DOI: 10.61782/fa.2023.0226

TORSO REFLECTION MODEL FOR DYNAMIC HEAD-RELATED TRANSFER FUNCTIONS

Jonas Reijniers* **Bart Partoens** **Herbert Peremans**
University of Antwerp, 2000 Antwerp, Belgium

ABSTRACT

Present-day virtual audio systems make use of 'static' Head Related Transfer Functions (HRTF), the head being fixed to the torso. If one wants to take into account head-above-torso movement, it has proven impractical to resort to a stored set of 'dynamic' HRTFs, as it would be too large. An alternative approach is to model the torso reflections separately and add these to the direct path impulse response. Here, we follow this approach, while limiting ourselves to modelling only the most prominent feature of the torso reflections: the time lag between direct path and torso reflection. We propose a simplified geometric model of the head, neck and shoulders and optimize its model parameters such that the modelled time lags obtained via ray-tracing correspond maximally with those that were extracted from a measured set of dynamic HRTFs. We show that: (1) the simulated time lags can be fit quite adequately to those extracted from the dynamic HRTFs; (2) the optimized geometrical model parameters have physical meaning, as they can be related to morphological features of the individual subject; and (3) the time lags can be calculated using a computationally inexpensive analytical expression, making it well suited for real-time implementation.

Keywords: *HRTF, torso, ray-tracing, snowman model, optimization.*

1. INTRODUCTION

The importance of torso reflections for sound localization has been studied before, albeit not extensively. Algazi *et al.* isolated the torso reflections from the direct path impulse responses (IR) and showed in a psychoacoustic experiment that torso reflections provide low-frequency cues for source elevation [1]. Moreover, using ray-tracing and simple geometrical models of the head and torso (the 'snowman' model), they were able to explain the torso reflection delays that give rise to these elevation dependent cues. In a follow-up study, using more elaborate techniques to solve Helmholtz equation, they established that the simple geometric models of head and torso indeed provide good approximations of the HRTF [2].

Another line of research is of a practical nature: how to combine the torso reflections with the direct path IRs. Algazi *et al.* took this approach to deal with the difficulty of measuring the HRTF at low frequencies, by replacing the latter by a modelled low-frequency HRTF [2, 3]. This mixed structural modelling was also considered for modelling individual HRTFs: each of the separate reflections (due to head, torso, pinna) is then modelled separately according to the individual's geometry and subsequently mixed together into an individualized HRTF [4].

These studies all assumed a static posture, with the head in the typical upright position. Guldenschuh *et al.* [5] studied torso reflection under variable head-above torso orientations (HATO). Using a dummy head, they measured the HRIR for different HATOs (yaw, see Fig. 1) and separated the torso dependent part from the torso invariant part. To deal with the high-dimensionality of the HRIR database due to variable HATOs, they proposed a mixed model, in which the invariant HRIR is supplemented with a dynamic, HATO dependent part, which they interpolated via linear regression. Brinkmann *et al.* [6] carried out sim-

****Corresponding author***: jonas.reijniers@uantwerpen.be.*

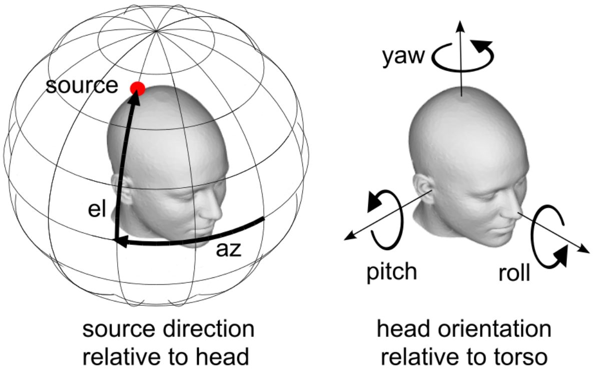
source
el
az
yaw
pitch
roll
source direction relative to head
head orientation relative to torso

Figure 1. Reference frames defining the source direction (left) and HATO (right).

ilar measurements on another dummy head (FABIAN), again under different HATOs (again yaw), but at a much higher resolution. Moreover, this database is freely accessible online [7]. In line with Guldenschuh's findings, they found that the torso-dependent variations were in general rather small compared to the direct path IR. However, in a static discrimination experiment, these variations were shown to be audible, at least for some directions [7].

