Audio Engineering Society

Convention Express Paper 232

Presented at the AES 156th Convention
2024 June 15–17, Madrid, Spain

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

HRTF Model Using Decoupled Rear-Front and Elevation Cues and Applications

Christof Faller[1], Michele Ebri[2], Alexis Favrot[1], and Alfonso Oliva[2]

[1]*Illusonic GmbH, Greifensee, Switzerland*
[2]*Ferrari S.p.A., Maranello, Italy*

Correspondence should be addressed to Christof Faller (christof.faller@illusonic.com)

ABSTRACT

A head-related-transfer-function (HRTF) model is presented, using delays, shelving filters for interaural-level-difference (ILD), and biquad equalizers for front-rear and elevation monaural cues. Based on averaged data from a few dozen measured HRTF sets, mappings are defined from azimuth and elevation to delays, ILD filters, and the monaural cues equalizers. The HRTF data implies that separate equalizers for front-rear and elevation cues, respectively, can approximate these. In addition to using the model directly to render binaural sound, the model can be used to convert omnidirectional-microphone array impulse responses or binaural-room-impulse-responses (BRIRs) to HRTFs, and to enable head-tracking for static HRTFs or BRIRs.

1 Introduction

Binaural rendering [1, 2] is relatively widely used for applications which present multi-channel surround audio over headphones, e.g. gaming, mobile video, and recent popular spatial audio music.

Without person-individualized head-related-transfer-functions (HRTFs), timbre and sound source localization can be impaired due to a mismatch between real signals at the subject's own ears and those simulated by the employed HRTFs [3, 4].

The goal of the work presented in this paper has been a generic model for non-individualized HRTFs. Timbral changes compared to direct playback of the signals (or downmixes thereof) over headphones shall be small, while providing interaural and monaural cues [5] for 3D localization of sources.

The use of a model as presented here has advantages compared to using a set of HRTF finite-impulse-response (FIR) filters:

- The model processes sound at any azimuth and elevation, thus no HRTF interpolation is needed for head-tracking.

- The computational complexity of the delay and infinite-impulse-response (IIR) filters used by the model is low.
- As the model parameters are determined by considering measured HRTF sets of 51 persons, it should represent good "average HRTFs", i.e. HRTFs which work reasonably well for many listeners.

We previously used a simple spherical head model (SHM) [6] for head-tracking [7]. The model presented here extends the SHM with rear-front and elevation monaural cues, i.e. equalizers which change the timbre, such that rear-versus-front and elevation are perceived. With only a SHM, any modeled HRTFs with directions on the same cone-of-confusion [5] yield identical results. That is, rear versus front and elevation are not differentiated.

In the following we are briefly explaining particularities of the presented model and reasoning for it:

- **Spherical Head Model:** The SHM is used to generate interaural-level-differences (ILDs) and interaural-time-differences (ITDs). The spherical model [6] is modified such that the sum of left and right HRTF power spectra is flat, such that timbre does not or only minimally change as a function of HRTF direction. The SHM is heuristically tuned to match ILDs and ITDs of measured HRTF data. If only this part would be used as a HRTF model, timbre changes would be minimal, but different directions on the cone of confusion could not be differentiated.
- **Rear-Front Monaural Cues:** Frontal horizontal directions use only the SHM to avoid coloration. That also means that for frontal directions, the headphones shall define the timbre. Rear directions are equalized with the following reasoning. Consider for example a horizontal frontal direction 30°. The rear direction with the same SHM ILDs and ITDs is $180^\circ - 30^\circ = 150^\circ$. The equalizer shall capture the spectral difference between 150° and 30°, as these aid front versus rear perception. The equalizers are determined based on averaged HRTF data.
- **Elevation Monaural Cues:** The elevation equalizers capture the spectral difference between a direction and its corresponding horizontal direction (direction vector projected to the horizontal plane). The model uses decoupled rear-front and elevation cues, that is, elevation cues are determined independently of azimuth.

