PERSONALIZED HEAD RELATED TRANSFER FUNCTION MEASUREMENT AND VERIFICATION THROUGH SOUND LOCALIZATION RESOLUTION

Michał Pec, Michał Bujacz, Paweł Strumiłło

Institute of Electronics, Technical University of Łódź
211/215 Wólczańska, 90-924 Łódź, Poland
michal_pec@o2.pl, bujaczm@p.lodz.pl, pawel.strumillo@p.lodz.pl

ABSTRACT

Filtering of sounds through head related transfer functions (HRTFs) is a common method for obtaining audio spatialization. HRTFs depend highly on an individual's anatomy, especially head dimensions and outer ear shape. The paper describes a system designed for efficient measurement of personalized HRTFs and verification of the collected data on a group of volunteers.

The main goal of utilizing personalized HRTFs was to obtain a high level of externalization, i.e. the illusion that a sound source is located outside one's head, as well as a high resolution of sound localization. Measurement of the HRTFs, using the constructed equipment in an anechoic chamber, was performed for 15 volunteers (9 of them blind, as our current research concerns electronic travel aids for visually impaired).

A series of trials were conducted, which verified the personalized HRTFs in externalization and localization quality. Average precision of localization of moving virtual sound sources reached 6.7° in azimuth and 10.6° in elevation. At the same time influence of other factors, such as the source type or movement, on sound localization was also tested.

1. INTRODUCTION

Living in the information age we constantly interact with numerous digital equipment which tests the limits of our perceptual abilities. Higher video resolutions, frame rates, 3D displays and virtual reality are all examples of this. Acoustic displays are also aiming to provide increasingly richer, more realistic experiences to listeners. This is especially evident in the evolution of different surround sound technologies utilizing multiple speakers. In this paper however we concentrate on a signal processing system, which allows to create realistic three dimensional sound using only stereophonic head phones.

The need for such a system arose during the process of designing an electronic travel aid for the blind (ETA) [1], which was to incorporate virtual 3D sound sources as part of its output. The only method to obtain spatialized audio on stereo speakers or headphones is filtering using head related transfer functions (HRTFs). These emulate the filtering introduced by human anatomical features, which allow us to perceive sounds three dimensionally using only our two ears.

Because HRTFs are a highly individualized characteristic, a system for their quick measurement for a number of users had to be designed and constructed.

After performing precise HRTF measurements of fifteen volunteers we were able to run a number of tests to verify the collected data and the future usefulness of the constructed equipment.

2. SPATIAL HEARING AND HRTFS

How is it that with only two ears we are able to precisely locate sound sources located anywhere in the three dimensional space? The first attempt to explain the mechanism of sound localization was the Duplex Theory, introduced by Lord Rayleigh early in the XX century [2]. According to this theory two main phenomena are responsible for sound localization:

- Interaural Time Difference (ITD) – the difference in the arrival time of a sound wave to each ear
- Interaural Level Difference (ILD) – the lowering of the sound's intensity in the ear further away from its source, more due to the effect of „head shadowing" rather than the marginally longer distance traveled.

The Duplex Theory, despite its correct assumptions, does not explain the mechanism of localizing sounds outside of the horizontal plane. So how to explain that a human can perceive a sound source's vertical position even with only one ear [3]? Other factors aside from ITD and IID must take part in 3D sound localization.

The answer lies in the sound's spectrum and the fact that the sound wave before reaching the eardrum undergoes multiple reflections and refractions from the shoulders, face and the outer ear. Different frequency components have different wavelengths, thus they interact differently with the anatomical features resulting in spectrum modifications that are unique for every direction in space and every listener. The relation modeling these modifications is called the Head Related Transfer Function (HRTF), and it is a function of frequency (ω) and two angles describing spherical coordinates: elevation (φ) and azimuth (θ).

$$HRTF = F(\omega, \varphi, \theta)$$

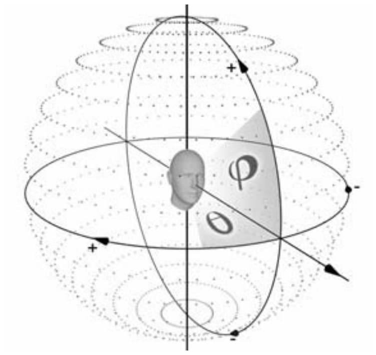

Figure 1 – Spherical coordinates used. Points on the sphere correspond to locations for which HRTFs were measured

Angles θ and φ determine the direction on the horizontal and vertical planes respectively as seen in Figure 1. When both are equal to zero, the source is located directly in front of the listener.

