Comparing orientation in simple HRTF-based virtual auditory space between sighted and visually impaired

Rund František · Elektrotechnika, Medicína

24.01.2011

Using own measured HRTF for each user is considered as the best way to obtain a high-quality perception of sound source virtual position. Without proper equipment the measuring process is very time consuming, so we decided to use the virtual auditory space with small resolution only at first. In this work, we got the HRTF in 15 positions for 6 different subjects with using MLS signal. Then a series of virtually positioned sound were created using MATLAB algorithm. Necessity of equalization the room, loudspeaker and headphones for later listening is also discussed.

Finally, we made a test sequence fixed to verify orientation in virtual acoustic space. Each subject had to listen to the sequence and mark positions in questionnaire, where he considers the sound source is located. Unknown to subject each sequence was doubled for estimation whether the subject is sure about sound source position or guesses. Wideband noise bursts modulated by low-frequency sine were chosen as stimulus. All 6 subjects whose HRTF was obtained were with no sight disability. Last subject was completely sightless, he did not attend the measuring procedure and HRTF sequence of another subject had to be used instead.

1. Introduction

Creating the three-dimensional sound for entertainment, commercial, and scientific systems by using HRTF is well-known nowadays. There have been many researches due to reach the best quality in virtual sound source positioning. HRTF is a complex function, which captures the spectral changes of signal that occurs when a sound wave propagates from sound source to the listener´s middle ear. The listener evaluates the spectral content and latency of both signals from left and right ear, and according to this estimates the location of sound source. Head-Related Transfer Function depends on frequency, azimuth, elevation, range and it also significantly varies from person-to-person [2]. We can write it down as H = f(φ, θ, ω, r, subject).

HRTF can be transferred through inverse Fourier Transformation into time domain, where it is represented as HRIR (Head Related Impulse Response), which represents the impulse response of path between sound source and entrance to the left (right) ear. If we know the HRIR for both channels, we can create a stereo signal from monaural source as implied for left channel by Eq. (1). It is necessary to keep both

channels separate, therefore headphones have to be used.

$$y_L(t) = HRIR_L(t).x(t)$$ (1)

HRTF is subjectively depended, so in general everyone must have own set of HRTF from required directions. There are two basic ways how to get personal HRTF. At first, we can measure a whole set. A disadvantage is that we need a net of measured points with sufficient density, so it requires a lot of time spending with measuring process. However, this approach provides the best results in final perception of sound. Other way is to create a mathematical model according to simple-measured anthropometrical parameters. Nowadays, models are constantly improved with good results, but in terms of sound source perception quality is worse than measured set.

In this work we got HRTF in 15 positions (5 azimuth and 3 elevation levels) for 7 different subjects. Measurement points were selected only in frontal area with 45 degrees step in horizontal and median plane from the center of view (θ = 0°, φ = 0°), as shown on Fig. 1. As the first step, it is enough positions for basic resolution in virtual auditory space, because we can combine directions of right / left and up / down.
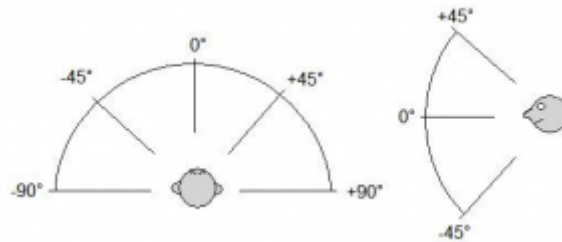

Figure 1: Measured points only for frontal area

2. Measurement of HRTF set

In order to get a proper HRIR set, we arranged a simple measuring set, which consists of tiltable loudspeaker mounted on extensible stand, swivel chair with calibrated pointer, two small microphones, sound card, amplifier, and measuring software. For particular types see Table 1. In best way all measurements of HRTS should be done in anechoic chamber, but we considered available baffled studio sufficient for our experiment.