Three configurations of the model are presented, for three different applications:

- Straightforward use of the model instead of HRTFs.
- Use of the model to convert 2-omnidirectional-microphone impulse responses (IRs) to HRTFs or binaural-room-impulse-responses (BRIRs) and enable head-tracking.
- Use of the model to enable head-tracking for static HRTFs or BRIRs.

The paper is organized as follows. Section 2 describes the proposed HRTF model. The three configurations of the model are described in Section 3. An application example, binauralization of a car with headtracking, is presented in Section 4 and conclusions are in Section 5.

2 HRTF Model

Consider the coordinate system in Figure 1. A listener's head is in the origin with a look direction in $x-$direction. Left and right ears lie on the $y-$axis. A direction vector $\mathbf{v}$ is defined by its azimuth α and elevation β angles. This direction vector defines the direction-of-arrival (DOA) of sound at the listener.

Used later, we are defining two additional angles. The angle between $\mathbf{v}$ and the $xz-$plane is

$$\phi(\alpha,\beta) = \frac{\pi}{2} - \arccos(\sin\alpha\cos\beta) \qquad (1)$$

and the angle between $\mathbf{v}$ and the $x-$axis is

$$\psi(\alpha,\beta) = \arccos(\cos\alpha\cos\beta)\,. \qquad (2)$$

To determine the model and its parameters the IRCAM HRTF Listen Database [8] was used. It contains HRTF sets of 51 different persons.

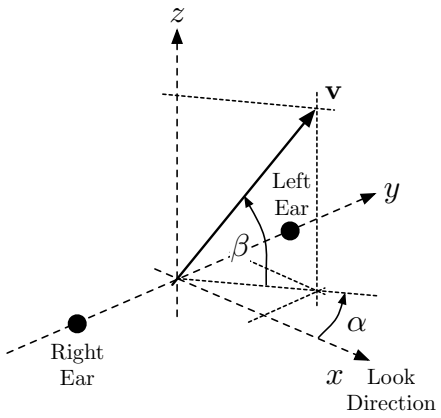


Fig. 1: Coordinate system, look-direction, and ear positions.

2.1 ILD and ITD Model

The SHM is symmetric around the $y-$axis connecting both ears. Thus ϕ (1) alone determines the transfer functions of arriving sound to ears. Consider the setup shown in Figure 2, where D is the diameter of the considered rigid sphere (head model).

The level transfer function between the sound pressure on a rigid sphere and free field plane wave sound pressure can be approximated by a one-pole-one-zero transfer function [6]. For the left ear, this transfer function in $s-$domain is

$$H(s,\phi) = \frac{A(\phi)s+B}{s+B}, \tag{3}$$

where $A(\phi) = 1+\sin\phi$ and $B = 2\frac{c}{D}$ with c being the speed of sound in air. The corresponding right ear transfer function is $H(s,-\phi)$.

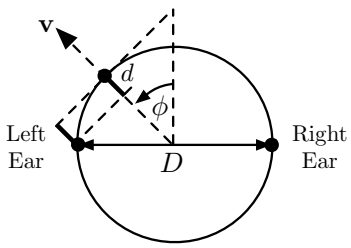


Fig. 2: Spherical head model (SHM).

Equation (3), modified such that the sum of left and right HRTF power spectra is flat, is

$$H(s,\phi) = \frac{\frac{A(\phi)}{\sqrt{A^2(\phi)+A^2(-\phi)}}s+B}{s+B}. \tag{4}$$

The time domain functions corresponding to $H(s,\phi)$ and $H(s,-\phi)$ are denoted $h(\phi)$ and $h(-\phi)$, respectively. Conversion of such filters to time-discrete digital filters is described in Section 5 of [7].

When sound reaches the sphere from a specific direction defined by $\mathbf{v}$, it has to travel by distance d farther to reach the left ear, illustrated in Figure 2. The delay corresponding to this additional travel distance d is

$$\tau(\phi) = \frac{D}{2c}(1-\sin\phi). \tag{5}$$

For the right ear, the delay is $\tau(-\phi)$.

