Acoustic signal localization through the use of Head Related Transfer Functions

Jaka SODNIK, Rudolf SUSNIK and Saso TOMAZIC
University of Ljubljana, Faculty of Electrical Engineering,
Laboratory for Communicational Devices,
Ljubljana, Slovenia

ABSTRACT

An acoustic image of space is an acoustically described visual image intended to help blind people orient themselves in space. Description is made with the aid of spatial sounds created using HRTF filters. HRTF filters are empirically acquired FIR filter sets that describe changes to the sound as it travels from its source towards the human eardrum. They include changes related to body shape, ears, ear canal, etc. Our research focused on finding the maximum resolution of the human auditory system when determining the location of a sound source in space. This is also the maximum resolution for creating an acoustic image. We were interested in minimum azimuth and elevation change resolution – we tried to establish the minimum angle between two sources that could still be detected. Resolution dependence on signal bandwidth was also measured. The results were encouraging, especially in the horizontal plane, where most of subjects were able to tell the difference between two sources only 5° apart. Edge resolution, with 80° – 90° azimuth, was still satisfactory if a wide bandwidth signal was used. If elevation is increased, the resolution deteriorates quickly and is no longer satisfactory. To address this problem, different coding should be used to create an acoustic image of elevation.

Keywords: Acoustic image, HRTF, spatial sound, spatial resolution, subspace.

1. ACOUSTIC IMAGE

Sight is one of the most important human senses for orientation and the acquisition of information about spatial properties. Blind people are thus forced to substitute the missing information with other senses. A new approach to the problem is to describe visual information by means of sound. This is called an acoustic image of space. An acoustic image is created with spatial sound synthesis, done in a way that most efficiently describes the visual image. Image is first captured (camera, radar, sonar) and converted to sound. One of the practical systems uses temporal scanning of the image from left to right. Azimuth (horizontal direction) is encoded with time delay, elevation (vertical direction) with sound frequency, and distance with sound amplitude. This is quite efficient, but unnatural to the brain. Our approach is to divide the visual field into subspaces, with boundaries set at -90° (left) to 90° (right) azimuth, and -40° (down) to 90° (up) elevation. An acoustic image is created with random subspace selection and the determination of distance to objects in the subspace. The probability of choosing a certain subspace is defined by means of a suitable dispersion function. This should ensure greater probability to spaces right ahead, in the centre of the image, and less probability to spaces at the boundaries of the image.

2. HRTFs

Spatial sound synthesis, a key part of creating an acoustic image of space, is done with the help of a set of HRTF (Head Related Transfer Functions) [3] [4].

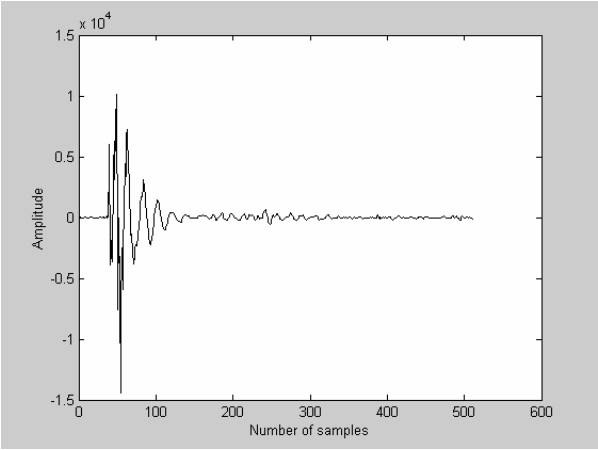

Figure 1: HRTF filter

This is an empirical set of FIR filters enabling the synthesis of a spatial sound that can be played through headphones. It describes changes to the sound as it travels from its source to the human eardrum. This means of course that the filters are different according to the shape of the head, size of the ear, shape of the ear canal, etc. These differences are usually not very important to the listener, so a general set is used. Spatial sound is created by filtering a sound signal with an appropriate HRTF pair (one for the left and one for the right ear).

We used MIT Media Lab (Cambridge) FIR filters for our research. They allow spatial sound synthesis for 710 different positions. The filters were measured in an anechoic chamber using a KEMAR dummy head with sensitive microphones in both ears. It was mounted on a rotating table, enabling measurements for different incident angles. The test signal consisted of a delta impulse, enabling researchers to use the measured responses directly as filter characteristics. Measurements were sampled at 44.1 kHz. The 0° elevation measurements (most useful to us) consisted of 72 FIR filters with 512 coefficients, enabling us to simulate spatial sound for a complete radius of 360° with 5° resolution. With bigger and smaller elevations, the number of sampled angles decreased.