Each subject was seated on the swivel chair, and both microphones were attached by plaster on the beginning of ear canal. For reducing the influence of canal cavity we used medical earplugs. At the beginning we asked the subject not to move and look straight ahead. After measuring process subject turns for another 45 degrees. This was done for 3 elevations of loudspeaker.

EASERA software using MLS measuring signal [3] was used for obtaining HRIR. The data was stored as stereo wav files and then processed in MATLAB 2006a version. It was necessary to check microphone position permanently and keep the system balanced, because even small position change caused incorrect microphone gain, what makes well perceptible offset in final virtual sound positioning. Measuring time for one subject was app. 40 minutes. More details of our HRTF measurement can be

found in [9].

Figure 2: Measuring of HRTF – a) whole measuring set b) detail on microphone attachment.

3. Equalization of measured HRTF

When we obtain HRIR, it includes even room response for measuring MLS signal, so HRIR is actually multiplied by strongly attenuated but still distinct reflections. We designed the in-room measurement for first reflection coming from the floor. It is necessary to eliminate all samples of HRIR after first reflection comes. That causes explicit HRIR distortion, but this variant is more accurate than the one with reflections included. There are different times of the reflection arrival, but we were operating only with average. It finally led to app. 8.5 ms duration of HRIR (from beginning) what is about 820 samples in using 96 kHz sampling frequency as shown on Eq. (2).

$$HRIR\ length = \frac{reflected\ sound\ efect}{sound\ velocity} \times sampling\ frequency \qquad (2)$$

Another adjustment is then needed to compensate influence of loudspeaker, which

distorts flat spectra of measuring MLS signal. The same compensation is also needed for headphones transfer function compensation, because this characteristic distorts HRTF too. In general, we have to compensate all elements presented during measurement and binaural listening between sound source and listener. All these adjustments were made in frequency domain by dividing Fourier transformation of measured HRIR with appropriate inverse transfer function.

Last step in creating virtually positioned sound is convolution between input signal and both appropriate adjusted HRIRs for left and right channel. After it both monaural signals are put together into one stereo wav file. Whole procedure of creating virtually positioned sound [out] from monaural source [x(t)] is depicted on Fig. (3).
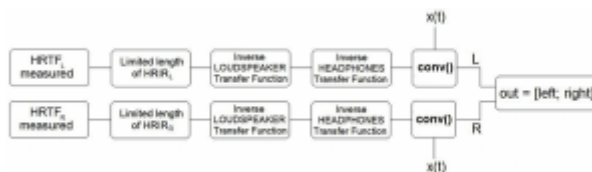

Figure 3: Scheme of HRTF equalization and generating stimuli

4. Stimuli creating

A wideband stimulus is considered as the most suitable for the best results in sound source position perception, because every frequency band is involved by different principle of localization cues [1]. Using a wideband stimulus we combine all localization mechanisms, so the final perception is supposed to be more specified. In [6] is used a White Gaussian Noise modulated (WGN) by sin(40Hz). Modulation causes constant amount of leading edges, which also improves sound source localization. Stimulus used in this experiment is depicted on Fig. (4).
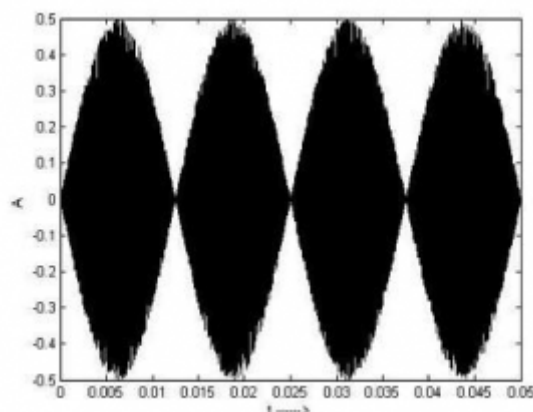

