Using Binaural Hearing for Localization in Multimodal Virtual Environments

Matti Gröhn, Tapio Lokki, Lauri Savioja

Telecommunications software and multimedia lab., Helsinki Univ. of Tech., P.O.Box 5400, FIN-02015 HUT, Finland

In this article we present the results of localization experiment accomplished in a real virtual room. The spatial audio reproduction methods used in experiment were HRTFs (for headphones), direct loudspeaker reproduction, and vector based amplitude panning. Also the effect of visual stimulus was explored.

BACKGROUND

Virtual environment is an interactive immersive multisensory 3-D synthetic environment; it uses position-tracking and real-time update of visual, auditory, and other displays (e.g., tactile). Currently spatial audio has been mostly used for sense of presence.

There are four common tasks in immersive visualization (localization, orientation, navigation, and data representation). In this research localization is defined as user ability to define direction and distance of the target. The other tasks and more detailed description of this research project as a whole is in article [1].

In this research we are comparing head-tracked binaural headphone reproduction (using HRTFs), direct loudspeaker reproduction, and vector based amplitude panning (VBAP) using loudspeakers [2]. The task of the user is to point to direction of the perceived sound source, and click the button of a interaction device. According to Djelani et al. [3] pointing is an appropriate method for localization experiments. In our test the user can freely move inside the virtual environment (as they typically do while they are using some application).

LOCALIZATION TEST

According to Blauert [4] the absolute lower limit for localization accuracy in front is one degree. In our experiment we have measured the localization accuracy inside a real virtual room, where there 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.

Test environment

Localization tests has been made in our virtual room [5] at Helsinki University of Technology. For spatial audio reproduction we use 15 Genelec 1029A loudspeakers. Alternatively headphones (Sennheiser 590A) can be used. Our audio reproduction system is described in article [6] in more detail.

Method

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. 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 reprocuded using one loudspeaker (LS), and 19 locations with VBAP reproduction.

In first two tasks there were no visual stimulus available on the screens. The order of headphone, and loudspeaker reproduction was randomized, in no visual stimulus, and with visual stimulus tasks. In the last two tasks the visual stimulus was a large 3D model of the estrogen receptor displayed on screens. (The locations of sound sources were not related with the molecule.)

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.

RESULTS

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 erros are shown in Table 1.

Table 1. Average and median values of azimuth and elevation errors for each reproduction type.

	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

We use an analysis of variance (ANOVA) model for the analysis. There was no significant difference between no-visual stimulus, and visual stimulus tasks. The visual stimulus was not related with sound source locations, which explains the result.

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 1.

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.

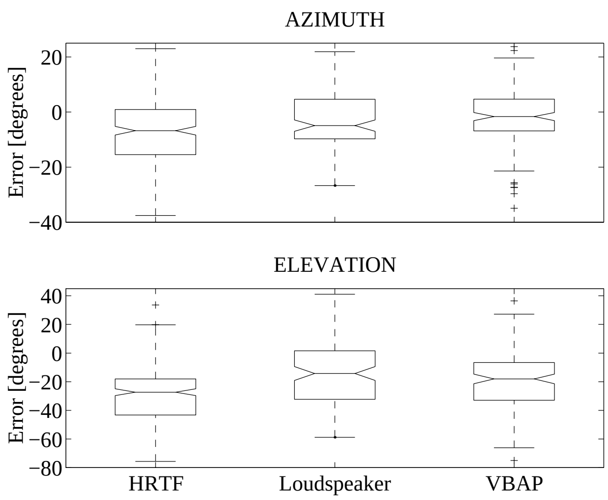

FIGURE 1. The results of the listening test with each reproduction system. The boxplot depicts the median and the 25%/75% percentiles.

CONCLUSIONS AND FUTURE RESEARCH

In a real virtual room with used reproduction methods the average azimuth error was 9.7 degrees, and elevation error 27.0 degrees. Non individualized HRTFs were significantly more inaccurate than loudspeaker reproduction methods. The VBAP was in this experiment as accurate reproduction method as direct loudspeaker reproduction. In this experiment the visual stimulus doesn't have on effect on localizaton accuracy.

Azimuth accuracy achieved in this experiment is convenient for our purposes. On the other hand more research should be carried out to find out how we can increase the elevation accuracy to convenient level.

Future research with other auditory stimuli should be conducted. Another area for future experiments will be testing the effects of integrated auditory and visual stimuli.

ACKNOWLEDGMENTS

This research was partly funded by Helsinki Graduate School in Computer Science and Engineering. The authors like to thank Juhani Forsman, and Tommi Ilmonen for their effort, and time with our audio reproduction system, and virtual environment.

REFERENCES

1. M. Gröhn, T. Lokki, L. Savioja and T. Takala, "Some Aspects of Role of Audio in Immersive Visualization", Proc. of SPIE, **4302**, San Jose, California, pp. 13 - 22 (2001).
2. V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning", Journal of the Audio Engineering Society **45** , nr 6, ss. 456-466 (1997).
3. 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**, ss. 1046 - 1053 (2000).
4. J. Blauert, "Spatial Hearing, The psychophysics of human sound localization.", The MIT Press, Cambridge, MA, USA, (1997).
5. J. Jalkanen, "Building a spatially immersive display - HUT-CAVE." Licenciate Thesis", Helsinki University of Technology, Espoo, Finland (2000).
6. J. Hiipakka, T. Ilmonen, T. Lokki, M. Gröhn and L. Savioja, "How to implement 3D audio in virtual environment", Proc. of Spie **4297B**, San Jose, California, (2001).