beacon Sounds on Navigation Performance in a Virtual Reality Environment," *Proc. 9th Int'l Conf. Auditory Display*, Boston Univ. Publications Dept., 2003, pp. 204-207.
4. J. Loomis, R. Golledge, and R. Klatzky, "Navigation System for the Blind: Auditory Display Modes and Guidance," *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 2, Apr. 1998, pp. 193-203.
5. S. Holland, D. Morse, and H. Gedenryd, "AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface," *Personal and Ubiquitous Computing*, vol. 6, no. 4, Sept. 2002, pp. 253-259.
6. G. Kramer, ed., *Auditory Display—Sonification, Audification, and Auditory Interfaces*, *Proc. Santa Fe Inst. Studies in Sciences of Complexity*, vol. XVIII, Addison and Wesley, 1994.
7. M. Gröhn, T. Lokki, and T. Takala, "Comparison of Auditory, Visual, and Audio-Visual Navigation in a 3D Space," *Proc. 9th Int'l Conf. Auditory Display*, Boston Univ. Publications Dept., 2003, pp. 200-203.
8. V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, June 1997, pp. 456-466.
9. J. Hiipakka et al., "Implementation Issues of 3D Audio in a Virtual Room," *Proc. 13th Symp. Soc. for Imaging Science and Technology/ Int'l Soc. for Optical Eng.* (IS&T/SPIE) *Electronic Imaging 2001*, SPIE, vol. 4297B, 2001, pp. 486-495.
10. M. Gröhn and T. Nyrönen, "Integration and Utilization of Different Visualization Methods and Devices in a Structure-Based Drug Design Process," *Proc. Int'l Soc. for Optical Eng.* (SPIE 2003), SPIE, vol. 5009, 2003, pp. 11-18.
11. Mathworks, "Matlab—Online Documentation," http://www.mathworks.com/access/helpdesk/help/helpdesk.html, 2004.



**Tapio Lokki** works as a teaching researcher in the Laboratory of Telecommunications Software and Multimedia at the Helsinki University of Technology, Finland. Lokki received his DSc degree in computer science from the Helsinki University of Technology. His research interests include room acoustic modeling, auralization, and audio in virtual reality applications.

**Matti Gröhn** is a visualization specialist in the CSC–Scientific Computing, Finland, the Finnish information technology center for science. He's the current president of the International Community for Auditory Display. His research interests include sonification, visualization, multimodal data representation, and virtual reality. Gröhn has an MSc in computer science from the University of Helsinki.

Readers may contact Tapio Lokki at Tapio.Lokki@hut.fi and Matti Gröhn at Matti.Grohn@csc.fi.