It can be clearly seen in Figure 2 showing sample HRTFs that the largest differences in the amplitude spectrum occur for the frequencies in the range of 2-10kHz, with most significant peaks in the range 3-6kHz. It can be speculated that sounds encompassing this area of the spectrum are most precisely localized. It has been proven that sounds which do not contain frequency components above 5kHz are very poorly localized [4].

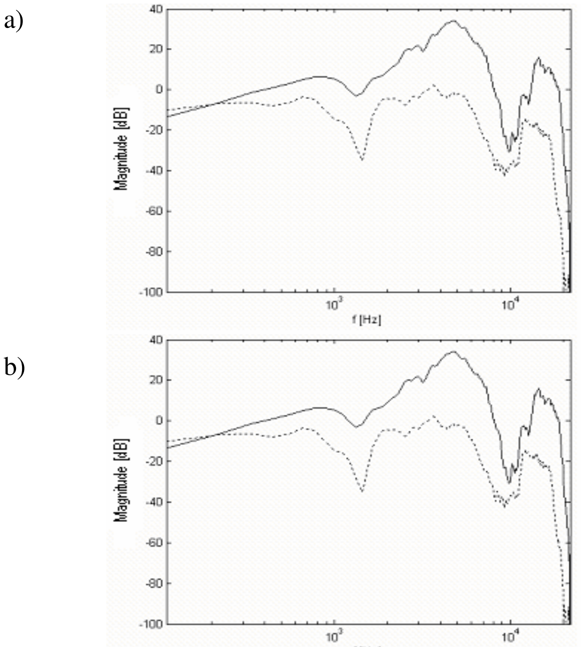

Figure 2 – Sample HRTFs for locations: a) θ=-40° i φ=-10° and b) θ=-40° i φ=18° ; solid line represents the left ear spectrum, the dashed line – the right ear spectrum.

2.1 Utilizing HRTFS

By incorporating HRTFs into the sound source it is possible to obtain the illusion of spatialization through standard stereo headphones. HRTFs are used in the construction of digital Finite Impulse Response (FIR) filters, which allow to recreate a sound wave near the entrance to the ear canal in the form, in the form it would have taken if it had been modified through reflections on the outer ear and shoulder.

A 2D array of paired filters (for left and right ears) corresponding to specific spherical directions is needed. A monophonic sound is filtered through the filter pair corresponding to desired azimuth and elevation coordinates. The filters emulate the audio distortions introduced by the head, shoulders and most importantly the ear pinnae, thus creating the illusion that the source has reached the listener from a particular direction in space.

Despite many general patterns, which can be used to form generalized transfer functions [3], the HRTFs are a highly individualized feature. The attenuation for a specific frequency band in a single direction's HRTF's spectrum can differ by as much as 20dB between two individuals. However, measurement of an individual's full HRTF set can be a long and tedious process. That is why there exists a number of very popular general HRTF sets recorded using mannequins, referred to as phantoms, with model ears and built-in microphones. Such solution is popular, because a phantom can be used in measurements lasting for hours, and allows for precise gathering of HRTF data.

We have found these general HRTFs insufficient for our purposes, mainly because a significant number of volunteers did not observe any externalization effects – the illusion of sounds originating outside of one's head, when using them. For this reason we decided to design a system for efficient measurement of personalized HRTFs.

3. HRTF MEASUREMENTS

Measuring of HRTFs is a complex process as the transfer function must be calculated for a large number of directions relative to the head. Our equipment used for HRTF measurements was designed and constructed in cooperation with the Technical University of Wrocław [5]. It allows for automatic measurements in the full azimuth range (θ=0° to 360°) with a step of 1° and a broad elevation range (φ=-45° to 90°) with a step of 9°. Although higher resolution is possible, it requires remounting of speakers and combining data from two or more measurements.