Nevertheless, it seems that this research has not yet materialized into a real-time implementation of variable torso reflections in a dynamic virtual audio system (VAS). The main difficulty for such an implementation is that allowing for variable torso reflection adds three extra dimensions to the HRTF. Whereas a typical HRTF requires two angles (for source direction), the HATO requires three extra angles (yaw, pitch, roll), see Fig. 1. Given this total of five dimensions, it is difficult to measure the HRTF sufficiently dense for each dimension as required for HRIR interpolation [5, 6].

Therefore, in this paper, we take a different approach. We also start from a set of HRIRs that have been measured for a limited set of HATOs. We then separate the torso invariant from the variant part in a novel way, by means of time-alignment and subsequent spherical harmonics (SH) representation. Next, instead of interpolating between torso reflections with sparsely sampled HATOs, we use the latter to *model* the torso reflections for *every possible* HATO. To this end, we elaborate on the strategy set out by Algazi *et al.* [1, 8] and model the torso reflections using two simplifications: (1) we consider only the time delay between direct path and torso reflection relevant; (2) this time delay is modeled by simple ray-tracing using an adapted version of the snowman model, where the head is modelled as a sphere and the torso as an ellipsoid. Our main contributions are adapting the snowman model for different HATOs and producing an analytical expression for the torso delay, which speeds up the calculation significantly. Finally, we optimize the model parameters of the snowman model. Whereas Algazi *et al.* [1] estimated the model parameters based on the geometry of the KEMAR manikin, we optimize these such that the simulated torso delays correspond maximally with those of the measured torso reflections. Hence, the model parameters are extracted directly from the measured HRIRs with variable HATOs. This way we set up the framework that should allow a rather simple dynamic implementation of (an approximation of a feature of the) torso reflections.

In this manuscript, we limit ourselves to presenting the outline of the model and the extraction of the model parameters. We will do this by means of the FABIAN data set [7].

2. METHOD

To avoid any ambiguity, we use the following notation:

$$\mathrm{HRIR}(\boldsymbol{\theta},\boldsymbol{\xi}) = \mathrm{hRIR}(\boldsymbol{\theta}) + \mathrm{tRIR}(\boldsymbol{\theta},\boldsymbol{\xi}), \qquad (1)$$

with $\boldsymbol{\theta}$ denoting the source direction (azimuth, elevation) and $\boldsymbol{\xi}$ the HATO (yaw, pitch, roll). As the HRIR is generally used for the transfer function due to the combined filtering of head and torso, we stick to this notation. The separate contributions to the HRIR due to the head (direct path) and the torso (reflected path) are then referred to as hRIR($\boldsymbol{\theta}$) and tRIR($\boldsymbol{\theta}, \boldsymbol{\xi}$) respectively.

2.1 HRIR data with variable HATOs

To be able to separate the torso reflections tRIR from the direct path hRIR, it is necessary for the HRIRs to have been measured for similar source directions $\boldsymbol{\theta}$ with different HATOs $\boldsymbol{\xi}$.

Such a database was obtained by Brinkmann *et al.* [6] and is available online [7]. The HRTF was measured on a custom build dummy head and torso (FABIAN dummy), for different rotations of the head with respect to the torso in the horizontal plane. According to our choice of coordinates, this means that the yaw angle was varied between $[-50^\circ, 50^\circ]$ in steps of 10°. The other angles which define the HATO, pitch and roll, were kept zero. Source directions were sampled with a resolution of 2° in elevation and 2° in azimuth, amounting to a total of 11,950 source directions for each of the 11 HATOs.

