
# HRTF modelling in due consideration variable torso reflections

M. Guldenschuh, A. Sontacchi and F. Zotter

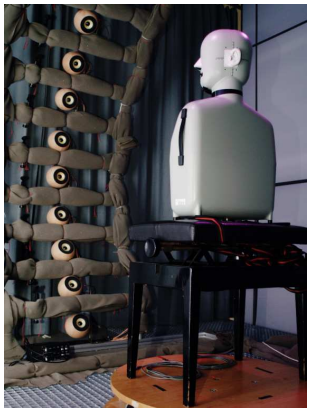
Institute of Electronic Music and Acoustics, Inffeldgasse 10 / 3, 8010 Graz, Austria
sontacchi@iem.at

Head related impulse response (HRIR) measurements using variable head-to-torso angles in the horizontal plane allow for an investigation of the influence of the torso on the HRIR. This influence is extracted from the HRIRs and denoted as torso related impulse response (TRIR). On the other hand, a remaining component can be extracted from the HRIR, which is independent of the shoulder orientation. We refer to this as basis HRIR. A closer look on the TRIRs reveals that they carry strong energy for certain geometries between sound source and shoulders only. Resulting from these geometric analysis, we present an analytic system, modelling the TRIRs by linear regression, and the basis HRIRs with spherical harmonics (SH).

# 1 Introduction

Head related transfer functions (HRTFs) are the frequency equivalent of the HRIRs. They describe the physical path from an acoustical source to the ears. These transfer functions are caused by various reflections, diffractions, pressure stasis, and shadowing at pinna, head, and torso, too. Virtual acoustics systems for headphones use HRIR filters to create an impression of externalisation and to sharpen the sound source localisation [8, Wenzel et al.]. The naturalness and coherence of such virtual acoustics systems increases by taking room responses into account, i.e. early reflections and reverberation [1, Begault et al.]. The first reflections however arise form the shoulders. In addition the torso causes shadowing and therefore colorations of the HRTF for sound source elevations below the horizontal plane [4, Kirkeby et al.]. Furthermore recent works on dummy heads show efforts to include the degrees of freedom that evolve from flexible head-to-torso positions [5, Lindau et al.].

Besides these efforts no study on the influence of different head-torso angles on the HRIRs are known to the authors so far. This paper shall more detailedly investigate this aspect and provides measures to asses the influence of the angle in the horizontal plane. We present a distinction between a basis HRIR, that is independent of the torso influence and a complementary torso related impulse response (TRIR), that includes the information of the torso position. We assume the torso influence on the HRIRs to gain relevance in dynamical situations. As a consequence the system representation should allow the implementation of a real time rendering system. Spherical harmonics representation (SHs) offers an adequate solution, regarding the problem of time varying interpolation between different HRIRs [7, Noisternig and Majdak][2, Evans et al.]. Hence the basis HRIRs, that cover the whole sphere will be modelled as SHs. For the cases of strong TRIRs, they will be modelled over an interpolation using linear regression. In a first step the measurement procedure and the data processing will be explained. The concept of basis HRIR and TRIR and how they can be modelled will be represented in section (3). Finally we will give a resumé and an outlook of further approaches.



(a) Measuring setup

| LS. | Angle |
|---|---|
| 1 | -41,3° |
| 2 | -33,9° |
| 3 | -25,3° |
| 4 | -17,3° |
| 5 | -8,3° |
| 6 | 0° |
| 7 | 9,2° |
| 8 | 18° |
| 9 | 27,5° |
| 10 | 37,1° |
| 11 | 48,1° |
| 12 | 58,2° |
| 13 | 69,2° |
| 14 | 79,6° |
| 15 | 90,1° |

(b) Elevation angles

Figure 1: Measuring setup with a table of the corresponding loudspeaker position in elevation direction.