The HRTF measuring equipment consists of a rotating chair with an adjustable head rest and microphone mounts, a set of 16 speakers mounted on an arch with a 1m radius, a digital camera used for calibration and an intercom for communication with the equipment operator sitting outside the anechoic chamber.

The measured HRTFs are given in the form of an impulse response (so the term HRIR – Head Related Impulse Response can also be used) with a number of coefficients ranging from 256 to 4096.

The measurement consists of recording the sounds produced by the speakers in various chair positions through two microphones placed at the entrances to a listener's ear canals. This allows to determine the influence of the individual's anatomy (the shape of the shoulders, head and pinnae) on the sound wave.

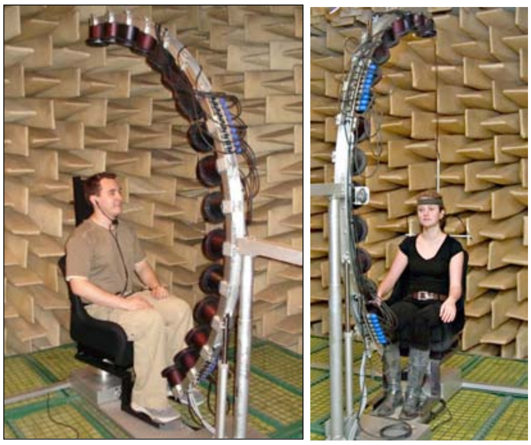

Figure 3 – HRTF measurement setup in an anechoic chamber.

Full HRTF measurements were performed for 15 persons, with a 5° step in azimuth and 9° step in elevation. With such resolution one measurement run lasted 10 minutes, and the whole measurement procedure for one person, including two runs and equipment fitting, took approximately half an hour.

3.1 Data processing and interpolation

In order to increase the HRTF resolution their interpolation was performed. A decision was made to interpolate using the HRIR coefficients, as this was the format of the data recorded by the equipment. However, the impulse responses also include the time delay caused by the differing propagation times to each ear. For interpolation in the time domain all impulse responses need to be equally delayed. To meet this condition, minimum phase components were extracted through windowing the cepstrum [6] of the impulse responses.

Before interpolating the data was sorted into a 3D array, indexed with azimuth (72 elements), elevation (16) and sample number of the impulse response (256). For each sample number bicubic spline interpolation was performed in the 2D array of azimuth and elevation.

The minimum phase conversion caused the loss of a very important localization factor: the ITD. The Woodworth formula for a spherical head model [7] was used for its reconstruction:

$$ITD = \frac{d \cdot (\theta + sin\,\theta) \cdot cos\,\varphi}{2v}$$

where: d – head diameter, v – sound velocity, θ – azimuth, φ – elevation. According to [7] this slight approximation introduces very little error into sound localization.

The obtained set of HRIRs with resolution of 1° both in azimuth and elevation was converted into a format supported by the SLAB environment [8] used in further verification and trials.

4. VERIFICATION TRIALS

Spatial hearing is a very subjective phenomenon, but a number of trial procedures were developed for verification of the usefulness of the collected data.

4.1 Externalization trials

A primary concern when dealing with spatial sound is the externalization phenomenon – the illusion of sound sources located outside of one's head when hearing them through head hones. Test of this phenomenon were performed with high-end HD-650 Sennheiser headphones. Participants were asked to point to a virtual source, which was orbiting around their heads. The source was either white noise or a chirp sound, depending on volunteer preference.

First trials with the HRTFs from the CIPIC database containing measurements from 45 different persons [9] ended with unsatisfactory results. Out of 8 volunteers who tested various HRTFs from the database only half observed any sound externalization, the remaining four heard the spatial sounds moving inside or on the surface of their heads.