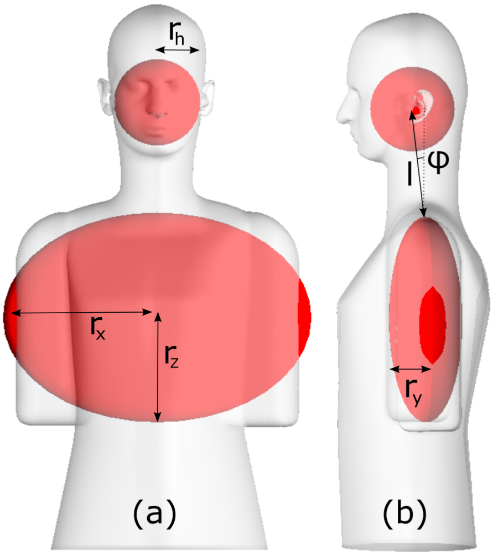

Figure 2. The dummy that was used for the HRIR measurements with variable HATOs is shown in light grey (in case yaw=0°). The geometrical model that is considered in this paper is shown in red; the six model parameters are indicated. The dimensions of the geometrical model are those obtained via the optimization described in the text.

2.2 Extracting the torso reflections

First we separate the torso reflection tRIR from the direct path hRIR. Algazi *et al.* [1] achieved this by reducing the filtering of the ear, either by removing the ear (in case of the KEMAR) or by means of a swimming cap (human subjects). Yet, as mentioned before, if the HRTF is measured for different HATOs $\boldsymbol{\xi}$, it is possible to separate the HRIR into a $\boldsymbol{\xi}$-invariant (hRIR) and $\boldsymbol{\xi}$-variant part (tRIR). This strategy was used by Guldenschuh *et al.* [5] by time aligning and averaging HRIRs that correspond with identical source directions but different HATOs. Here, we take a similar approach, but use SH representation instead of averaging:

1. For each of the measured HRIRs, the source direction is expressed with respect to the head reference frame.

2. Next, all the measured HRIRs are pooled into one single data set, irrespective of source direction or HATO.

3. Each of the measured HRIRs is then time-aligned on the direct path IR, based on first-onset alignment [9].

4. Next, we consider the source direction only (and temporarily forget about HATO) and express the resulting set of aligned HRIRs in SH basis functions with truncation order 15, using Tikhonov regularization as in Ref. [10]. This results in a set of SH coefficients from which we reconstruct the HRIR for each measured direction $\boldsymbol{\theta}$. As all measured HRIRs with different HATOs were pooled, HRIRs of similar source direction but different HATO will be represented by the same SH representation. The SH representation captures that part of the HRIRs that is *invariant* to the exact HATO and only depends on the source direction. The resulting interpolated HRIRs are denoted as hRIR($\boldsymbol{\theta}$).

5. Subtracting the hRIR($\boldsymbol{\theta}$) from the measured HRIR($\boldsymbol{\theta}, \boldsymbol{\xi}$), we then obtain tRIR($\boldsymbol{\theta}, \boldsymbol{\xi}$), i.e., the part that could not be well presented by the SH representation, due to the variable torso reflections corresponding with different HATOs.

An example of the results of applying this procedure is shown in Fig. 3.

2.3 Modelling torso reflections

Modelling only delays As the torso reflections turn out to be fairly complex, we simplify according to the following strategy: instead of trying to reproduce the torso reflections with high fidelity, we focus on a single feature of the torso reflections: the amount they lag behind the direct path response. We assume that this time delay is the most relevant perceptual feature, a conjecture that is also corroborated by psychoacoustic experiments performed by Algazi *et al.* [1]. They showed that when the direct path hRIR (modelled by a spherical head model) was supplemented with a torso reflection that was modelled as a delayed direct path hRIR (with a fixed 9.5dB lower intensity), this clearly resulted in improved elevation localization. As modelling delay time is also easier than modelling the frequency content of the torso reflection, it seems sensible to first try to model the torso reflection delays, implement these in a dynamic VAS and see whether such a first-order approximation already makes a

