University *of Ljubljana*
Faculty *of Mathematics and Physics*

SEMINAR I$_a$ - 1. YEAR, 2ND CYCLE

Head-Related Transfer Function

Author:
Tilen POTISK

Advisor:
doc. dr. Daniel SVENŠEK

Ljubljana, 9$^{\text{th}}$ January, 2015

Abstract

Head-related transfer function (HRTF) is a function used in acoustics that characterizes how a particular ear (left or right) receives sound from a point in space. A pair of two transfer functions, one for each ear, is used for sound localization which is very important for humans. In this seminar ways of measuring and application of HRTFs are presented. At the end I will also present problems of eliminating cross-talk caused by two loudspeakers.

Contents

1 Introduction

Humans have an amazing ability to spatialize and locate the sound. Having such an ability has a clear evolutionary advantage. The auditory system uses several different cues for locating sound source, such as time and level differences between both ears or spectral information. Some animals may also use different cues such as the effect of ear movements. Knowing how our body features filter sound is important in reproducing binaural sound.

First, I will define head-related transfer function as a function in the frequency domain. I will briefly explain basics of spatial hearing and present several different cues of the auditory system for locating sound source described in [1]. In the second part I will present a technique used to measure the head related transfer function. Following the Ref. [2], I will present the so called cross talk cancellation, which occurs when reproducing binaural recording.

2 Head-Related Transfer Function

Head-related transfer functions (HRTF) capture transformations of a sound wave propagating from the source to our ears. Some of the transformations include diffraction and reflections on the parts of our bodies such as our head, pinnae, shoulders and torso. As a consequence, with these two functions we are able to create the illusion of spatially located sound [3].

HRTF is a Fourier transform of a head-related impulse response (HRIR). It should be emphasized that it is a complex function defined for each ear, having both information about the magnitude and the phase shift. The HRTF is also highly dependent on the location of the sound source relative to the listener, which is a main reason we are able to locate the sound source.

We will denote the impulse responses for the left and the right ear in the time domain as $h_L(t)$ and $h_R(t)$, respectively. In the frequency domain we will denote the responses by corresponding capital letters - $H_R(\omega)$ and $H_L(\omega)$.

Let the function $x(t)$ describe the pressure of the sound source and let functions $x_L(t)$ and $x_R(t)$ be the pressure at the left and the right ear, respectively. In the time domain, the pressure at the ears can be written as a convolution of the sound signal and the HRIR of the corresponding ear:

$$x_{L,R}(t) = h_{L,R}(t) * x(t) = \int_{-\infty}^{\infty} h_{L,R}(t - \tau)x(\tau)\mathrm{d}\tau. \tag{2.1}$$

In the frequency domain, convolution is transformed into multiplication:

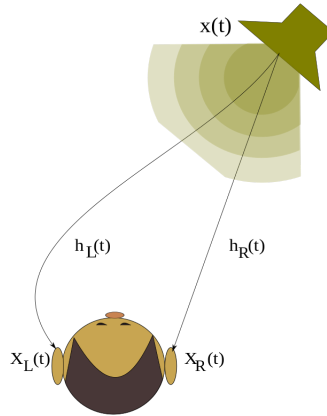$$X_{L,R}(\omega) = \mathcal{F}\left(h_{L,R}(t) * x(t)\right) = H_{L,R}(\omega)X(\omega). \tag{2.2}$$

Figure 1: Filtering of a signal $x(t)$ by two separate transfer functions $h_L(t)$ and $h_R(t)$ [4].

Figure 1 shows schematicly the propagation of a sound wave from the sound source to the ears with the conventional notations for filtering functions and pressure impulses.

2.1 Spatial hearing

Spatial hearing is an important aspect of auditory perception. It allows us to orient ourselves in space and also suppress sound reflections from walls of the room that interfere with the direct sound.