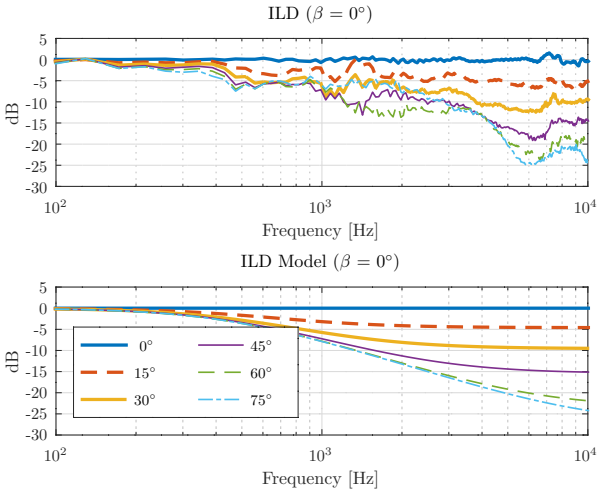


Fig. 3: ILD averaged over 51 HTRFs (top) and corresponding model (bottom).

Considering the above, ILDs and ITDs of left and right HRTFs with a specific azimuth α and elevation β are modeled as

$$\begin{aligned}\tilde{h}_l(t) &= h(\phi)(t-\tau(\phi))\\ \tilde{h}_r(t) &= h(-\phi)(t-\tau(-\phi)).\end{aligned}$$

The described model was compared to the HRTF data averaged over the 51 HRTFs. For ILDs and ITDs, head diameter D in the model was modified to match the data better (D of a human head is usually about 140 mm):

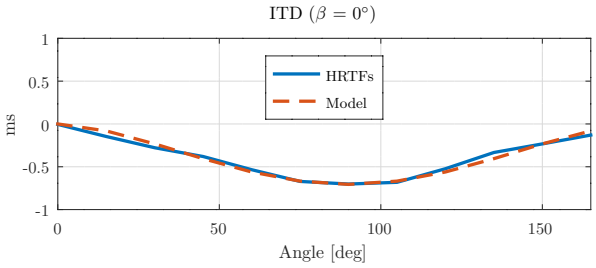


Fig. 4: ITDs averaged over 51 HTRFs and corresponding model.

- For ILDs: $D = 175$ mm.
- For ITDs: $D = 238$ mm. The ear-delays (5) have been computed assuming free field. Due to the rigid sphere (head) between the ears, it is expected that a larger D is needed for modeling. Additionally $\sin\phi$ in (5) is replaced with $|\sin\phi|^{1.5}\text{sign}(\sin\phi)$ to match the measurement data better.

Figures 3 and 4 show the averaged HRTF data and the model data for ILDs and ITDs, respectively.

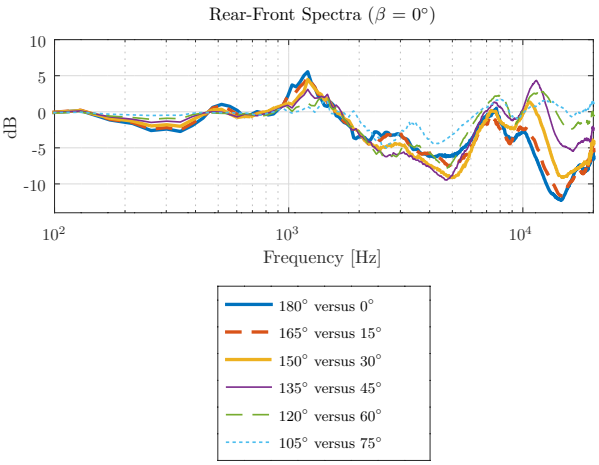


Fig. 5: HRTF spectra of rear directions $180° - \alpha$ normalized with corresponding spectra of front directions α, horizontal plane, averaged over 51 HRTFs.

2.2 Spectral Cues: Rear-Front

In order that timbre of sound follows the signature of the headphones, no spectral cues are applied in frontal directions $-90° \leq \alpha \leq 90°$. The reason for adding spectral cues at all is to improve front-rear discrimination. Figure 5 shows HRTF spectra of rear directions $180° - \alpha$ normalized with the corresponding front directions α, in the horizontal plane ($\beta = 0$).