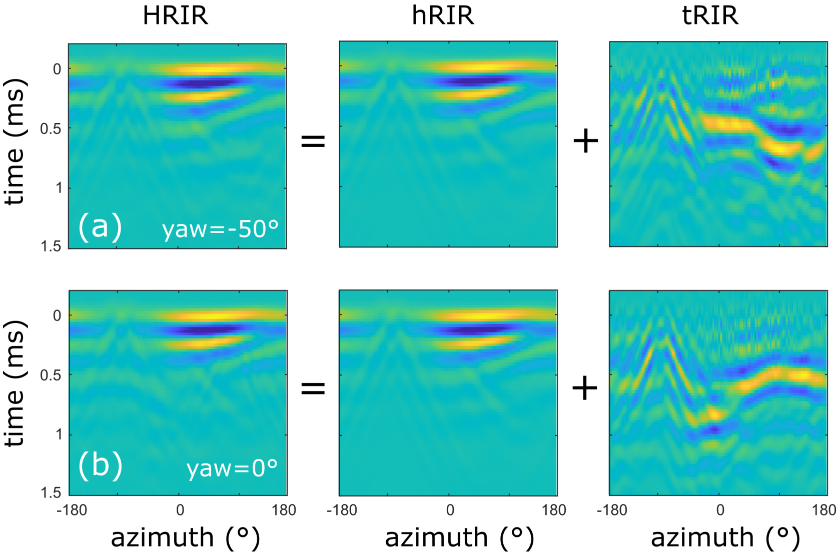

Figure 3. Example of how the HRIR is separated into the hRIR and tRIR. The HRIR is shown for zero elevation as function of azimuth angle for two different HATOs: with (a) yaw=$-50°$ and (b) $0°$. The hRIR is the part that is captured by the SH representation (identical for both HATOs), the residual part is the tRIR which is HATO dependent. Note that the tRIR has been scaled by a factor 10.

perceptual difference, before modelling the torso reflection with higher fidelity.

Geometric model We opt for a simple geometric model, similar to the one introduced by Algazi *et al.* [1]. We consider a spherical head model, where the ears are located at opposites sides of a spherical head with radius r_h, the centre of the head at their midpoint (i.e., we do not consider any additional ear offsets). The centre of the head is connected via a straight neck with length l to the torso, which is modelled as an ellipsoid with dimension r_x, r_y and r_z, see Fig. 2. We assume that the head can only rotate around the neck joint (and not around the inter-aural axis), which is located at the top of the ellipsoidal torso. The final model parameter is φ, which is the angle between the modelled neck and the vertical axis, when the subject is looking straight ahead (zero elevation).

Ray-tracing As in Ref. [1], we use simple geometric ray-tracing to model the time delay between the direct response and the torso reflection. Given a particular source direction, HATO and set of geometric model parameters, we calculate the direct path length (source-ear) and the torso reflected path length (source-torso-ear). This requires finding the point of reflection $\boldsymbol{t}$ on the ellipsoid, given the ear position $\boldsymbol{e}$ and source position $\boldsymbol{s}$. In Ref. [1] this is achieved using a fairly complex procedure, where torso delays are calculated by stepping systematically over the surface of the ellipsoid to find the source direction that would cause a reflection at that point. Using a SH interpolation, the delay can then be calculated for any source direction. In case of a dynamic VAS, where the HATO is to be updated in real-time, this approach is no longer feasible. To allow for real-time implementation, Algazi *et al.* took a different approach in a follow-up paper [8]. Assuming a spherical head and torso and assuming a source infinitely far, they reduced the problem to a nonlinear equation, which was then solved numerically.

Analytical solution It is possible though to calculate the point of reflection $\boldsymbol{t}$ on an ellipsoid *analytically*, for any two points $\boldsymbol{s}$ (source) and $\boldsymbol{e}$ (ear), irrespective of their distance to the torso. This speeds up the calculation significantly and is also applicable to sources nearby. Such an analytical result is obtained as follows:

1. express $\boldsymbol{s}$ and $\boldsymbol{e}$ in the coordinate system of the torso, with the ellipsoid centre as $\mathbf{0}$ and the axes along the main axes of the ellipsoid;
2. if s_y is negative, mirror $\boldsymbol{s}$ and $\boldsymbol{e}$ along the xz-plane;
3. scale such that the ellipsoid is transformed to a unit

sphere;