The position of the sound source can be specified with three variables: the azimuthal angle ϕ in equatorial plane also called the left/right dimension, the polar angle θ also called the vertical dimension and distance r from the listeners head to the sound source.
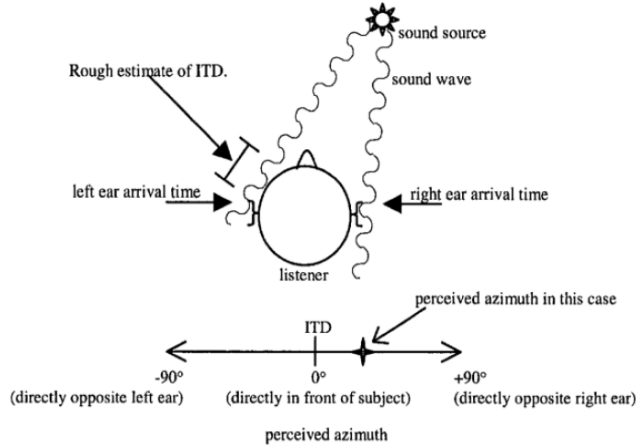
Figure 2: Basic concept of interaural time differences (ITD)[1].

Various psychoacoustic experiments have shown that the following cues contribute to the spatial localization of the sound source [5]:

- The interaural time difference (ITD) is the time difference of the arrival of sound waves at left and right ear canals. This cue is dominant for frequencies below 1.5 kHz. Figure 2 shows the basic concept of the ITD caused by the sound waves coming from a specific azimuthal angle.

- The interaural level difference (ILD) is the pressure level difference between left and right ears. This cue is important for frequencies above 1.5 kHz.

- The spectral cues encoded in the shapes of our pinnae. This cue is important for frequencies above 5 to 6 kHz. The importance of these cues will be emphasized in the next subsection.

The frequency below which the ITD cue is dominant (ν_c) can be estimated from the average diameter of the head (d) and the speed of sound (c):

$$\nu_c = \frac{c}{d} \approx 1.5 \text{ kHz}, \tag{2.3}$$

where we used $c = 340 \frac{\text{m}}{\text{s}}$ and the average diameter of the human head $d = 25$ cm.

2.1.1 Cone of confusion

Since these two cues (ITD and ILD) only explain the perception of azimuth, there are a number of locations along curves of equal distance [1]. This set of points is called the *cone of confusion* because the location of the sound waves originating from these points are indistinguishable as it is shown on figure 3. It turns out that brains use another cue which is used to locate sounds in the median plane. Our pinnae are shaped to collect sound and also change spectral profile of a sound. Depending on the origin of the sound source, certain frequency intervals get enhanced while others get attenuated.
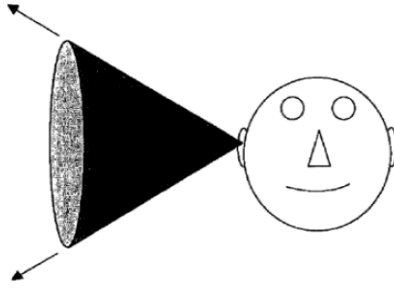
Figure 3: The cone of confusion [1].

2.2 Measurement of HRTFs

Measurement of HRTFs is expensive, since a typical set up requires high quality audio equipment like speakers and headphones. It is also time consuming and is usually performed in anechoic room to reduce the effect of a room geometry. To measure the transfer functions we produce an input signal from a given space location and measure the output at the entrance of the left and right ear canals. The desired function can then be obtained from equation (2.2):