Figure 4: Modulated noise stimulus in time domain

After filtration with HRTF resp. convolution with HRIR the output signal spectra (both channels) is uniquely shaped according to appropriate HRTF for desired direction. Spectral notches and peaks can be seen on characteristic frequency bands. Filtration and shaping of narrowband spectra of WGN stimulus for one channel and one direction is shown on Fig (5).
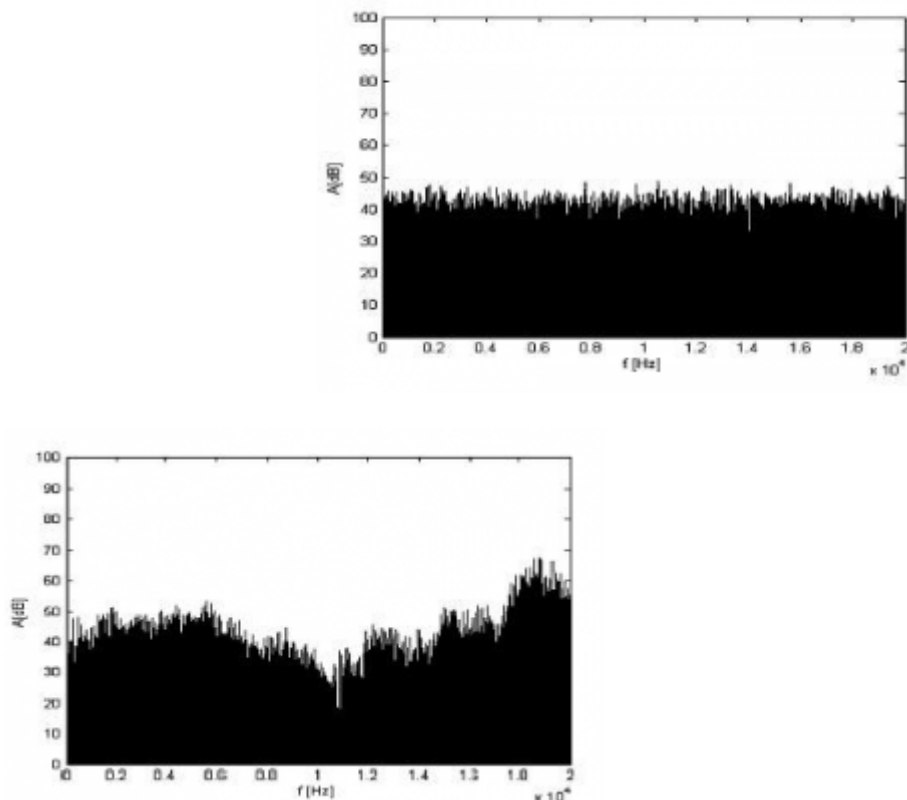
Figure 5: Changing spectral parameters of noise-like stimulus using HRTF

5. Subjective test of sound source location perception

For final testing of perception a sequence of positioned WGN wav files was created. The sound source was virtually positioned (ideally) into the same locations, where all 15 HRIRs were measured. We tried to verify whether compensation of headphones and loudspeaker is really needed and how it actually affects the final perception. All sequenses were made in 3 variants: measured HRTF, compensated loudspeaker, and compensated headphones and loudspeaker. We used AKG K 55 headphones for this test.

The Virtual Auditory Space we used can be represented as shown on Fig. (6). Every subject firstly listened to the tutorial sequence, which went through all 15 points in order: A1, A2, A3, B1, B2, ... and so on. We considered this important, because informal test shown that first contact with virtually positioned sound can make the subject confused. On this tutorial sequence the subject was allowed to set the volume to feel comfortable.