4. rotate such that $\boldsymbol{s}'$ is on the positive y-axis, i.e. $\boldsymbol{s}' = (0, s'_y, 0)$ and then rotate around the y axis such that $\boldsymbol{e}' = (0, e'_y, e'_z)$ is in the yz-plane;

5. the problem is now reduced to two dimensions: find the shortest path between points (e'_y, e'_z) and $(s'_y, 0)$ via point $(\cos\gamma, \sin\gamma)$ on the unit circle. If we assume that $\beta = \tan^{-1}(e'_z/e'_x)$, this boils down to solving

   $$s'_y = \|\boldsymbol{e}'\| \csc\gamma \left[s'_y \sin(2\gamma - \beta) + \sin(\beta - \gamma)\right]$$

   for γ. This equation can be rewritten as a quartic equation in γ, which can be solved analytically. Next, select the real solution $\boldsymbol{t}' = (0, \cos\gamma, \sin\gamma)$ for which the summed distance is shortest;

6. do the inverse of the sequence of transformations above to arrive at $\boldsymbol{t}$, the desired point of reflection on the ellipsoid.

Having this point of reflection $\boldsymbol{t}$, the time delay Δt is the difference between the direct path and the reflected path length, divided by the speed of sound. If the line connecting the ear $\boldsymbol{e}$ and the source $\boldsymbol{s}$ or the ear $\boldsymbol{e}$ and the torso point of reflection $\boldsymbol{t}$ intersects the spherical head, we use the great circle distance between the point of intersection and the ear (we use the shortest distance on the surface of the spherical head, instead of the Euclidean distance through the head). Note that if the line connecting $\boldsymbol{s}$ and $\boldsymbol{e}$ intersects the ellipsoid, there is no direct path, and consequently $\Delta t = 0$.

2.4 Model optimization

Optimization criterion The proposed geometric model has six model parameters (see Fig. 5) which we want to optimize such that the produced time delays correspond best to those observed in the tRIRs. To this end, it is necessary to make explicit what is meant by 'best'. Given a certain set of model parameters, we first calculate the corresponding time delays Δt for each of the N tRIRs. For each HRIR measurement, we use its particular source direction and HATO to calculate the left and right time delay Δt_i^L and Δt_i^R by means of the analytical formula derived above. Next, to assess the quality of the the model output, the tRIRs are evaluated at the corresponding time delays and summing over these values, we arrive at the following quality measure (objective function)

$$Q = \sum_{i=1}^{N} \sum_{k=L,R} \mathrm{tRIR}_i^k(\Delta t_i^k) \tag{2}$$

to be maximized. Indeed, the higher the Q-value, the better the overall torso peak timing is reproduced by the model. Note that in order to give preference to the highest peak timing, the tRIR is represented in the linear domain.

As it is yet unclear which torso reflections will be most important, we consider two different optimization criteria. First, we consider the tRIR as is. As a consequence, strong torso reflections will dominate the Q-value, compared to weaker ones. Such a choice of objective function implicitly assumes that the stronger the reflection, the more perceptually relevant it is, no matter the strength of the direct path response. Secondly, one could also argue that it is the strength of the torso reflection relative to that of the direct path response that is relevant. Indeed, the higher the energy ratio of the reflection to the direct response, the larger the amplitude of the comb filter modulations and thus the easier it is to perceive.

Optimization strategy We optimize the model parameters by brute force. We select a random subset of 5000 tRIRs from the merged data set, vary each of the model parameters over a wide range (with 1 cm and 1° resolution), calculate for each parameter configuration the corresponding Q value and select the configuration for which the Q is highest. Such a strategy is feasible since the time delays can be solved analytically.

3. RESULTS

3.1 Extracting the torso reflections

In Fig. 3, as an illustration, we have shown the measured HRIRs for two different HATOs and their respective hRIR and tRIR, if we follow the procedure outlined in Sec. 2.2. The torso reflection tRIR has far less energy than the direct path hRIR, therefore, in the figure it was scaled by a factor 10.