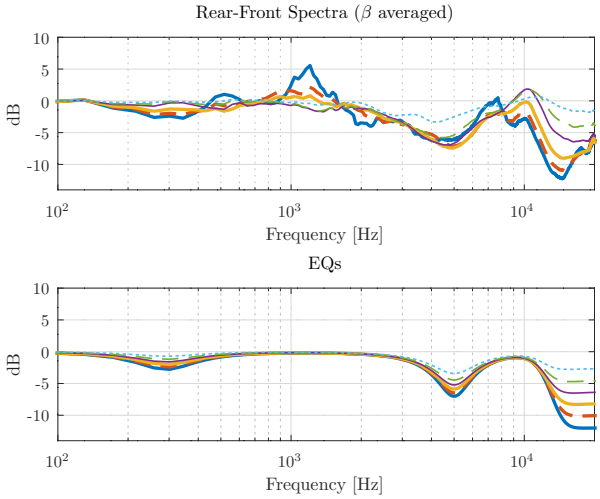


Fig. 6: Top: HRTF spectra of rear directions $180° - \phi$ normalized with corresponding front directions ϕ, averaged over all α and β resulting in ϕ (1) and 51 HRTFs. Bottom: biquad filters modeling he rear-front spectral differences.

Given a DOA $(\alpha,\ \beta)$, its angle to the forward axis x is ψ (2). ψ indicates how much a DOA is in front or back.

Given a specific ψ for a rear direction $(180° - \alpha,\ \beta)$, $(\alpha,\ \beta)$ is the corresponding frontal direction, for which rear-front cues are of interest.

The top panel of Figure 6 shows similar data as shown in Figure 5, but as opposed to considering α and the horizontal plane, ψ is considered and the data is averaged over all $(\alpha,\ \beta)$ corresponding to ψ. The averaged cues are reasonably similar (a bit weaker) to those only in the horizontal plane (Figure 5). Therefore, we are considering rear-front spectral cues only in one dimension ψ.

The bottom panel of Figure 6 shows the modeled rear-front spectra using biquad filters as specified in Table 1.

Table 1: Rear-front spectral cues biquad filters.

Peaking	290 Hz	−2.8 dB	Q= 1.6
Peaking	5000 Hz	−7 dB	Q= 3.0
High-Shelving	14000 Hz	−12 dB	
High-Shelving	14000 Hz	−12 dB	

The dB gain values in the table represent the strongest cues, rear ($\alpha = 180°$, $\beta = 0$) versus front ($\alpha = 0$, $\beta = 0$). The equalizers for the other rear-front cues are the same, with their dB gains decreased. The rear-front cues filter is denoted $e_{rf}(\psi)$.

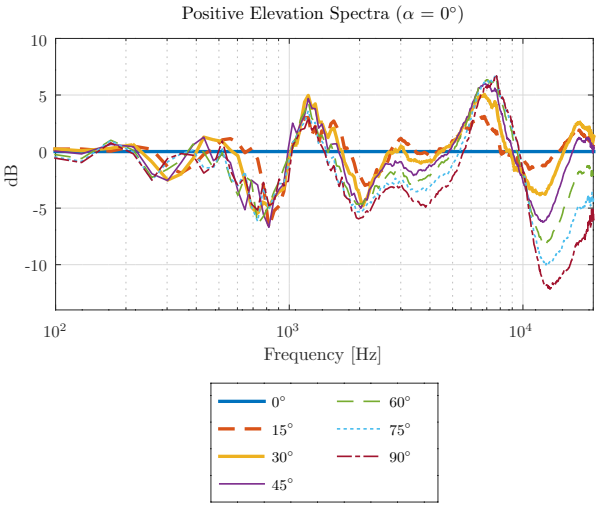


Fig. 7: HRTF spectra relative to elevation zero, azimuth zero, averaged over 51 HRTFs.

2.3 Spectral Cues: Elevation

Figures 7 and 8 show HRTF spectra of positive and negative elevations β for azimuth $\alpha = 0$, respectively, normalized with forward look-direction x.