$$H_{L,R}(\omega) = \frac{X_{L,R}(\omega)}{X(\omega)}. \tag{2.4}$$

Because HRTFs vary depending on the angle of incidence of the sound waves, it is necessary to excite signals from all directions where the HRTF ought to be measured. For this purpose a set of loudspeakers are mounted on a semicircular rotating hoop which excite the signals used in measurements (figure 4). The subject's head is at the center of the semicircular hoop and the interaural axis is aligned with the axis of rotation of the hoop. To ensure that the subject doesn't move the head too much, a set of lasers pointing into ear canals notify the subject to align the head with the axis of the hoop.
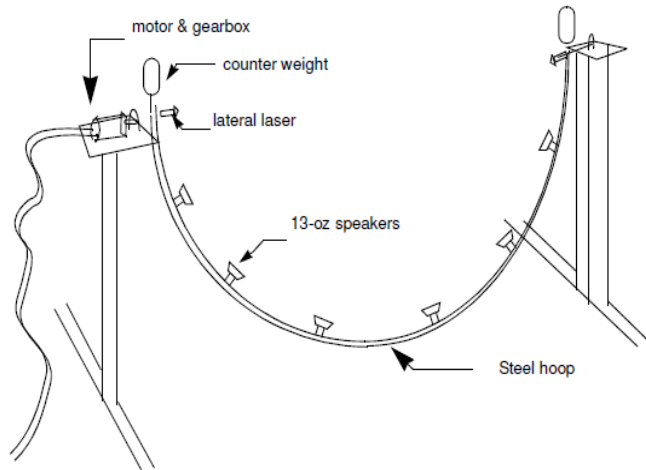

Figure 4: Measurement of HRTFs with a semicircular rotating hoop of loudspeakers [6].

For recordings a replica of the human head (dummy head) is often used (figure 5). It is shaped of an average human head. The sound waves reaching the dummy head undergo

4

approximately the same transformations on their way to the ear channels, as if they were reaching the listeners head [7].

Figure 5: An example of the dummy head used in measuring HRTFs [8].

There are many different signals that can be used. Most popular are the impulse responses, exponential sweep signals and pseudo-random noise signals.

If the distance from the sound source to the center of the head is greater than 1.0 m, the HRTF is approximately independent of distance and so the measurements are called far field measurements, where the loudspeaker can be approximated as a point source. When the distance is less than 1.0 m the HRTFs vary with distance and are thus called near-field HRTFs [5]. Near-field measurement is more difficult because you need a near-field point source and so an ordinary loudspeaker is no longer a suitable choice, due to its size, directivity and multiple scattering between source and the subject. Secondly, measurement is much more time consuming since measurements of HRTFs at various distances are required due to the strong distance dependency of the near field. This means that it is almost impossible to obtain a near-field transfer function of a human subject.

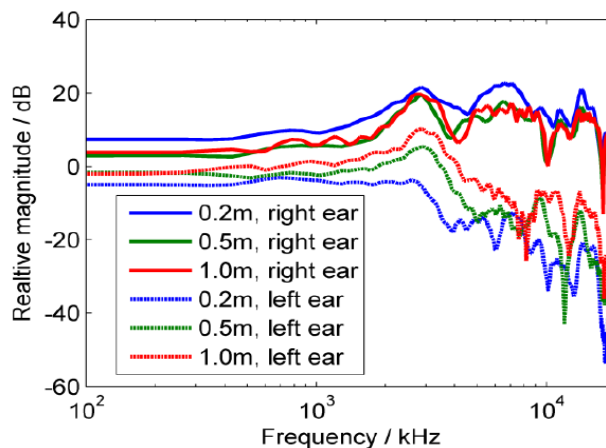Figure 6: HRTF magnitudes obtained with at the left and tight ears of the dummy head at $r = 0.2$ m, 0.5 m, 1.0 m and $(\theta, \phi) = (90°, 0°)$ [5].

The figure 6 show HRTF magnitudes at $r = 0.2$ m, 0.5 m, 1.0 m and $(\theta, \phi) = (90°, 0°)$.

We can see clearly that the magnitudes vary with distance from $r = 0.2$ m to $r = 0.5$ m, and vary less with distance from $r = 0.5$ m and 1.0 m.

2.3 Headphone reproduction

In headphone reproduction we can accurately render the virtual sound source if the sound pressures for original sound source are exactly replicated at our eardrums. However, there are numerous experimental results that indicate the following subject-dependent errors in perceiving virtual source position:

1. Reversal error (front-back confusion). Virtual source intended in the front hemisphere is perceived at mirror position. Sometimes there is also confusion with up and down source positions.