To get an idea of the energy contained in the tRIRs, we have plotted in Fig. 4 the energy of the right ear hRIR and tRIR for the 5000 randomly sampled HRIRs that were used for the model optimization. The energy of the hRIR varies smoothly over the source directions and is highest for the ipsilateral ear. The energy of the tRIR shows a much higher local variation, because neighbouring sampled source directions have different HATOs. Still, we

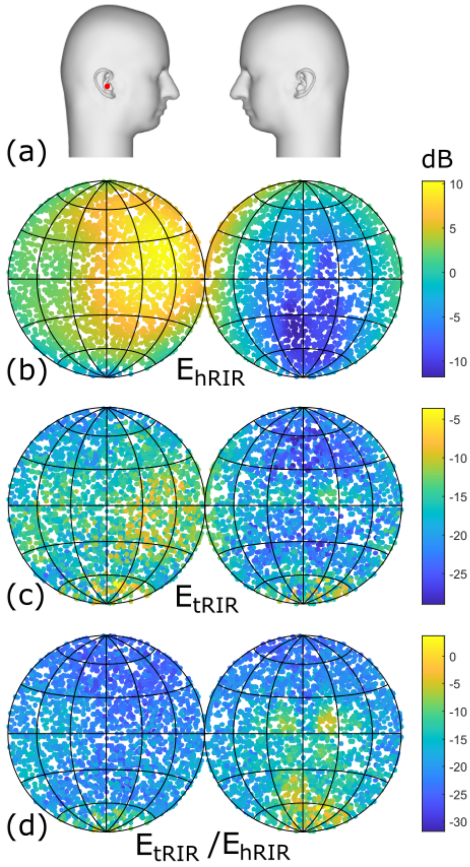

Figure 4. The energy of the right ear hRIR, tRIR and their ratio, shown for 5000 HRIRs which were randomly sampled from the merged HRIR dataset with different HATOs. The data is plotted using a Lambert azimuthal equal area projection method.

see that overall the energy of the tRIR is still higher on the ipsilateral side, although less so for higher elevations (compared to the hRIR). Yet, if we look at the energy of the tRIR relative to that of the hRIR, we notice that this ratio is higher on the contralateral side. In general, though, the energy of the tRIR is (much) smaller than that of the hRIR.

3.2 Model optimization

In Table. 1 we show the optimized geometrical model parameters for the two different optimization criteria: based on the tRIR scaled by the energy of the hRIR (relative) and the tRIR as is (absolute). We see that the choice of the exact criterion only slightly affects r_y and r_z; all other parameters are the same.

Table 1. Optimized model parameters (see Fig.5) for three different optimization criteria: scaled tRIR, absolute tRIR and scaled monaural (left ear) tRIR.

	relative	absolute	monaural
r_h (cm)	7	7	7
φ (°)	7	7	7
ℓ (cm)	18	18	18
r_x (cm)	25	25	25
r_y (cm)	6	7	6
r_z (cm)	17	17	16

To compare the optimized geometrical model with the actual torso that was used for the measurement of the HRIRs, we show an overlay of the two surface meshes in Fig. 2, aligned on the inter-aural axis. We notice that the head radius is similar to that at the ear level (notice the dark red at the entrance of the ear canal), yet the neck is slightly longer and the r_y dimension of the torso is somewhat smaller than that of the FABIAN torso. But overall, we see that ray-tracing on a simplified geometrical model results in a snowman model that, given the simplicity of the model, is in fairly good agreement with the actual torso.

We want to emphasize the fact that these model parameters were optimized based solely on the torso delay times and that e.g. the ITD, which is a good predictor for head size, was not taken into account. To prove this point, we have also included in Table. 1 the optimized model parameters in case only monaural (left ear) tRIRs were considered, showing that the obtained values are almost identical.