Further attempts were made later on personalized HRTFs for 15 volunteers (9 of them blind, but the influence of their disability was not the subject of this study). Most of the trial participants were able to clearly perceive virtual sound sources outside their heads and accurately track the orbiting source. A few participants had problems with precise localization of the source, but did perceive it to be externalized. Two participants did not externalize the sounds right away, but were able to do so after a short training session in which they were able to observe on the computer where the virtual source was located. Only one of the 15 volunteers was unable to observe externalized audio. The collected results are presented in Table 1.

Table 1 – Results of externalization trials

Level of externalization	Persons
Full externalizaiton and localization of sources	8 (4 blind)
Externalization, without precise localization	4 (2 blind)
Externalizaiton after training with visual feedback	2 (2 blind)
Lack of externalization	1 (1 blind)

4.2 Localization of static virtual sources

For further studies we selected five volunteers which had no problems with sound externalization. Each participant sat in front of a large paper screen with a 1cm grid. Different virtual sources located in the frontal hemisphere were presented to the volunteer using his previously measured HRTFs. The volunteer was to point to the screen where he perceived the sound to be coming from. The coordinates were then converted into polar form and the accuracy of the perceived azimuth and elevation was calculated. We concentrated on sources within the front 100°x100° central area, as this will be the region in which our ETA will place virtual sound sources informing a blind user of obstacles [1].

The sound used for these trials was the vowel „a" synthesized with different base frequencies (from 60 to 200Hz) and different amounts of modulated noise. The synthesizer used is detailed in [10]. Reference runs with real sound sources and HRTFs not belonging to the user (if ones allowing externalization were found) were also made. The results are presented in Table 2 and Figure 4.

Table 2 – Results of static localization trials

Source type	Average azimuth error	Average elevation error	Azimuth correlation	Elevation correlation
"a" 200Hz +20% noise	6.2°	14.7°	0.91	0.69
"a" 200Hz +40% noise	7.2°	12.7°	0.92	0.82
White noise	6.8°	11.5°	0.95	0.80
Real source	3.0°	4.8°	0.995	0.993

4.3 Localization of dynamic virtual sources

Although average errors in the static trials were relatively small and a number of patterns could be observed, the problem remained with quite frequent large errors and inconsistencies in measurements. Since humans perceive moving or changing sounds the best [11] we decided to take advantage of that fact in our trials.

In addition to the synthesized vowel we also used chirp sounds with a wider spectrum, one ranging from 500Hz to 16kHz and another in the 500Hz to 8kHz range. The sound source oscillated with a velocity of 10°/s between two points in one of three trajectories:

- horizontally in a 20° range
- vertically in a 20° range
- diagonally - moving in both directions at once

Trial participants were asked to point to the perceived extremes of the oscillating virtual sources' paths (thus two measurements were recorded each time). Again, reference runs were made with a real sound and other non-personalized HRTFs. The trial results are summarized in Table 3 and sample measurement data shown in Figure 5.

Table 3 – Results of dynamic localization trials

Source type	Average azimuth error	Average elevation error	Azimuth correlation	Elevation correlation
"a" 200Hz +40% noise	5.7°	12.8°	0.95	0.67
Chirp 0,5-16kHz	6.7°	10.6°	0.95	0.85
Real source	2.5°	3.4°	0.998	0.995

5. DISCUSSION OF RESULTS

Despite the relatively small group of tested volunteers, a number of clear trends and patterns could be observed.

First of all, the experiments proved that the measurement of personalized HRTFs could be both efficient (<30min) and provide useful data of high quality, resulting in a much greater externalization rate and more precise localization.