2. Elevation error. The perceived direction of the sound source in the median plane is usually elevated.

3. In-head localization. The virtual sound source is perceived inside the head which is not a natural hearing experience.

It is known that the front-back and vertical localization depend more on high frequency cues [5]. Unfortunately these are also highly individual-dependent. The possible source of errors can be errors in binaural recording/synthesis, such as lack of headphone equalization and non-individualized HRTF processing. Using individual HRTFs reduces localization errors.

2.4 Room reflections

Reflections exist in most real rooms and are essential to spatial auditory perception. Incorporating the effects of reflections into the so called virtual auditory environment brings many advantages: ability to recreate the spatial auditory perception in a room, reducing the in-head localization (mentioned in the previous subsection) in headphone presentation and perceiving virtual source distance [5].

A basic method to include reflections is simulating the physical propagation of sound from the source to the listener inside a room (binaural room impulse response - BRIR) and then convolving the input stimulus with BRIR.

3 Cross-talk cancellation

Let us assume our stereo signal was encoded with the head-related transfer function and includes all relevant cues. We can measure this using a dummy head with microphones placed at the entrance of the ear canals. This is also called binaural recording.

If we delivered the signal on each of the channels of the stereo signal (only and directly) to the corresponding ear, the listener would hear an accurate $3-$D reproduction of the recorded sound signal. A partial cancellation of sound signals at both ears is needed, since the sound from one channel is also heard on contralateral ear (crosstalk), which is shown on the figure 7. Prior to loudspeaker reproduction, the original signals should be filtered so as to cancel the transmission from each loudspeaker to the contralateral ear [2].

The principle of crosstalk cancellation can be described in matrix notation. The height of the filter matrix H is the number of ears and the width is the number of loudspeakers. In the following analysis we will use two ears and two loudspeakers with the amplitudes Y_L

and Y_R for the left and the right channel, respectively. There is a relation between outputs at both channels and the desired amplitudes at our ears, X_L and X_R:

$$H \begin{pmatrix} Y_L \\ Y_R \end{pmatrix} = \begin{pmatrix} X_L \\ X_R \end{pmatrix}. \tag{3.1}$$

The matrix H is the matrix of all possible HRTFs. Elements of the matrix are geometry dependent i.e. if we change the locations of the loudspeakers and our position in the room the matrix H changes:

$$H = \begin{pmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{pmatrix} \tag{3.2}$$
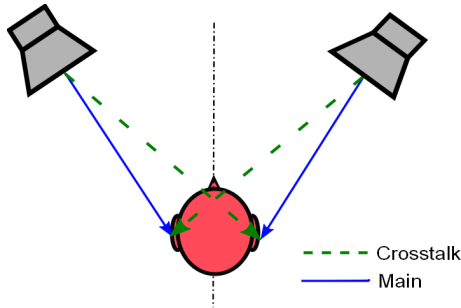

- - - - Crosstalk
——— Main

Figure 7: Cross-talk at left and right ear caused by interference from right and left loud-speaker, respectively [10].

It should be emphasized that the amplitudes Y_L and Y_R are the transformed amplitudes and we are to find this transformation, denoted C:

$$\begin{pmatrix} Y_L \\ Y_R \end{pmatrix} = C \begin{pmatrix} E_L \\ E_R \end{pmatrix}, \tag{3.3}$$

where E_L and E_R stand for amplitudes of the original sound signal measured directly at the entrance of the left and right ear canals, respectively. Vector of the two pressure amplitudes $\mathbf{x} = [X_L, X_R]^T$ can be shortly written as a transformation of vector of amplitudes $\mathbf{e} = [E_L, E_R]^T$:

$$\mathbf{x} = HC\mathbf{e}. \tag{3.4}$$

We require that $\mathbf{x} = \mathbf{e}$. This is possible only if

$$HC = I, \tag{3.5}$$

where I is an identity matrix.