# 2 Measurement and data processing

## 2.1 Measurement

To determine a HRTF, two measurements are necessary. One of them has to be done in the virtual centre point of the head and is called reference measurement. The second one can be done at any measurement point within the ear channel. For a more detailed accomplishment, we refer to [6, Møller]. The measurement setup contains of a vertical array of 15 loudspeakers, arranged in a segment of a circle form -41° to 90° with a radius of 1.5 m. The precise angles of the arrangement can be seen in figure (1(b)).

Our measurements were done with a Brüel & Kjær head and torso simulator (HATS) 4128, that was positioned on a turntable in the middle of the measuring station. Measurements were done with exponential sweeps of 2 seconds, ranging from 20 Hz to 20 kHz. More about the properties of exponential sweeps can be read in [3, Farina]. For every of the 15 angles of elevation ($\vartheta$), 36 angles of azimuth ($\varphi$) have been measured in steps of 10 degrees. This measurement procedure was repeated for horizontal head-torso angles ($\delta$) ranging from 0 to 90 degree in steps of 10 degrees. As a forward rotation of the shoulder on the left ear side, results in a backward rotation on the right ear side, the total region of considered head-torso angles reaches from -90 to +90 degrees.[1]

[1] Positive $\delta$ denote the forward turned shoulder and negative the backward turned shoulder.

## 2.2 Data processing

The measured sweeps get deconvolved by dividing their spectrum through the spectrum of the original sweep (that was sent to the loudspeakers). They were transformed into the frequency domain by a $2^{17}$ point FFT.

$$H_{\varphi,\vartheta}(e^{j\theta}) = \frac{S_{\varphi,\vartheta}(e^{j\theta})}{S(e^{j\theta})} \tag{1}$$

The HRTFs are derived over the division of the frequency response measured in the ear (dependent on $\varphi$ and $\vartheta$) through the measurement of the reference point (only dependent on $\vartheta$).

$$HRTF_{\varphi,\vartheta}(e^{j\theta}) = \frac{H_{\varphi,\vartheta}(e^{j\theta})}{H_{\vartheta}(e^{j\theta})} \tag{2}$$

Finally the HRIRs were reduced with a Han window to a length of 128 samples. This corresponds to the time, sound needs to travel 1 meter approximately. Reflections that might have occurred due to the turntable are therefore windowed out.

# 3 Dynamical model with a basis HRIR and an additive TRIR

To highlight the influence of the torso to the HRIR, we extracted the torso relevant parts out of the measured HRIR. The remaining basis HRIR is blind to the torso position and therefore independent to the head-torso angle. The basis HRIRs are obtained by averaging over all HRIRs of different head-torso angles for a specific sound source direction.

$$\tilde{h}_{\varphi,\vartheta}[t] = \frac{1}{10}\sum_{i} h_{\varphi,\vartheta,\delta_i}[t] \tag{3}$$

The complementary difference signal, i.e. the TRIR

$$s_{\varphi,\vartheta,\delta}[t] = h_{\varphi,\vartheta,\delta}[t] - \tilde{h}_{\varphi,\vartheta}[t] \tag{4}$$

is dependent on the head-torso angle and carries the information of the torso influence. Figure (3) shows an averaged basis HRIR and the corresponding TRIRs. In our system the basis HRIR and the TRIRs are modelled separately. The basis HRIR is represented over spherical harmonics (3.1) and the TRIRs will be derived over linear regression and is added to the basis HRIR, like it can be seen in figure (2).

## 3.1 Basis HRIR in the SH domain

SHs represent the solutions of the angular part of the wave equation in polar coordinates. Any continuous and bounded function on a sphere can be expressed as a weighted expansion of SHs. Our basis HRIRs cover the whole sphere[2], therefore we can choose to represent them in the SH domain.

$$\tilde{h}(\varphi,\vartheta)[t] = \sum_{m=1}^{(N+1)^2} c_m[t] Y_m(\varphi,\vartheta) \tag{5}$$

[2]except for elevation angels lower than -41°.

