Measurements of auditory navigation in virtual acoustic space

Rudolf Susnik, Jaka Sodnik, Goran Bobojevic and Saso Tomazic
University of Ljubljana, Faculty of Electrical Engineering, Slovenia, Europe

ABSTRACT

Auditory navigation in virtual acoustic space is a well-known aid for helping blind people orient in space. Methods used for generating virtual acoustic space are also useful in reproduction of realistic sounds in surrond audio systems, video games, etc. In order to design an efficient virtual acoustic space or acoustic image, many psycho acoustic facts should be considered.

To prove theoretical presumptions in practice, many tests need to be done. We developed an experimental system for testing variety of parameters. Using the experimental system, one can virtually place the sound source in almost any location in front of the listener's head. The system uses either pre-recorded sound files or produces space sound in real time with the use of arbitrary spatial sound model (i.e. the way of presenting Head Related Transfer Functions – HRTF) and arbitrary sound sources.

The main goal of the research was to determine the possibilities of auditory navigation and to examine results of acoustic image resolution obtained in previous researches. Fifty test subjects underwent several experiments in which they were asked to locate the various sound sources. Experiments showed that the major problems were related to resolution. Test subjects were able to approximately find location very quick, but they had problems locating it more precisely.

1. INTRODUCTION

Sight is one of the most important human senses for orientation and the reception of information about spatial properties. However, we can also represent information about space in other ways. For example, visually impaired people are forced to substitute the missing information with other senses, for example hearing. The implementation of describing visual information by sound is called a virtual acoustic space or acoustic image of space. The acoustic image is suitable for a wide range of applications, for example, acoustic representation of the computer screen (i. e. graphic user interface) being just one of them.

The acoustic image is created by using spatial sound synthesis. Representation of computer screen with an acoustic image is a possible solution for development of an experimental system for acoustic image resolution measurements. Similar interfaces are also known as aids for blind users which they use for navigation over computer screen.

No matter variety acoustic aids for blind people, like graphic user interfaces for blind people (GUIB) already exist [8, 9], the goal of our research is to evaluate acoustic navigation ablities in acoustic images created with different models, to determine acoustic image resolution and to improve acoustic imaging methods. By the use of appropriate measuring tool we can perform measurements, considering variety parameters.

In this paper we present results of auditory navigation measurements using different spatial models, non-speech and speech signals. Since a useful measuring tool is necessary for performing experiments, we also concentrate on our experimental system and discusse some facts about spatial sound.

2. SPATIAL SOUND AND HEAD RELATED TRANSFER FUNCTIONS

The representation of spatial sound with the use of headphones is the key part of creating an acoustic image in our experimental system. This is done with the help of a set of HRTF (Head Related Transfer Functions) [5]. HRTFs are an empirical set of filters which describe the impact of the communication channel (i.e. channel between the sound source and ears) on the sound [6, 7]. HRTFs are realized with the use of different methods – they can either be measured, derived from different models etc. HRTFs are measured with the use of microphones inserted in the test subject’s ears. This means that the filters differ due to the shape of the listener’s head, size of the ear, shape of the ear canal, etc. Since these differences are usually not very important, a general set of selected HRTF model is used.

In our research, we used MIT Media Lab FIR (Finite Impulse Response) filters [1] and a resonator model of HRTF [2] realized with IIR (Infinite Impulse Response) filters. MIT Media Lab filters allow spatial sound synthesis for 710 different positions.

The resonator model of HRTF set is intendend to reduce the number of coefficients in standard set of FIR filters. The model used in this research [2] consists of a set of IIR filters and tends to imitate the amplitude characteristic of

MIT FIR filters, while the phase characteristic is realized with a minimum phase shift.

3. EXPERIMENTAL SYSTEM

The graphic user interface of the experimental system is presented as a panel defined with a 2-dimensional system of Cartesian coordinates – horizontal and vertical. Since the space around user's head is shaped as a sphere, transformation between Cartesian and spherical system of coordinates is required. The space of interest in our research and also in the experimental system is limited to the front half of the sphere. When the mouse pointer reaches the right edge of the screen, a sound coming from right is heard. Similar happens when the mouse pointer reaches the left edge. Since MIT HRTFs cover elevation in the range from -40^{o} up to 90^{o}, the situation in the vertical dimension is necessarily different. The vertical dimension of the computer screen is therefore transformed into the mentioned limits and does not regularly represent a rectangular graphic user interface's panel. Virtual space covered by the experimental system is depicted in figure 1.