The final orientation test consists of 30 virtually positioned sounds. Every sound was introduced by 1 kHz tone non-positioned signalization "beep" and three times repeated. After that the subject had 6 seconds to fill a gap on questionnaire with number of a sample. Each following sample in sequence had to differ at least in one step in elevation and one in azimuth for bigger subjective sound source movement. After first 15 samples, when each position occurred, the whole sequence was repeated without subject´s knowledge. Comparing results in both same sequences tells more whether the subject guesses or is sure about virtual sound source position.
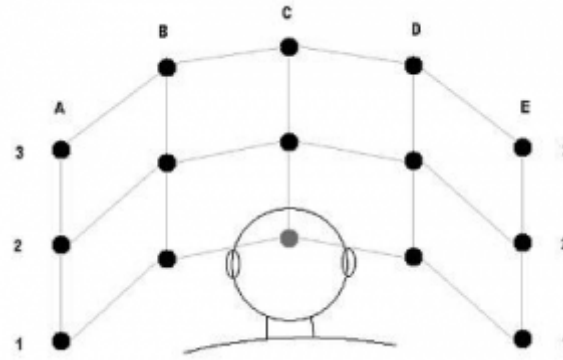
Figure 6: Virtual auditory space scheme – rear view

Results of this task were not as good as we predicted. We thought about almost 100% accuracy because of quite big distances between measuring points, but it was only 10 – 46%. In Table 2 we can see factual information. During the tests we took notice of very strong sensitivity on microphone gain offset. It is necessary to take care of gain balance in the beginning of measurement in "C" position – both microphones must be symmetric fixed, because even app. 3-4 mm deviation causes as many as 20° error in localization in azimuth plane.

All subjects were able to distinguish side of incoming sound, but results in median plane were not so precise. Columns "azim." and "elev." in Table 1 show RMSE for every subject for both planes. Values are related to step of 45°. In sequence 2 (with compensations) an externalization effect [4] is perceptible when we hear the virtual sound source out of the head, so it makes the source more "real" even the stimulus is only wideband noise. Subjects DS and MB, who were also authors, show better results although they didn´t know the sequence order in advance. We think it is important to get used to virtual positioning and sound character first for improving orientation in virtual acoustic space.

Table 1: RESULTS OF ORIENTATION IN VIRTUAL AUDITORY SPACE

subject	Sequence 1			Sequence 2		
	RMSE_az	RMSE_el	correct	RMSE_az	RMSE_el	correct
SM	0,66	1,15	4	0,73	1,11	6
BM	0,71	0,88	10	0,58	1,05	7
PS	0,98	0,8	6	0,63	1,18	3
TS	0,86	1,11	5	0,93	1,13	3
BK	0,88	1,03	6	0,73	1,24	6
DS	0,48	0,84	10	0,48	0,58	14
ZB	0,18	0,97	11	0	0,93	10
average	0,68	0,97	7,4	0,58	1,03	7,0

Finally the test sequences 1 a 2 were tested on visually impaired subject ZB. It was impossible to let him filling a questionnaire so the subject was asked to imagine a net of real sources, as shown in Fig. 6, and determine the coordinates after each stimulus. Subject had also to show predicted position by hand. This test was recorded on CCD

camera for later assessment. A scheme of measuring workplace is shown on Fig. 7, real picture of measuring setup on Fig. 8.
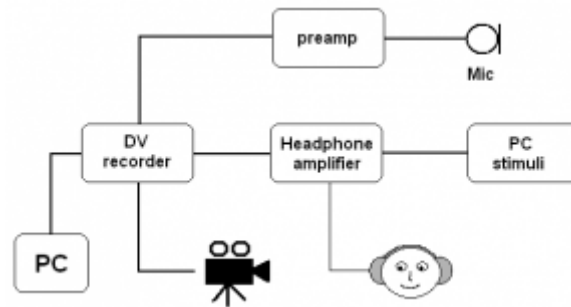

Figure 7: A scheme of measuring workplace

Results of testing subject ZB were surprising. Although he was not able to see a scheme of HRTF measuring points, as in Fig. 6, his position determining was very precise. These results are marked red in Table 2. There was no mistake in azimuth plane for sequence 2, only one mistake in sequence 1 and elevation error was almost an average. This could be caused by using sequence personalized for subject DS, because subject ZB used "other´s" pinna, which is just the most important body part in localization cues in median plane. These results also confirmed the subject has to be used to extract important spatial information even only through hearing system.