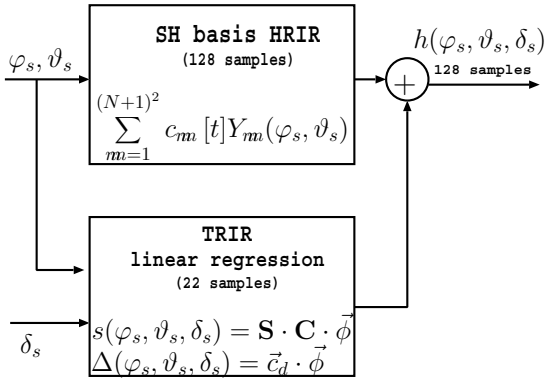



Figure 2: Block diagram of the dynamical system.

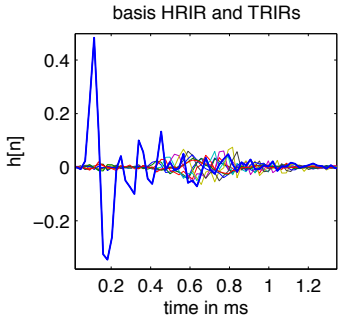



Figure 3: The averaged basis HRIR at 90° azimuth and 0° elevation with corresponding TRIRs.

The letter $m = 1 \ldots (N+1)^2$ stands for a natural number linearly indexing the spherical harmonics $Y_n^m(\varphi,\vartheta)$:

$$\begin{aligned} &\forall n = 0 \ldots N : m = -n, ..., n \\ &m = n^2 + n + 1 + m \end{aligned} \tag{6}$$

$Y_m(\varphi,\vartheta)$ are the spherical harmonics and $c_m$ are the weights that represent the directivity of the ear in this SH domain. Detailed treatise about spherical harmonics can be found in [9, Williams]. We can form a matrix containing the spherical harmonics for all discrete measurement points $(\varphi_k, \vartheta_k)$

$$\mathbf{Y} = [Y_m(\varphi_k, \vartheta_k)]_{M\times(N+1)^2} \tag{7}$$

where $M$ represents the number of measurement points. Rewriting equation (5) in matrix notation yields:

$$\vec{h}[t]_{[M\times 1]} = \boldsymbol{Y} \cdot \vec{c}[t]_{[(N+1)^2\times 1]} \tag{8}$$

The vector on the left hand side denotes the amplitude distribution of the basis HRIRs over the whole sphere[3] for every time sample $t$. The coefficient vector on the right hand side contains all SH coefficients. Its length depends on the order of the system ($(N+1)^2$). The order $N$ of a SH system itself depends on the resolution of the spherical sampling. As explained in [2, Evans] it can be simplified to the dependence of the resolution between 0° and 180° in one of the polar coordinates. Our measuring setup consists of $M_\vartheta = 15$ elevation angles. This would lead to a maximum order of $N = M_\vartheta - 1 = 14$. Due

[3]at our discrete measurement points respectively.

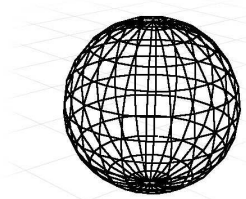


Figure 4: Partitioning of the sphere due to non equidistant sampling. Picture taken from: http://en.wikipedia.org/wiki/Sphere

to numerical reasons, the missing measurement points under -41.1° have to be interpolated, as it can be read in the following section. Therefore the maximum order can be increased to $N = 17$.

### 3.1.1 Calculation of the SH weights

To use the SHs as analytical description of our basis HRIRs the weighting vector $\vec{c}$ has to be determined for every time sample $t$. This can be done over the least squared error (LSE) algorithm. The error of the system can be easily written down as:

$$\vec{e} = (\vec{h} - \boldsymbol{Y}\vec{c}) \tag{9}$$

The quadratic error follows to:

$$\|\vec{e}\|^2 = (\vec{h} - \boldsymbol{Y}\vec{c})^H(\vec{h} - \boldsymbol{Y}\vec{c}) \tag{10}$$

But we have to consider that our measurement points were not equidistantly distributed. Figure (4) sketches the scheme of a spherically symmetric, but not equidistant partitioning. It can be seen that measurement points at the equator cover the largest areas. Therefore the quadratic error has to be weighted with the spherical surface segments that are covered by the measurement points. These weights are written down in the diagonal matrix $\boldsymbol{W}$.