The top panels of Figures 9 and 10 show HRTF spectra of positive and negative elevations normalized with corresponding horizontal directions, averaged over all azimuths. Compared to only frontal elevation cues (Figure 7), the azimuth-averaged cues are weaker but show clear correlation. Thus our model is using elevation cues only as a function of elevation, independently of azimuth.

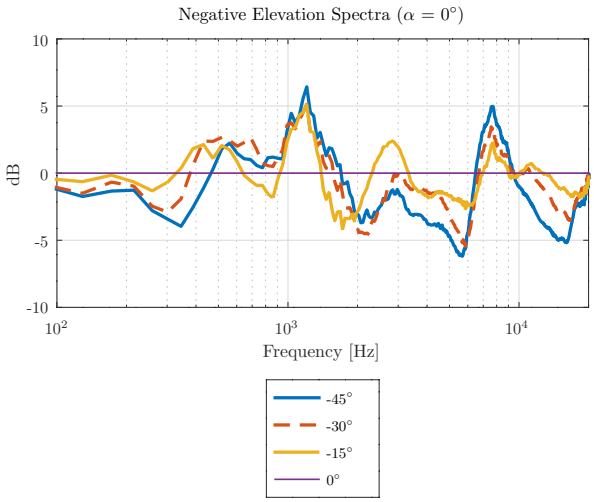


Fig. 8: HRTF spectra relative to elevation zero, azimuth zero, averaged over 51 HRTFs.

Table 2: Positive elevation spectral cues biquad filters.

Peaking	720 Hz	−2.8 dB	Q= 2.5
Peaking	1250 Hz	−7 dB	Q= 2.5
Peaking	8400 Hz	−2.8 dB	Q= 3
Peaking	13000 Hz	−7 dB	Q= 2.5
High-Shelving	600 Hz	−12 dB	

Table 2 specifies the biquad filter parameters used for 90° elevation. For smaller elevations the biquad gains magnitudes are decreased. The resulting filters for different elevations are shown in the bottom panel of Figure 9.

Table 2 specifies the biquad filter parameters used for −45° elevation. For negative elevations with smaller magnitudes the biquad gains magnitudes are decreased. The resulting filters for different elevations are shown in the bottom panel of Figure 10.

The elevation cues filter is denoted $e_e(\beta)$. Depending on β this may be the filter of Table 2 ($\beta > 0$) or 3 ($\beta < 0$).

3 Model Configurations

3.1 Binaural Rendering and Head-Tracking

Figure 11 shows the different processing steps of the model and the direction-related input data used for each

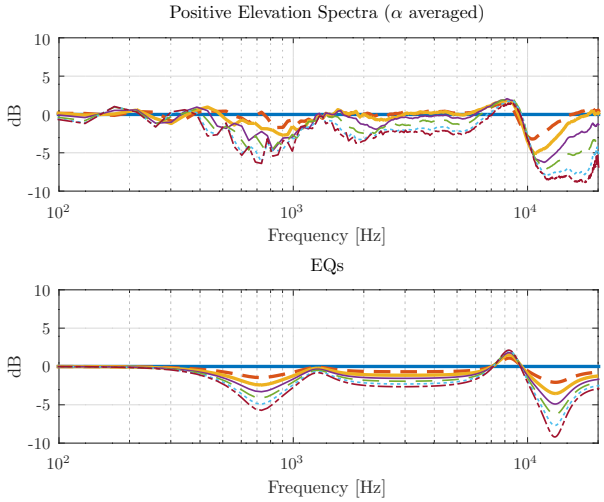


Fig. 9: Top: HRTF spectra related to elevation zero, averaged over all azimuths and 51 HRTFs. Bottom: biquad filters modeling positive elevation spectral differences. Legend: see Figure 7.