Also an "acoustic pointer" operated by joystick was presented to visually impaired subject ZB. It allows the user to rotate a virtual sound source in azimuth plane in range 0 – 360 degrees. The subject noticed it is very easy to point on real sound sources in Virtual Auditory Space (VAS) because of the virtual source is externalized by HRTF. More details can be found in [7]. b

Figure 8: Two pictures taken during the test with visually impaired subject ZB – a) subject was monitored by CCD camera for later assessment of the sequence b) moving in VAS with joystick

6. Results

Head Related Transfer Function was measured for 15 locations on 7 subjects. HRTF is very sensitive for any gain offsets, so the system configuration has to be permanently checked. After that test sequence of noisy-like stimulus using unique HRTF set for each subject was made in order to verify orientation in Virtual Auditory Space. Differences in perception for three types of equalization were tested (none, loudspeaker, loudspeaker + headphones). Compensation brings an externalization effect, which moves the sound source perception out of subject´s head, so it makes the source more real. It also improves the resolution in virtual space, but in out experiment it was not proved because of unwanted gain offset.

The same test was made on visually impaired subject ZB which had very good results in spite of his test sequence was used from another subject. Presence of the externalization effect in using equalized sequence was also noticed.

Final results of this experiment were not sufficient, because we expected much more precise orientation with certain position determination for sighted subjects. Now we want to extend this experiment by more precise and dense measuring of HRTF with using a head-tracking system, because possibility to make head movements during localization which shift the sound source position improves subject´s estimation [4]. Also possibility of "learning-to-hear" virtual positioned sound, as mentioned in section 5, is desirable to verify. This research is aimed to develop interfaces of assistive technologies (image sonification, virtual navigation, …) for visually impaired.

Acknowledgements

References

1. Wenzel, E. M., Arruda, M., Kistler, D. J., Wightman, F. L., „Localization Using Nonidividualized Head-Related Transfer Functions," J. Acoust. Soc. Am., vol. 94, pp. 111-123 (July 1993)
2. Algazi, R., Aveando, C., Thomson, D., "Dependence of Subject and Measurement Position in Binaural Signal Acquisition," J. Audio. Eng. Soc., vol. 47, no 11, pp. 937- 947, Nov. 1999
3. Kadlec, F., "Zpracování akustického signálu", ČVUT FEL, Praha 2002
4. Wersényi, G., "Localization a HRTF-based Minimum-Audiable-Angle Listening test for GUIB applications," in Electronic Journal «Technical Acoustics», 2007
5. Susnik, R., Sodnik, J., Tomazic, S., "Measurements of Auditory Navigation in Virtual Acoustic Space," Perceptual Interfaces and Reality Laboratory, UMIACS, University of Lubljana, Slovenia, 2004
6. Susnik, R., Sodnik, J., Tomazic, S., "Sound Source Choice in HRTF Acoustic Imaging", University of Lubljana, Slovenia, 2003
7. Rund, F.: Audio Pointer – Joystick & Camera in Matlab, Technical Computing Bratislava 2010, 2010, pp. 1-3
8. Rund, F., Štorek, D., Glaser, O.: GUI for Comparing Perception of Sound Adjusted by Measured or Modeled HRTF, Technical Computing Bratislava 2010, 2010, pp. 1-5
9. Rund, F., Štorek, D., Glaser, O., Barda, m.: Orientation in Simple Virtual Auditory Space Created with Measured HRTF, Technical Computing Bratislava 2010, 2010, pp. 1-7

Coauthors of this paper are Ing. Dominik Štorek, O. Glaser, M. Barda, Czech Technical University in Prague, Faculty of Electrical Engineering, dept. of Radioelectronics, Technická 2, 166 27 Praha 6, Czech Republic