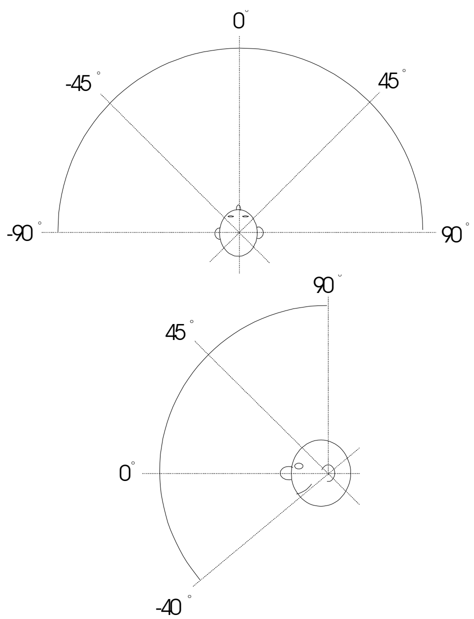

Fig. 1: Virtual space covered by experimental system.

The key object for navigation over the screen is mouse pointer, so the temporary position of the mouse should be known every moment. The position is signalized with virtually positioned sound source which changes its location according to the mouse's moves. Since a certain fixed time is needed to reproduce any sound, the problem of sound source choice becomes important. Fast mouse moves seriously reduce the resolution, especially when the duration of the sound is long.

As mentioned before, the space around the listener is shaped as a sphere. Figure 2 depicts a part of our experimental system – the front left quarter of the sphere. The white frame with a net of coordinates represents the virtual space in which the user can move his sound source. Due to research purposes, the system also has some additional functions; for example, the temporary position of the mouse pointer is displayed and the users can choose between several HRTF models and arbitrary sound sources (recorded into .wav file). The most common test is a case where sound is placed into a fixed position (reference point) and user then try to reach that position by listening to the sound and moving the mouse. It is also possible to adjust the reference point's neighborhood; i.e. when the user moves the mouse into the vicinity of the reference point, a sound appears signaling that a reference point was found. Another feature of the experimental system is the record of the mouse pointer's trace. Pressing the start button resets the trace counter and starts recording the mouse pointer trajectory. This is useful in to observe the length of path (trajectory) needed to reach the reference point.

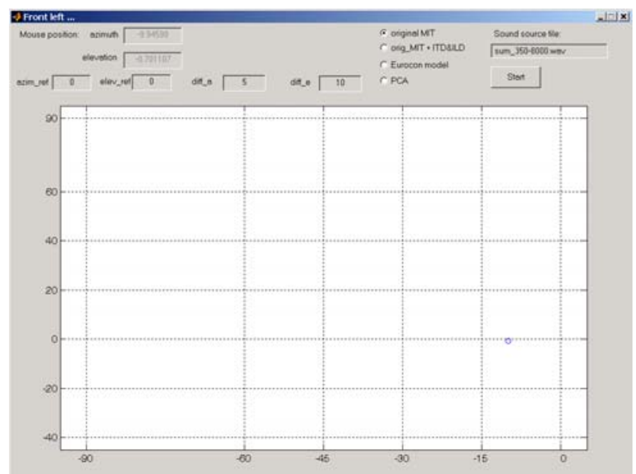

Fig. 2: Graphic user interface of the experimental system.

Virtual positioning of any sound source in an arbitrary position could be done by different methods. Real time filtering can be performed when there are no special processor limits for data processing. Obviously, this is the most flexible way, since any sound source and any HRTF model can be used. Unfortunately, such real time processing sometimes fails and makes delays even on the fastest computers. Another way is playing pre-recorded .wav files, which demands quite a large disk space. A typical .wav of noise lasting 500ms or less used for acoustic image tests is about 50kB long.

4. MEASUREMENT PROCEDURE

Many papers deal with graphic user interface for blind people [8, 9, 10] implemented by means of acoustics. But they do not report detailed information about navigation abilities in conjunction with acoustic image resolution.

We performed an experiment of navigation abilities in acoustic image of the computer screen. The experiment was performed on a sample of 50 test subjects – secondary school students (15–19 years old) with normally developed sight and hearing and no previous experience with virtual spatial sounds. The aim of the experiment was to check the possibilities for acoustic navigation. Measurements presented in this paper are continuation of experiments given in [4].