3.3 Time delay validation

To assess the quality of the time delays Δt produced by the model, in Fig. 5, we plot these on top of the tRIRs for three different HATOs: yaw = $-50°, 0, 50°$. We show the data for different elevations (ranging between $[30°, -30°]$ in steps of $6°$) and for each elevation the data are plotted for azimuths covering the full circle. We notice that the tRIRs vary quite significantly for different HATOs, but the model output Δt manages to trace the major peak quite well, especially for those tRIRs that contain a lot of energy

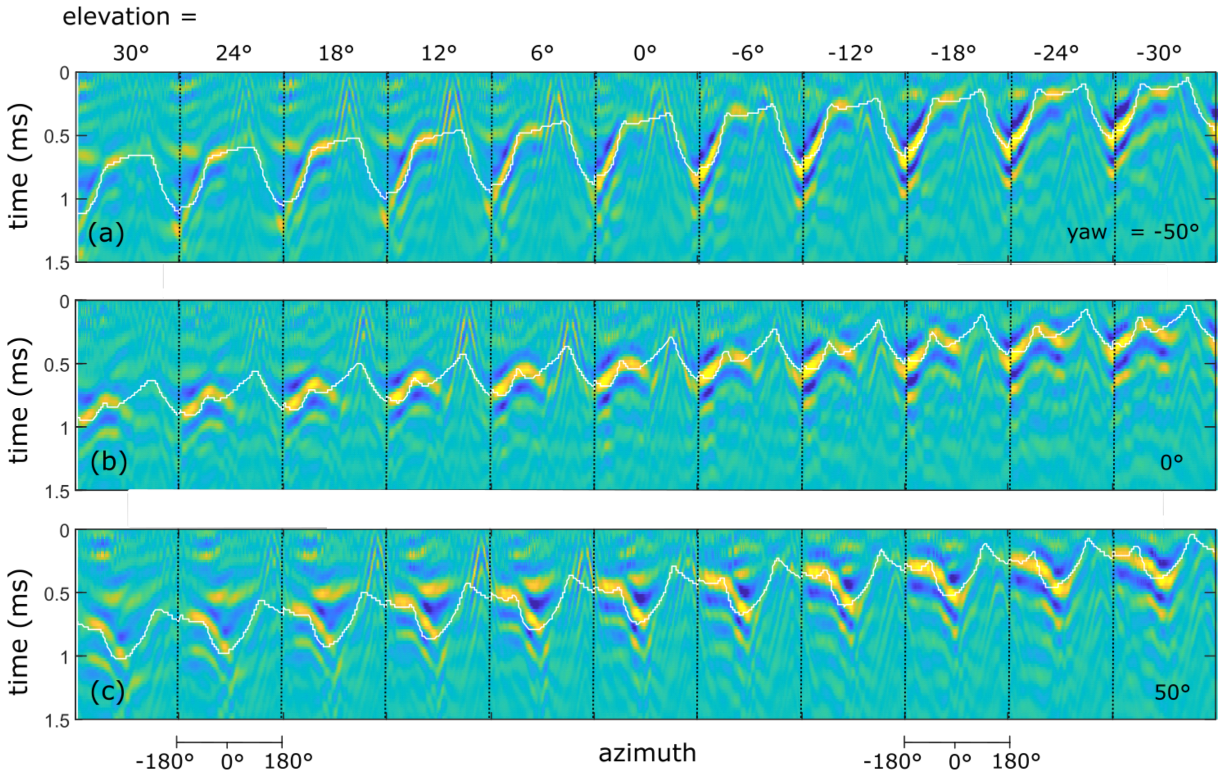

Figure 5. The right ear tRIR(θ, ξ) is shown for three different HATOs as function of azimuth and elevation. The delay time produced by the optimized model is shown in white.

(bright yellow). Hence, we conclude that ray-tracing on a geometrical model does a fairly good job at reproducing the torso time delays.

4. DISCUSSION

We have shown that one can optimize a simple geometric model such that, using ray-tracing, one can reproduce the torso reflection delay quite well, at least for the higher intensity torso reflections.