3.1 Symmetric loudspeaker set-up

We wish to analyze a simple system consisting of two loudspeakers and two ears, where we make an assumption that the signals from the loudspeakers propagate to our ears as in the free field. The impulse response at our ears will be simply just a time-delayed and attenuated delta function. We will analyze a symmetric loudspeaker set-up as shown on figure 8. The two loudspeakers will be point sources radiating a sound wave of frequency ω with pressure amplitudes:

$$P(r, \omega) = \frac{i \omega \rho_0 q}{4\pi} \frac{e^{-ikr}}{r}, \tag{3.6}$$

where ρ_0 is the air density, $k = 2\pi/\lambda$, λ is wavelength and q the source strength. Equation (3.6) can be written as

$$P(r,\omega) = Y(\omega)\frac{e^{-ikr}}{r}, \tag{3.7}$$

where we defined a quantity Y, which is proportional to the source strength

$$Y(\omega) = \frac{i\omega\rho_0 q}{4\pi}. \tag{3.8}$$

At the left ear the air pressure due to the two sources add up as

$$X_L(\omega) = \frac{e^{-ikl_1}}{l_1}Y_L(\omega) + \frac{e^{-ikl_2}}{l_2}Y_R(\omega) \tag{3.9}$$

and similarly at the right ear we have

$$X_R(\omega) = \frac{e^{-ikl_2}}{l_2}Y_L(\omega) + \frac{e^{-ikl_1}}{l_1}Y_R(\omega). \tag{3.10}$$

Because of the symmetric set-up of loudspeakers, there are only two different distances l_1
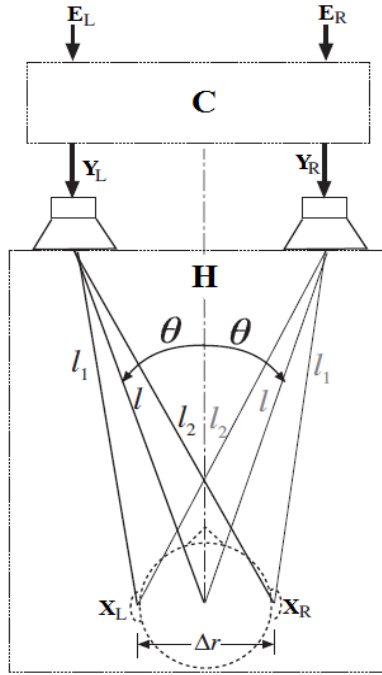

Figure 8: An example of symmetric loudspeaker set-up [2].

and l_2 in the equations. It is convenient to define

$$\Delta l = l_2 - l_1 \text{ and } g = \frac{l_1}{l_2}. \tag{3.11}$$

From the geometry on the picture 8 it follows that

$$l_1 = \sqrt{l^2 + \left(\frac{\Delta r}{2}\right)^2 - \Delta r l \sin(\theta)} \text{ and } l_2 = \sqrt{l^2 + \left(\frac{\Delta r}{2}\right)^2 + \Delta r l \sin(\theta)}. \tag{3.12}$$

Another important parameter is the characteristic time τ_c

$$\tau_c = \frac{\Delta l}{c}. \tag{3.13}$$

The equations for the pressure amplitudes can be written in matrix form as

$$\begin{pmatrix} X_L(\omega) \\ X_R(\omega) \end{pmatrix} = \alpha \begin{pmatrix} 1 & g e^{-i\omega\tau_c} \\ g e^{-i\omega\tau_c} & 1 \end{pmatrix} \begin{pmatrix} Y_L(\omega) \\ Y_R(\omega) \end{pmatrix}, \tag{3.14}$$

where we have defined

$$\alpha = \frac{e^{-i\omega l_1/c}}{l_1}, \tag{3.15}$$

which just introduces a phase shift (or time delay in time domain) divided by the constant l_1. The HRTF of our loudspeaker set-up for a fixed position of the listener can be extracted from equation (3.14):

$$H = \alpha \begin{pmatrix} 1 & g e^{-i\omega\tau_c} \\ g e^{-i\omega\tau_c} & 1 \end{pmatrix}. \tag{3.16}$$