$$\|\vec{e}_w\|^2 = (\vec{h} - \boldsymbol{Y}\vec{c})^H\boldsymbol{W}(\vec{h} - \boldsymbol{Y}\vec{c}) \tag{11}$$

To find the minimum of this function, it has to be derived after $\vec{c}$ and set to 0, which leads to the following solution:

$$\vec{c} = (\boldsymbol{Y}^H\boldsymbol{W}\boldsymbol{Y})^{-1}\boldsymbol{Y}^H\boldsymbol{W}\vec{h} \tag{12}$$

Calculating the weighted pseudo inverse $(\boldsymbol{Y}^H\boldsymbol{W}\boldsymbol{Y})^{-1}$ demands a regular matrix $\boldsymbol{Y}$. The missing measurement points under -41.1° however introduce linear dependencies, that reduce the achieved maximum order. To compensate this effect the missing basis HRIRs were interpolated. At the poles, all the HRIRs have to converge to one HRIR realization. We calculated the mean over all basis HRIRs at $\vartheta = -41°$ and declare the resulting signal as basis HRIR at -90°. Four further HRIRs between -90 degree and -41 degree were linearly interpolated in $\vartheta$ for every azimuth position. Padding the polar gap with interpolations increases the maximum order to $N = 17$. However the HRIRs of higher elevation angles benefit more form the increase than the

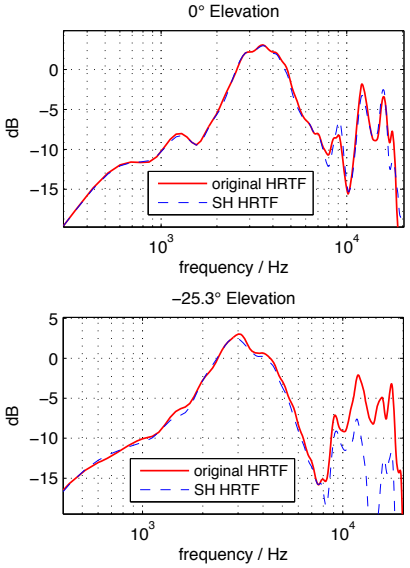



Figure 5: The spectrum of the original HRTF compared with the one of the SH representation for 0° and -25° elevation respectively.

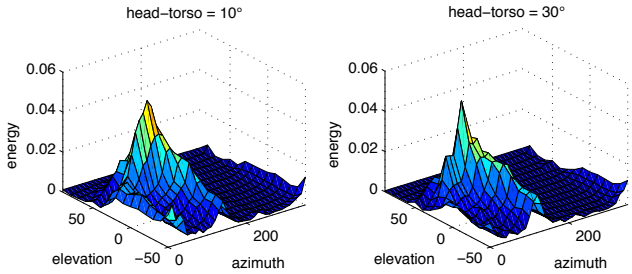



Figure 6: Energy distribution of the TRIRs for head-torso angles of 10° and 30°.

HRIR close to the interpolated ones. This can be seen by comparing the frequency response of HRIRs of different elevation angles in figure (5). Generally the upper cut off frequency in dependence of the truncation $N$ can be specified as:

$$f_c = \frac{Nc}{2\pi r} \tag{13}$$

where $c$ is the speed of sound and $r$ the radius of the sweet spot.

## 3.2 TRIR modelling

### 3.2.1 Analysis of TRIRs

As a first step the TRIRs are analysed in regard to their energy values. Only a few constellations of sound source direction and head-torso angle cause strong TRIRs. Figure (6) sketches the energy distribution over all elevation and azimuth angles for the head-torso angles $\delta = 10°$ and $\delta = 30°$, respectively.