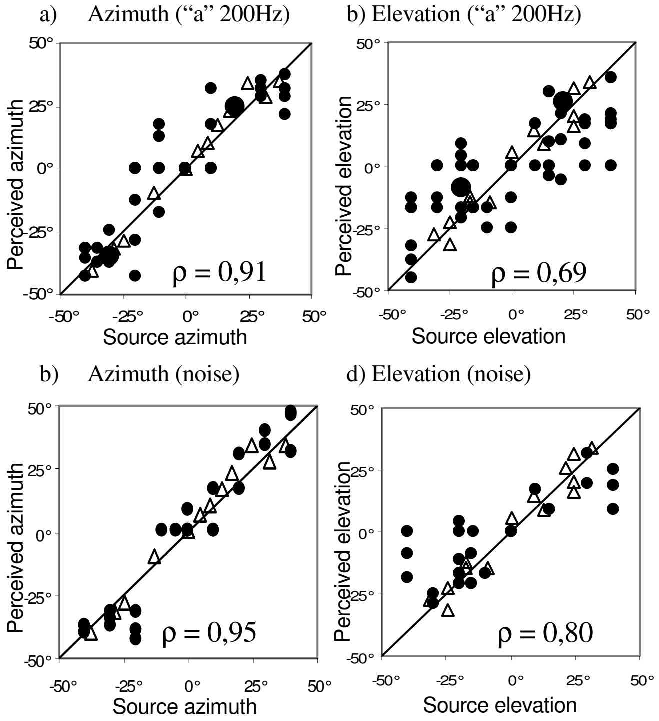

Figure 4. Localization trial data for static virtual sources for one of the volunteers. Triangles denote a reference trial with real sounds.

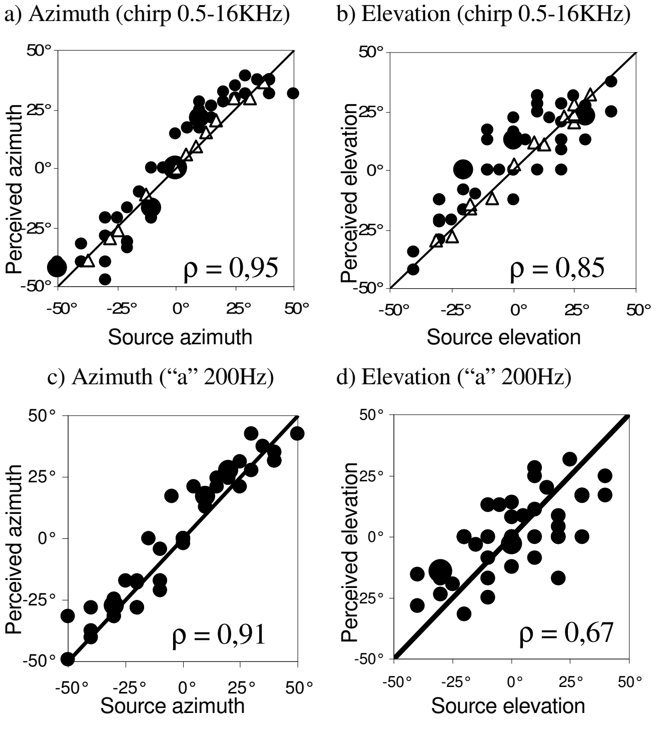

Figure 5. Localization trial data for moving virtual sources for one of the volunteers. Triangles denote a reference trial with real sounds.

Second of all, the trials showed us which types of sounds were best localized and the limits of localization resolutions we can expect to achieve in the future. In accordance with theory, sources with a wide and flat spectrum were localized more precisely (white noise or chirp), but there were exceptions to the rule. Sounds modulated with a large amount of white noise tended to spatially "smear", and were harder to localize in azimuth. As theorized, moving sources were more precisely localized than static ones. We expect to achieve even better resolutions with the incorporation of active-feedback from a head tracker, allowing for head movements inside the virtual acoustic environment.

Thirdly, the phenomenon of spatial separation of wide spectrum sounds was observed and might be the subject of further studies. A number of volunteers perceived two separate sources when only one was given. The source seemed to be split into its high-frequency components that were precisely localized, and the low frequency part which seemed to fade into a uniform background. This occurred for the vowel "a" modulated with noise at lower frequencies and with chirp sound when the sweeping frequency was high.

Lastly, the trials pointed out to us problems that could be worked on in the future. For example, all the azimuth errors seemed to show a tendency for sources to be perceived more to the sides than they actually were. We suspect that this might have been caused by the introduction of a too large ITD through the use of the Woodworth formula. Decreasing the head diameters used in the conversions to slightly below the actual measured values improved this tendency, but did not fixed the problem entirely. The results in Figures 4 and 5 are from measurements with the corrected head diameter, but the tendency for the perceived absolute angle values to be larger more often than smaller still remains.