3.1.1 Perfect crosstalk cancellation

A perfect crosstalk cancellation filter theoretically yields complete cancellation of crosstalk at the entrance of ear canals of the listener, for all frequencies. We see that in order to achieve this, equation (3.5) requires that $HC = I$. The filter for the perfect crosstalk cancellation can be in our simplified example analytically expressed:

$$C = H^{-1} = \frac{\alpha^{-1}}{1 - g^2 e^{-2i\omega\tau_c}} \begin{pmatrix} 1 & -g e^{-i\omega\tau_c} \\ -g e^{-i\omega\tau_c} & 1 \end{pmatrix} \tag{3.17}$$

An important measure of the performance of crosstalk filters is the ratio of the diagonal and off-diagonal terms in the matrix $R = HC$:

$$\chi_{RL}(\omega) = \frac{R_{RR}(\omega)}{R_{LR}(\omega)} \quad \text{or} \quad \chi_{LR}(\omega) = \frac{R_{LL}(\omega)}{R_{RL}(\omega)}. \tag{3.18}$$

In the case of a perfect crosstalk cancellation, we have by definition $R = I$ and therefore

$$\chi_{RL}(\omega) = \chi_{LR}(\omega) = \infty. \tag{3.19}$$

3.1.2 Sensitivity to perturbations

The eigenvalues of matrices are in general sensitive to perturbations [11]. Small changes in the matrix elements can severely change the eigenvalues. Similarly, in matrix inversion problems (e.g. $\mathbf{x} = H^{-1}\mathbf{b}$), the sensitivity of the solution ($\mathbf{x}$) to a small change in the input argument ($\mathbf{b}$) is characterized by the so called condition number κ, given by

$$\kappa(H) = \|H\|\|H^{-1}\|, \tag{3.20}$$

where $\|.\|$ is the spectral norm, which is equal to the largest singular value of the matrix [11]. The condition number is given by the ratio of the largest to smallest singular values (σ_1 and σ_2) of the matrix H. In our example, we have two singular values:

$$\sigma_1 = |\alpha|\sqrt{1 + g^2 - 2g\cos(\omega\tau_c)}$$
$$\sigma_2 = |\alpha|\sqrt{1 + g^2 + 2g\cos(\omega\tau_c)}. \tag{3.21}$$

We can see that the values of σ_1 and σ_2 of our HRTF, defined in equation (3.16), are a function of $\omega\tau_c$ so the appropriate ratio of singular values should be taken when calculating the condition number, either σ_1/σ_2 or σ_2/σ_1, whichever is the highest:

$$\kappa(H) = \max\left(\frac{\sigma_1}{\sigma_2}, \frac{\sigma_2}{\sigma_1}\right). \tag{3.22}$$

The following minima and maxima can be found:

$$\kappa(H) = \frac{1+g}{1-g} \text{ at } \omega\tau_c = n\pi, \ n = 0, 1, 2, 3, \ldots$$
$$\kappa(H) = 1 \text{ at } \omega\tau_c = \frac{n\pi}{2}, \ n = 1, 3, 5, \ldots, \tag{3.23}$$

which can also be seen on figure 9. Typical values of g are close to 1 so the maxima of $\kappa(H)$ are very high. In the limit $g \to 1$ the inversion problem becomes ill-conditioned at $\omega\tau_c = n\pi$, which occurs when the path lengths from loudspeakers to the corresponding ear differ by half the wavelength at frequency ω. The filter matrix then becomes non-invertible and infinitely sensitive to errors:

$$\lim_{g\to 1} \kappa(H) = \infty. \tag{3.24}$$

If the listener were to move his head slightly, this would change the elements of the matrix H. This would result in a loss in crosstalk cancellation control at certain frequencies (which form a harmonic series). When a crosstalk filter boosts some frequencies, we say that they bring spectral coloration i.e. affect the sound colour of the recording.
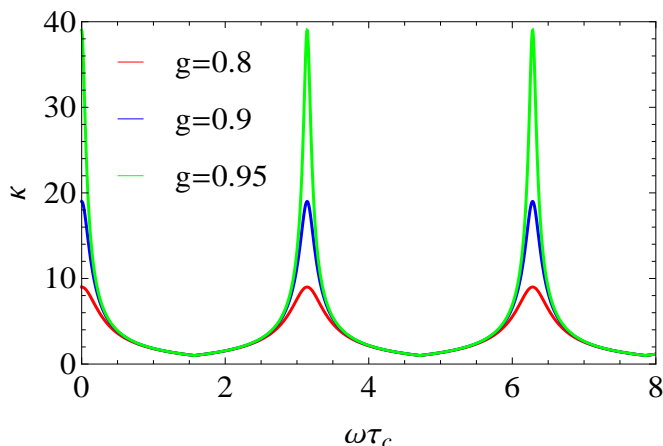

Figure 9: Condition number κ as a function of $\omega\tau_c$ at three different values of g.

There is a way to control the norm of the approximate solution but at the price of loss in the accuracy of solution. We want to minimize the spectral coloration for a desired minimum crosstalk cancellation performance. This will not be further discussed but let us just mention that in essence, a nearby solution to the matrix inversion problem is sought [2]:

$$C = [H^\dagger H + \beta I]^{-1} H^\dagger, \tag{3.25}$$

where the superscript $\dagger$ denotes the adjoint matrix and $\beta > 0$ is the so called regularization parameter which causes that the matrix H is not the exact inverse of C.

4 Conclusion

In this work, I have presented the basic concepts behind acoustical response of the listeners head described by the head-related transfer function. Some principles of spatial localization and techniques for measuring HRTFs were presented. It must be emphasized that the HRTFs are not unique. Different people have different shapes of head and torso and therefore the transformations of sound waves from the sound source to the ear canals of the listeners are different. In many cases the HRTF of the dummy head is measured instead and we usually take left-right symmetric functions, since the left-right asymmetries are not ideal for all subjects.

Measurement of HRTF is important, since it can be used in cancellation of a crosstalk. I have shown some problems relating the perfect crosstalk cancellation such as ill-conditioned matrix inversion. A way to minimize the spectral coloration, but at the price of loss in the accuracy of the solution, was briefly mentioned.

References

[1] S. Park: *Three Main Spatial Techniques of Spatial Mixing System*, (see hdl.handle.net/2123/10644) (7. 11. 2014)

[2] E. Y. Choueiri: *Optimal Crosstalk Cancellation for Binaural Audio with Two Loudspeakers*, Journal of the Audio Engineering Society, 2008

[3] B. Groethe, M. Pecka, D. McAlpine: *Mechanisms of Sound Localization in Mammals*, Physiol Rev 90: 983–1012, 2010 doi:10.1152/physrev.00026.2009

[4] en.wikipedia.org/wiki/Head-related_transfer_function (7. 11. 2014)

[5] Xiao-li Zhong and Bo-sun Xie (2014). *Head-Related Transfer Functions and Virtual Auditory Display*, Soundscape Semiotics - Localization and Categorization, Dr. Hervé Glotin (Ed.), ISBN: 978-953-51-1226-6, InTech, DOI: 10.5772/56907.

[6] E. Grassi, J. Tulsi, S. Shamma: *Measurement of Head-Related Transfer Functions Based on the Empirical Transfer Function Estimate*, Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA, July 6-9, 2003

[7] A. Illenyi, G. Werseny: *Evaluation of HRTF Data using the Head-Related Transfer Function Differences*, Forum Acusticum, Budapest, 2475-2479 (2005)

[8] broadcastaudiotechnology.wikia.com (9. 11. 2014)

[9] T. Weissgerber, K. Laumann, G. Theile, H. Fastl: *Headphone Reproduction via Loudspeakers using Inverse HRTF-Filters*, NAG/DAGA International Conference on Acoustics 2009

[10] code.google.com/p/virtudroid/wiki/Principle (9. 11. 2014)

[11] James E. Gentle: *Numerical Linear Algebra for Applications in Statistics*, Springer (1998)