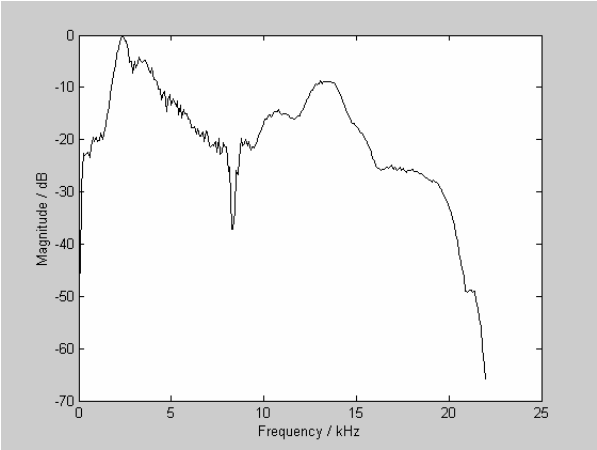

Figure 2: HRTF filter spectrum

Our research primarily investigates the ability of the human auditory system to distinguish the location of a sound source. Previous research has mostly examined the ability to pinpoint directions and the factors that influence our ability to pinpoint the position of a sound source.

Key factors influencing the process of localisation of a sound source are [2]: inter-aural time delay, inter-aural level difference, the inter-aural spectrum, the mono-aural spectrum, etc. The importance of these factors depends primarily on the signal spectrum. The most appropriate signal for measuring these factors is white Gaussian noise, filtered with appropriate band pass filters and divided into frequency bands: 0.3–5 kHz, 0.3–7 kHz, 0.3–10 kHz, etc. Inter-aural factors seem to be the most important for positioning a sound source, regardless of the type of source and acoustic environment.

3. DESCRIPTION OF THE PROBLEM

This paper describes the measurement of the maximum resolution of the human auditory system when trying to determine the position of a sound source. This information is needed to divide visual space into subspaces in the best possible manner. Resolution is the minimum distance between two sound sources that can still be distinguished even though the signal from both sources is the same.

4. MEASUREMENTS

An experiment was done on 70 subjects of all age groups (15–60 years old) with normally developed hearing and sight and no previous experience of virtual spatial sounds and HRTFs. The test environment was developed using the Matlab programming language.

The test signal was filtered white Gaussian noise (600 ms long) using four band pass filters: 350–2800 Hz, 350–8000 Hz, 2000–8000 Hz and 1000–8000 Hz.

Resolution was measured in the space that is most important for acoustic imaging – the visual field (azimuth of -90° to 90° and elevation of -45° to 90°). Twenty-five points were obtained to position our virtual sources:

- azimuth: 0°, -45°, 45°, 90° and -90°
- elevation: 0°, -45°, 45°, 90° and -90°

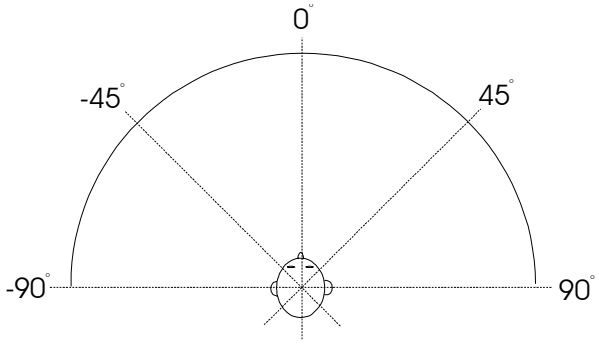

Figure 3: Azimuth

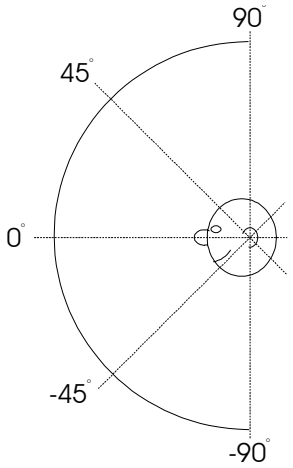

Figure 4: Elevation

Prior to measurements, all subjects were introduced to spatial sounds played through headphones. The introduction consisted of three parts:

- random sound playback from azimuth of 0°, 90° and -90° (intended to introduce HRTFs to the listener)
- random sound playback from azimuth of 45°, 0° and -45° (intended for the listener to get a sense of direction)
- sequential sound playback from azimuth of 90°, 45°, 0°, -45°, -90°
- sequential sound playback from azimuth of 0°, 15°, 30°, 45°, 60°, 90°
- sequential sound playback from azimuth of 0° and elevation of -20°, 0°, 20°, 40°, 60°, 90°

The resolution of the human brain was determined with each subject trying to find the correct sound source from several virtual sources a small distance apart sequentially playing the same signal.