First, all subjects were introduced to spatial sound played through the headphones. They were introduced to the experimental system and they were asked to test the system on their own by moving the mouse over the computer screen. The purpose of this exercise was to acquaint the test subjects with spatial sound played through the headphones. We used different sound sources and different spatial models (MIT HRTF, resonator model).

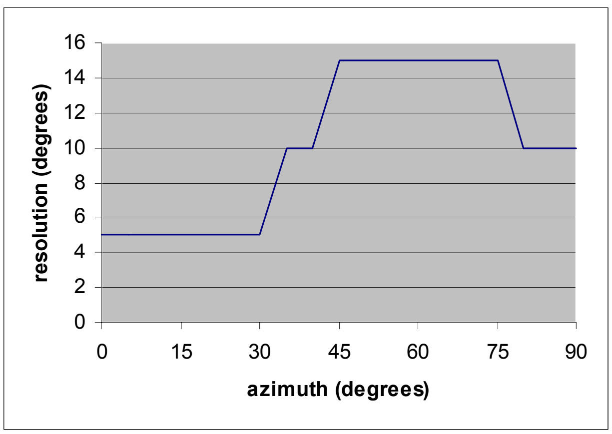

Fig. 3: resolution in dependence of azimuth

The second part of the experiment was to check test subjects' abilities to locate a sound source located in arbitrary virtual location. The experimental system randomly selected a virtual position and played a sound from that position – the so-called reference point. All reference points were selected using uniformly distributed random function. Test subjects were asked to locate the sound source by moving mouse pointer from the origin (azimuth 0^{o} and elevation ^{o}) to the reference point as straight as possible. While the test subjects were moving the mouse they heard a sound coming from the current position of mouse pointer. They were also able to refresh the knowledge of reference point by clicking on the right button of the mouse. The path (trajectory) of the mouse was recorded and then used as a measure to value navigation abilities.

The reference point's neighborhood (i.e. how close to the reference point mouse pointer must be to be recognized as a hit) was chosen according to the previous researches of acoustic signal localization [4]. Figure 3 depicts range of neighbourhood in dependence of azimuth. As we can see, resolution is the best in front of the listener. Since lack of information about resolution in vertical dimension we stated 20^{o} for all cases.

5. RESULTS

Three different sound sources were used in all test procedures:

- white noise, duration 20 ms;
- white noise, duration 200 ms;
- speech signal.

Spatial sounds were created using two different spatial models:

- MIT HRTF FIR filters [1];
- Resonator model [2].

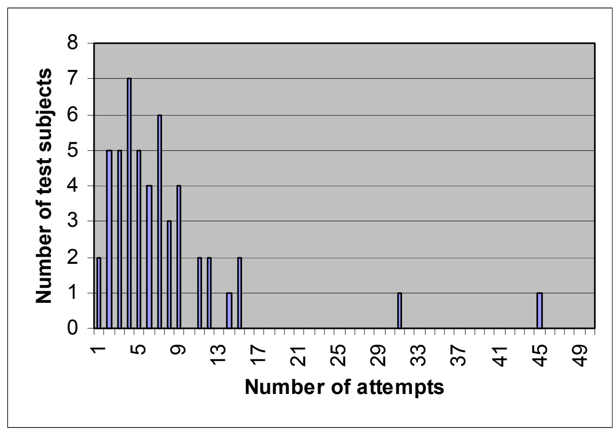

Fig. 4: Number of attempts needed to locate white noise (model: MIT HRTF).

The results of navigation ability tests are gathered in figures 4, 5 and 6. Graphs in figures 4 and 6 represent results, achieved by using the white noise signal with duration of 200 ms and different spatial models (MIT HRTF FIR filters, Resonator model), graph in figure 5 shows results for speech signal, spatialized through the use of MIT HRTF FIR filters.

The length of the trajectory in the test means the number of attempts to locate the reference point. The test subject's first attempt depended on a subjective estimate where the reference position might be, therefore describing absolute resolution. The subsequent attempts tried to approach reference position. Of course, every single test subject has a different strategy of searching the reference point. Two very different strategies were noticed – short moves from the origin (point of first attempt) toward the reference point or large moves which often resulted in rotating around the reference point.