However it is difficult to examine the energy distribution in our 4 dimensional space ($\varphi, \vartheta, \delta$, energy) over sketches of 19 different head-torso angles. It is more suitable to plot the surface of TRIRs with same energy values over the variables $\varphi, \vartheta$ and $\delta$. In figure (7)

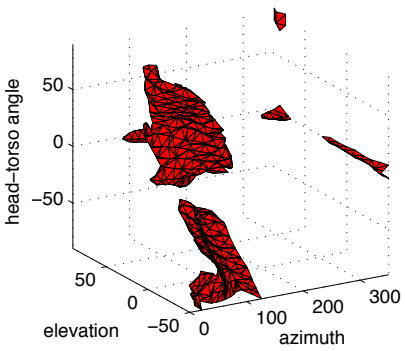



Figure 7: Measurements were done for different elevation, azimuth, and head-torso angles. The figure sketches the area of equally strong TRIRs in this 3-dimensional space. The depicted surface enclosures the TRIRs, whose energy lies 3 dB under the energy of the strongest TRIR.

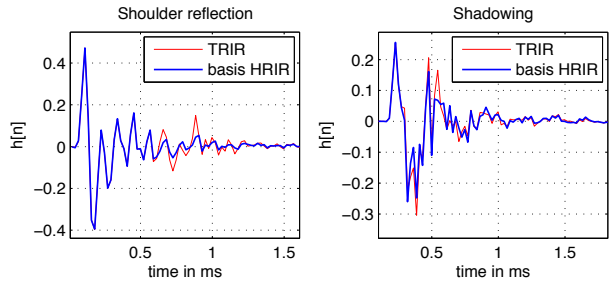
the area of the TRIRs at a threshhold of -3 dB relative to the energy of the strongest TRIR is outlined. Analysing these TRIRs with high energy values, shows that it is necessary to distinguish between the forward and the backward turned shoulder. While the first case causes patterns of reflections form sound source elevations between -20˚ and 40˚ (figure (8(a))), the second case shows shadowing at elevations[4] below -10˚ (figure (8(b))). Besides, it can be seen in figure (7) that the head-torso angle, that limits the sketch, changes with the azimuth angle. The sound source which provokes the strongest reflections always lies slightly behind the direction, where the shoulder is pointing to. If we stand in a normal position, sound sources from the left (around 120˚of azimuth) will cause strong reflections, if we turn the left shoulder forward (e.g. to 10˚ azimuth) also the sound that causes strong reflections will incident from the front (at 30˚ approximately). This is quite comprehensible. Only if sound source, sound sink and the normal vector of the reflecting surface lie in one plane, strong reflections can occur. This relation between the head-torso angle ($\delta$) and the azimuth angle ($\varphi$) can be used to determine the conditions that cause strong reflections. Hence we introduce a new variable $\varsigma = \varphi + \delta$ that helps to describe the limits of the region of strong TRIRs, which are, for the forward turned shoulder:

$$\begin{aligned} 100^\circ \leq &\varsigma \leq 140^\circ \\ -20^\circ \leq &\vartheta \leq 40^\circ \end{aligned} \tag{14}$$

For the backward turned shoulder we observe the effect of shadowing, but we still can define the same conditions for the region of interest like before. In this case, the strongest shadowing occurs if the sound source lies slightly in front of the direction, where the shoulder is pointing to. This is obvious, too. Only if the shoulder lies between sound source and the ear it can cause shadowing.

$$\begin{aligned} 10^\circ \leq &\varsigma \leq 50^\circ \\ &\vartheta < -10^\circ \end{aligned} \tag{15}$$

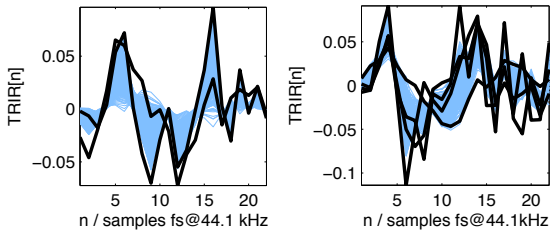
[4]not forgetting, that our measurements only reach to -41˚




(a) Shoulder reflection at 20˚ and the forward turned shoulder

(b) Shadowing at -41˚ elevation and the backward turned shoulder

Figure 8: Reflection and shadowing due to the torso.