There are discrepancies, though, which are partly due to the simplified geometrical model of the head and torso and partly to the ray-tracing approximation. Indeed, the ray-tracing approximation works well if the wavelength is small compared to the dimensions of the reflective object, which a condition that is not satisfied for the torso reflections. As is visible on Fig. 3 and Fig. 5, the spectrum of the torso reflection attains its maximum around 4 kHz (period of $\approx$ 2.5 ms), which corresponds to a wavelength of about 9 cm, which is rather large compared to the dimensions of and the distances between the head and torso. Moreover, we only consider one single path (the shortest), while in reality sound waves may be arriving with similar intensity via different routes. The resulting interference hampers the estimation of a timing delay; it even makes the concept of a well defined time delay meaningless.

In the near future, we will apply the presented methodology to dynamically measured HRIRs, obtained through the measurement procedure described in Refs. [10, 11]. As this method measures the HRIR of real human subjects under different HATOs, this should allow to estimate the torso reflection model parameters of that individual such that both the hRIR *and* the tRIR can be individualized. Moreover, this will allow to evaluate the proposed analytical model in case of HATOs with variable nonzero pitch and roll angles.

In a next step, we will then implement these individualized torso reflection delays (and the individual hRIR) in a head-tracked VAS and investigate whether these torso reflections make a perceptual difference, and if so, how they may improve the experience of virtual 3D audio.

5. ACKNOWLEDGEMENTS

This research was supported by the Research Foundation Flanders (FWO) under Grant no. G023619N and the Agency for Innovation and Entrepreneurship (VLAIO).

6. REFERENCES

[1] V. Algazi, C. Avendano, and R. Duda, “Elevation localization and head-related transfer function analysis at low frequencies,” *J. Acoust. Soc. Am.*, vol. 109, no. 3, pp. 1110–1122, 2001.

[2] V. Algazi, R. Duda, R. Duraiswami, N.A.Gumerov, and Z. Tang, “Approximating the head-related transfer function using simple geometric models of the head and torso,” *J. Acoust. Soc. Am*, vol. 112, no. 5, pp. 2053–2064, 2002.

[3] C. P. Brown and R. O. Duda, “A structural model for binaural sound synthesis,” *IEEE transactions on speech and audio processing*, vol. 6, no. 5, pp. 476–488, 1998.

[4] M. Geronazzo, S. Spagnol, and F. Avanzini, “Mixed structural modeling of head-related transfer functions for customized binaural audio delivery,” in *IEEE 18th International Conference on Digital Signal Processing (DSP)*, pp. 1–8, 2013.

[5] M. Guldenschuh, A. Sontacchi, F. Zotter, and R. Holdrich, “HRTF modelling in due consideration variable torso reflections,” *J. Acoust. Soc. Am.*, vol. 123, no. 5, pp. 3080–3080, 2008.

[6] F. Brinkmann, R. Roden, A. Lindau, and S. Weinzier, “Audibility and interpolation of head-above-torso orientation in binaural technology,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 931–942, 2015.

[7] F. Brinkmann, A. Lindau, S. Weinzierl, M. Müller-Trapet, R. Opdam, and M. Vorländer, “A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations,” *J. Acoust. Soc. Am.*, vol. 65, no. 10, pp. 841–848, 2017.

[8] R. Algazi, R. O. Duda, and D. M. Thompson, “The use of head-and-torso models for improved spatial sound synthesis,” in *AES 113th convention*, (Los Angeles, USA), pp. 1–18, 2002.

[9] A. Andreopoulou and B. F. G. Katz, “Identification of perceptually relevant methods of inter-aural time difference estimation,” *J. Acoust. Soc. Am.*, vol. 142, no. 2, pp. 588–598, 2017.

[10] J. Reijniers, B. Partoens, J. Steckel, and H. Peremans, “HRTF measurement by means of unsupervised head movements with respect to a single fixed speaker,” *IEEE Access*, vol. 8, pp. 92287–92300, 2020.

[11] J. Reijniers, B. Partoens, and H. Peremans, “DIY measurement of your personal HRTF at home: Low-cost, fast and validated,” in *Audio Engineering Society Convention 143*, (New York, USA), 2017.