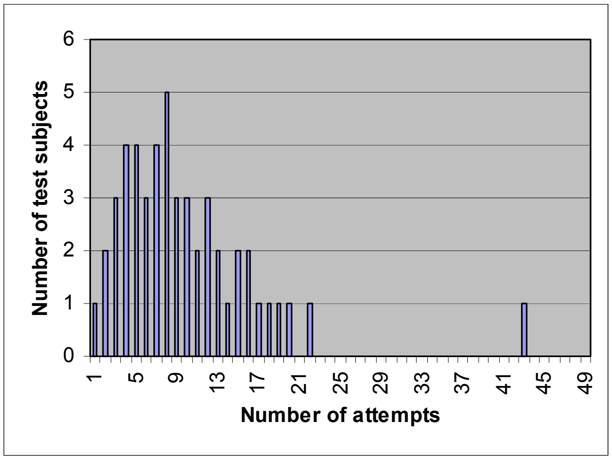

Fig. 5: Number of attempts needed to locate speech signal (model: MIT HRTF).

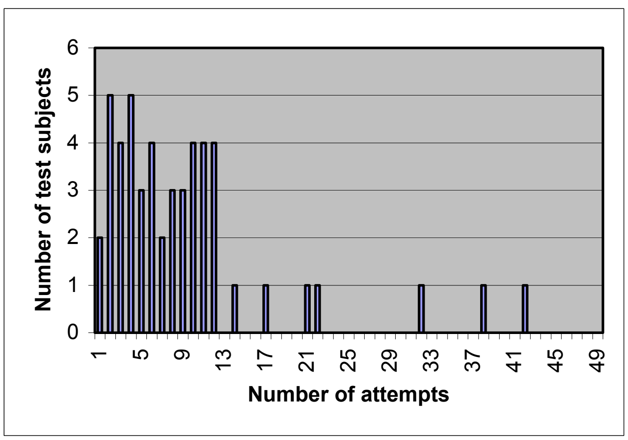

Fig. 6: Number of attempts needed to locate white noise (Resonator model).

Results show that majority (2/3) of test subjects were able to locate reference point in less than 10 attempts for all test cases. The number is quite great, but it is also necessary to take into consideration test subjects had no previous experiences with spatial sound played through headphones. As found in previous researches [2] people are able to approximately define the direction of sound in the first attempt. But they also need some further attempts to precisely localise sound. This fact represents a pretty large limitation, so in further researches it would be necessary to value acoustic image resolution for such situations and find a way to give optimal amount of information to the user.

6. CONCLUSION

Results prove perspective ideas for auditory navigation using acoustic image. While evaluating results we need to take into consideration the limits of acoustic image and test subjects' previous non-experience with acoustic image, the main barrier established until today is the insufficient resolution. In addition, it would also be necessary to precisely adapt coordinates. HRTFs are measured in a spherical system of coordinates which does not simply comply with the rectangular shape of our measuring tool.

7. REFERENCES

[1] B. Gardner, K. Martin, "HRTF Measurements of KEMAR Dummy-Head Microphone", MIT Media Lab Perceptual Computing – Technical Report #280, May 1994.

[2] R. Susnik, J. Sodnik, A. Umek, S. Tomazic, "Spatial sound generation using HRTF created by the use of recursive filters", Eurocon 2003, vol. 1, pp. 449-453, Slovenia, September 2003.

[3] K. Itoh, M. Shimizu, "GUI Objects Represented by New Localized Sounds using HRTF", HCI International 2003, Greece, June 2003.

[4] J. Sodnik, R. Susnik, S. Tomazic, "Acoustic signal localization through the use of Head Related Transfer Functions", CCCT '03, pp. 100-103, Florida, July 2003.

[5] C. I. Cheng and G. H. Wakefield, "Introduction to head-related transfer functions (HRTF's): representations of HRTF's in time, frequency, and space (invited tutorial)", Journal of the Audio Engineering Society, vol. 49, no. 4, pp. 231-249, April 2001.

[6] G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane", J. Acoust. Soc. Am., vol. 62, no. 1, pp. 157-167, July 1977.

[7] V. R. Algazi, C. Avendano, R. O. Duda, "Elevation localization and head-related transfer functions analysis at low frequencies", J. Acoust. Soc. Am, vol. 109, no. 3, pp. 1110-1122, March 2001.

[8] P. B. L. Meijer, "Seeing with Sound for the Blind: Is it Vision?", Conference on Consciousness, Tucson, Arizona, USA, April 8, 2002.

[9] P.B.L. Meijer, "An Experimental System for Auditory Image Representations", IEEE Transactions on Biomedical Engineering, Vol. 39, No. 2, pp. 112-121, Feb 1992.

[10] T. Lokki, M. Grohn, L. Savioja, T. Takala, "A Case Study of Auditory Navigation in Virtual Acoustic Environments", ICAD 2000, USA, April 2000.