The first part of the resolution measurements was done in the horizontal plane (0° elevation), with main virtual sound sources at azimuth of 0°, 45°, -45°, 90° and -90°. Neighbouring sources were ±5°, ±10°, ±15°, ±20°, ±25° and ±30° apart.

The second part of the resolution measurements was done with 50° elevation and with the same azimuths as in the first part. The smallest distance between sources was 8°, due to insufficient HRTFs resolution (filters would have to be approximated to achieve better resolutions).

The third part of the resolution measurements was done with 90° elevation (above the head). Sources were placed at azimuths of 0°, 30°, 60° and 90°.

The last part of the resolution measurements was done with 0° azimuth. Sources were placed at elevation at 0°, 10°, 20°, 30°,-10°,-20° and -30°.

5. MEASUREMENT RESULTS

Test subjects were exposed to spatial sound through headphones for the first time, which resulted in a very poor spatial awareness. Most problems were caused with front/back confusion. We believe this happened because all subjects were normally sighted. The fact that the sound sources were not visible confused them into feeling that the sources were behind them most of the time.

The first part of the measurements (horizontal plane - 0° elevation) was very encouraging: 87% of subjects were able to differentiate sound sources only 5° of azimuth apart.

	Successful perception		
	Main sound source at		
Source distance in ^{0}	Azimuth 0^0	Azimuth 45^0	Azimuth 90^0
30	100 %	100 %	100 %
25	100 %	100 %	100 %
20	100 %	100 %	100 %
15	100 %	95,7 %	100 %
10	100 %	91,3 %	95,7 %
5	87,0 %	39,1 %	52,2 %

Table 1: Perception in horizontal plain

Interestingly, resolution at 90° azimuth is better than at 45° azimuth.

The second and third parts of the measurements were done with elevations of 50° and 90° and the same azimuths as in the first part.

	Successful perception		
	Main sound source at		
Source distance in ^{0}	Azimuth 0^0	Azimuth 45^0	Azimuth 90^0
48	100 %	100 %	100 %
40	100 %	100 %	100 %
32	100 %	100 %	100 %
24	100 %	95,7 %	93,2 %
16	89,4 %	76,3 %	76,7 %
8	58,7 %	17,2 %	10,5 %