(a) $\delta \geq 0$

(b) $\delta < 0$

Figure 9: The figures show the initial TRIRs (thick black) and all the interpolations for the forward turned shoulder and the backward turned shoulder respectively.

Due to symmetry reasons of the human body, the regions also exist for $360^\circ - \varsigma$.

### 3.2.2 Linear regression method

For our further considerations, as well as for our model, we only take the reduced set of the previously described strong TRIR into account. In addition the length of the TRIRs can be reduced to 22 samples. This leads to the following two questions.

- How to derive the correct TRIR without using a large data base?
- How to find the correct delay time after which the 22 samples of the TRIR have to be added to the basis HRIR?

It can be noticed that the impulse shapes of the TRIRs vary between few pronounced contours. This leads us to a simple interpolation method using linear regression. Weighting the depicted initial TRIRs in dependence of the angles $\varphi$, $\delta$ and the recently introduced variable $\varsigma$ yields the interpolation of of all the needed TRIR shapes. Again we have to distinguish between the foreward and the backward turned shoulder. While for positive head-torso angles it is sufficient to take two impulses as initial vectors, modelling the TRIRs for the backward turned shoulder demands 4 initial TRIRs. The initial TRIRs and the interpolated TRIRs are depicted in figure (9). These initial vectors are multiplied with the sum of the weighted input variables, whereas a linear regression of order 2 also takes the squares of the input variables into account. In the following the linear regression is deduced exemplarily for the case of two

initial TRIRs $s_1$ and $s_2$.

$$s_{[n]} = (a_1\vartheta + a_2\vartheta^2 + a_3\delta + a_4\delta^2 + a_5\varsigma + a_6\varsigma^2)s_{1[n]} + \\ +(b_1\vartheta + b_2\vartheta^2 + b_3\delta + b_4\delta^2 + b_5\varsigma + b_6\varsigma^2)s_{2[n]} \quad (16)$$

This equation can easier be written down in matrix form.

$$\begin{pmatrix} s_{[1]} \\ s_{[2]} \\ \vdots \\ s_{[22]} \end{pmatrix} = \begin{pmatrix} s_{1[1]} & s_{2[1]} \\ s_{1[2]} & s_{2[2]} \\ \vdots & \vdots \\ s_{1[22]} & s_{2[22]} \end{pmatrix} \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ b_1 & b_2 & b_3 & b_4 & b_5 & b_6 \end{pmatrix} \begin{pmatrix} \vartheta \\ \vartheta^2 \\ \delta \\ \delta^2 \\ \varsigma \\ \varsigma^2 \end{pmatrix} \quad (17)$$

To determine the coefficients $a_i$ and $b_i$ (subsumed in eq.(18) to the matrix $\boldsymbol{C}$) we need a matrix containing all considered input variable combinations $\Psi_{[6\times B]}$ on the one hand side and a matrix that consists of all TRIRs at the specified angles $\boldsymbol{S}_\Psi$ on the other hand side. $\boldsymbol{S}$ denotes the matrix of the two initial TRIRs.

$$\boldsymbol{S}_{\Psi[22\times B]} = \boldsymbol{S}_{[22\times 2]} \cdot \boldsymbol{C}_{[2\times 6]} \cdot \Psi_{[6\times B]} \quad (18)$$

Again the coefficients are gained over the LSE approach leading to:

$$\boldsymbol{C} = (\boldsymbol{S}^H\boldsymbol{S})^{-1}\boldsymbol{S}^H\boldsymbol{S}_\Psi\Psi^H(\Psi\Psi^H)^{-1} \quad (19)$$

The same algorithm is used to determine the TRIRs for the backward turned shoulder with the only difference that 4 initial TRIRs are needed for the interpolation. The delay times ($\Delta$) are calculated over linear regression with 6 coefficients, too.

$$\vec{\Delta} = \vec{c}_d \cdot \Psi \quad (20)$$

### 3.2.3 Alternative evaluation method

In order to consider psychoacoustic insights on the sensitivity of the ears, we evaluated the TRIRs in third octave bands in regard to masking levels. The resulting areas, depicted in figure (10), are not as continuous as in the previous case and expand for some cases to higher elevation angles. In principle, the additional areas can be modelled by the same interpolation technique that was presented in section (3.2.2), by of course increasing the number of initial TRIRs and coefficient.