6. SUMMARY AND CONCLUSIONS

During the course of our work on an electronic travel aid for the blind we encountered the need to be able to present 3D environments by means of spatial acoustic displays. The only way to provide spatialized audio through stereo headphones is the use of head related transfer functions (HRTFs). Initial trials with general or non-personal HRTFs showed a low rate of sound externalization among a number of volunteers, creating the need for personalized HRTF measurements if spatial audio was to be used in our system.

Equipment for efficient measurement of quality personalized HRTFs was designed and constructed. The collected data was interpolated to cover a higher resolution of azimuth and elevation angles.

Full personal HRTF measurements were done on 15 volunteers, who later participated in various trials. The first trial dealt with externalization, and personalized HRTFs proved to have provided clearer effect than non-personal ones. Further trials showed that with the collected HRTFs virtual spatial sound sources could be localized with an average error of 6.8° and 11.5° for azimuth and elevation respectively, and an error of 6.7° and 10.6° for moving sources. These results are noticeably better than those in some similar studies utilizing generalized HRTFs or the CIPIC database, which ranged between 13° and 20° for azimuth [12,13], especially that [13] points out lower localization accuracy by blind individuals.

The experiments have confirmed a number of theories about parameters influencing the presentation of spatial sounds and proved future usefulness of the employed HRTF measurement system.

ACKNOWLEDGEMENTS

This work has been supported in part by the Ministry of Science and Higher Education of Poland research grant no. 3T11B 03827 in years 2004-2007 and in part by the Mechanism for Support of Innovative Ph.D. Student Research financed by the European Social Fund and the Polish Ministry of Economy.

REFERENCES

[1] Paweł Strumiłło, Paweł Pełczynski, Michał Bujacz, Michał Pec, "Space Perception By Means Of Acoustic Images: An Electronic Travel Aid For The Blind", 33rd International Acoustical Conference, Slovakia, 2006.

[2] F. A. Everest, *The master handbook of acoustics.* McGraw-Hill, USA, 2001, pp. 67-68.

[3] R. O. Duda, "Auditory localization demonstations", *Acustica acta acustica*, Vol. 82, , 1996, pp. 346-355.

[4] F. L. Wightman, D. J. Kistler, "Factors Affecting the Relative Salience of Sound Localization Cues", edited by R. H. Gilkey, T. R. Anderson, *Binaural and Spatial Hearing in Real and Virtual Environments*, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, 1997, pp. 1-24.

[5] P. Plaskota, P. Pruchnicki, "HRTF automatic measuring system" 53 Open Acoustics Seminar, Zakopane, 2006.

[6] A.V. Oppenheim and R. W. Schafer, *Discrete-time signal processing.* Prentice Hall, 1975, pp. 501-511.

[7] R.O. Duda, W.L. Martens, "Range-dependence of the HRTF for a spherical head" *J. Acoust. Soc. Am.*, Vol. 104, No. 5, 1998, pp. 3048-3058.

[8] J.D. Miller, http://human-factors.arc.nasa.gov/SLAB/

[9] *CIPIC Interface Laboratory*: http://cipic.ucdavis.edu

[10] Paweł Strumiłło, Paweł Pełczynski, Michał Bujacz, "Formant-based speech synthesis in auditory presentation of 3d scene elements to the blind", 33rd International Acoustical Conference, Slovakia, 2006.

[11] M. Kato, H. Uematsu, M. Kashino and T. Hirahara, "The effect of head motion on the accuracy of sound localization", *Acoustical Science and Technology*, Vol. 24, No.5, 2003, pp. 315-317.

[12] J. Scarpaci, H. Colburn, J. White, "A system for real time auditory space" 11th International Conference on Auditory Display, Ireland, 2005.

[13] A. Afonzo, B. Katz, "A Study of Spatial Cognition in an Immersive Virtual Audio Environment: Comparing Blind and Blindfolded Individuals" 11th International Conference on Auditory Display, Ireland, 2005.