Table 2: Perception at elevation 50^0

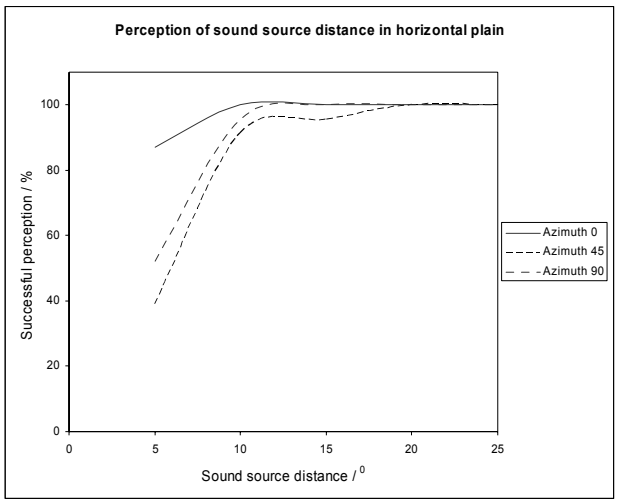

Figure 5: Perception in horizontal plain

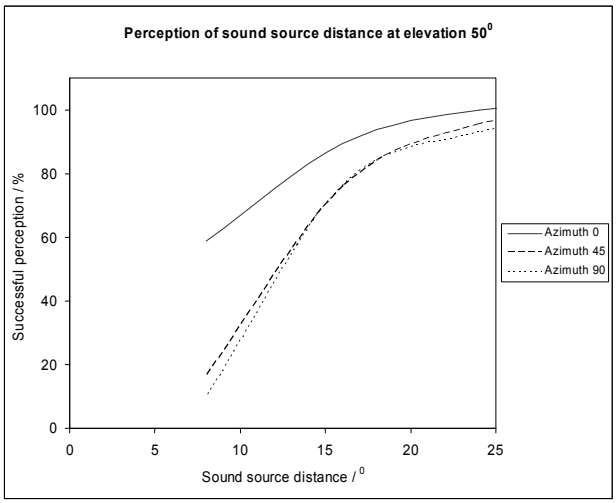

Figure 6: Perception at elevation 50^0

	Successful perception		
	Main sound source at		
Source distance in ^{0}	Azimuth 0^0	Azimuth 45^0	Azimuth 90^0
60	24,3 %	12,5 %	0 %
30	0 %	0 %	0 %

Table 3: Perception at elevation 90^0

The signal bandwidth (the pass band of the filter that was used on the white Gaussian noise) had a big impact on perception ability. In the centre of the horizontal plane, an 80% success rate was achieved with a quite limited spectrum (350–2800 Hz). The boundaries frequently required a relatively broad band signal (2000–8000 Hz or 1000–8000 Hz) for the same success rate.

6. DISCUSSION AND CONCLUSION

The best results were achieved in the horizontal plane with an azimuth of 0°. This is also the centre of the human visual field, which is advantageous for an acoustic description of the visual image. An acceptable resolution for this coding is at least 5°, and can even be increased with certain limitations in signal choice. The resolution is much lower at the boundaries (azimuth of 80–90°). The subjects tested were mostly able to tell the difference in the sound, but were unable to determine the position of the source (direction of change of azimuth or elevation) with any precision.

The resolution also reduces with increasing elevation. At an elevation of 45°, the source direction changes are still detectable, but the resolution decreases to 20–25°. At a 90° elevation, the changes were no longer detectable.

We also found that several repetitions of a test considerably improved the results because the subjects got used to the virtual spatial sounds.

Acoustic imaging with the aid of HRTFs can be sufficiently accurate for practical use, especially in the horizontal plane. Azimuth (positions to the left and right) can be efficiently coded, but elevation coding (positions up and down) does not provide enough resolution; another way of coding it must therefore be found. We found that most subjects reported the higher frequency signals to be positioned higher (at high elevations). This fact could be used to code the elevation with signal frequency. On the other hand, this would be disadvantageous because the narrower bandwidths of the signals reduces the azimuth coding resolution.

7. REFERENCES

[1] Bill Gardner and Keith Martin, **HRTF Measurements of a KEMAR Dummy-Head Microphone**, MIT Media Lab Perceptual Computing – Technical Report #280, May, 1994

[2] Andre van Schaik, Craig Jin, and Simon Carlile, Computer Engieering Lab, Department of Electrical and Information Engineering, Auditory Neuroscience Lab, Department of Physiology, **Human Localization of Band-Pass Filtered Noise**, EWNS 1999

[3] Corey I. Cheng and Gregory H. Wakefield, Introduction to Head-Related Transfer Functions (HRTF'S): **Representations of HRTF's in time, frequency, and space (invited tutorial)**, University of Michigan, Department od Electrical Engineering and Computer Science, Ann Arbor, Michigan, U.S.A.

[4] V.R. Algazi, R. O. Duda and D. M. Thompson, **The Cipic HRTF Database**, October 2001, New Platz, New York

[5] V. Ralph Algazi, University of California, Davis, Pierre L. Divenyi, Speech and Hearing Research Facility, V.A. Martinez, Richard O. Duda, San Jose State University, **Subject Dependent Transfer Functions in Spatial Hearing**, 1997

[6] Pavel Zahorik, Doris J. Kistler, and Frederic L. Wightman, Waisman Center, University of Wisconsin – Madison, **Sound Localization in Varying Virtual Acoustic Environments**, Proceedings of the Second International Conference on Auditory Display, ICAD 1994

[7] Dieter Beheng, **Sound Perception**, Deutche Welle, Radio Training Centre, DWRTC Cologne 2002

[8] M. Cowling, R. Sitte, **Sound Identification and Direction Detection in Matlab for Surveillance Applications**, Griffith University, Faculty of Engineering and Information Technology, Queensland, Australia