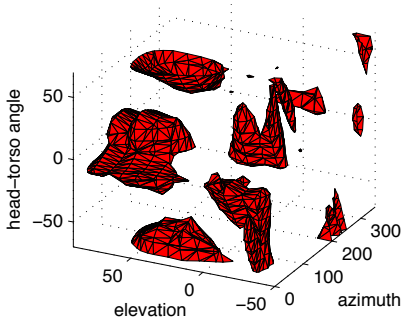



Figure 10: The area of the strongest TRIRs, analysed in third octave bands.

# 4 Conclusion and Outlook

We measured HRIRs for 10 different horizontal head-torso angles. By averaging over these head-torso angles we derive a basis HRIR that is independent of the torso position. The difference signal (i.e. the TRIR) between the torso dependent HRIR and the averaged basis HRIR represents the influence of the torso to the HRIR. These TRIRs are evaluated in regard to their energy. We depicted the constellations of sound source direction and torso orientation that cause the strongest TRIRs. For this specified set of constellations the TRIRs are modelled over an interpolation using linear regression. The basis HRIRs are modelled by spherical harmonics.

Adding the TRIRs to the basis HRIRs produces noticeable differences. We assume these cues to gain importance in dynamical situations. Therefore a dynamical implementation of the model shall be realized to evaluate the perceptive relevance. Listening test can answer how far and which TRIRs are worth to be modelled, which TRIRs rather cause sound colorations and which improve the sensation of externalisation. In a next step the model could be extended to a nodding and side bended head. Despite the insufficient dispostion of these 3 degrees of freedom one could think of modelling these TRIRs as hyperspherical harmonics on a 5 dimensional sphere, too.

# References


[1] Durand R. Begault, Elizabeth M. Wenzel, and Mark R. Anderson. "Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source". *J. Audio Eng. Soc.*, 49, No. 10:904–916, October 2001.

[2] Michael J. Evans, James A. S. Angus, and Anthony I. Tew. "Analyzing head-related transfer function measurements using surface spherical harmonics". *Acoustical Society of America*, 104, No. 4:2400–2411, October 1998.

[3] Angelo Farina. "Simultaneous Measurements of Impulse Response and Distortion with a Swept-Sine Technique". *AES 108th convention, Paris*, February 2000.

[4] Ole Kirkeby, Eira T. Seppälä, Asta Kärkkäinen, Leo Kärkkäinen, and Tomi Huttunen. "Some Effects of the Torso on Head-Related Transfer Functions". *122nd AES-Convention, Vienna*, May, 2007.

[5] Alexander Lindau and Stefan Weinzierl. "FABIAN - Schnelle Erfassung binauraler Raumimpulsantworten in mehreren Freiheitsgraden". *DAGA 2007 Conference, Stuttgart*, March 2007.

[6] Henrik Møller, Michael F. Sørensen, Dorte Hammershøi, and Clemen B. Jensen. "Head-Related Transfer Function of Human Subjects". *J. Audio Engeneering Society*, 43, Issue 5:300–321, May 1995.

[7] M. Noisternig and P. Majdak. "Implementation kopfpositionsbezogener Binauraltechnik". *Diploma Thesis, TU Graz, Institute of Electronic Music and Acoustics*, 2002.

[8] Elisabeth M. Wenzel, Joel D. Miller, and Jonathan S. Abel. "Sound Lab: A Real-Time, Software-Based System for the Study of Spatial Hearing". *AES 108th convention, Paris*, February 2000.

[9] E. G. Williams. "Fourier Acoustics". *Academic Press, San Diego*, 1999.