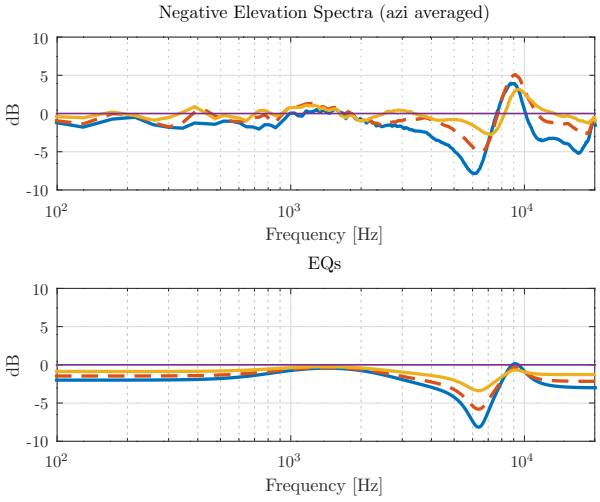


Fig. 10: Top: HRTF spectra related to elevation zero, averaged over all azimuths and 51 HRTFs. Bottom: biquad filters modeling negative elevation spectral differences. Legend: see Figure 8.

Table 3: Negative elevation spectral cues biquad filters.

Peaking	6400 Hz	−6 dB	Q= 4
Peaking	9000 Hz	4 dB	Q= 3
Low-Shelving	800 Hz	−2 dB	
High-Shelving	2000 Hz	−3 dB	

block. The HRTF model can be used for conventional binaural audio applications. Particularly head-tracking is easy, as one can specify the desired directions precisely. When head-tracking is used, the delays and IIR-filters have to be programmed such that no artifacts occur when they are updated in real-time.

3.2 Binauralization of 2-Mic-Array IRs with Head-Tracking

Impulse responses (IRs) to two omnidirectional-microphones are measured for a few directions, e.g. a surround sound setup, as shown in Figure 12. In the following it is explained how to binauralize this setup, including head-tracking.

These IRs feature ITD cues. If the ITD model D differs from the microphone distance D_m, the microphone ITDs are corrected to match the model ITDs. The microphone IRs do not feature ILD and spectral cues, thus HRTF model is used to apply these.

Delay between both microphone IRs is

$$\tau_m(\phi) = \frac{D_m}{2c}\left(1 - \sin\phi\right). \tag{6}$$

Thus, when the left microphone IR is delayed by

$$\Delta\tau(\phi) = \tau(\phi) - \tau_m(\phi) + \tau_0 \tag{7}$$

and the right microphone IR by $\Delta\tau(-\phi)$ then their ITDs are similar to the ITDs of the HRTF model. The offset τ_0 is chosen such that $\Delta\tau(\phi)$ is never negative. The microphone distance D_m should not deviate too much from the D for ITDs, such that acoustic reflections and diffuse sound are rendered faithfully.

The IRs from the two microphones are converted to binaural audio with a modified model, shown in Figure 13.

Head-tracking Consider a DOA with head-tracking $(\alpha,\ \beta)$ and a corresponding DOA without head-tracking $(\alpha_0,\ \beta_0)$. In the following ϕ denotes $\phi(\alpha,\ \beta)$ and $\phi_0 = \phi(\alpha_0,\ \beta_0)$, respectively.

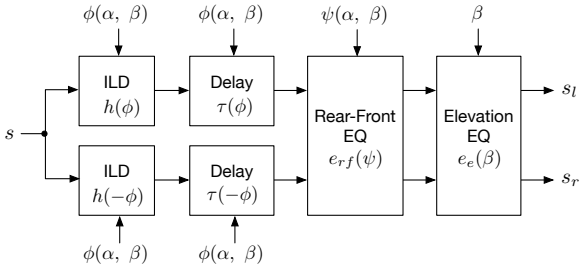


Fig. 11: Processing steps of the HRTF model and directional information used by each.

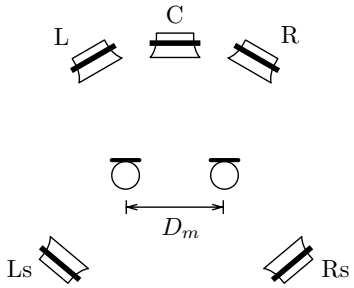


Fig. 12: Example setup, where IRs of a surround sound system to 2-omnidirectional-microphones are measured.

The desired delay for left ear is $\tau(\phi)$ (5) and its actual delay is $\tau_m(\phi_0)$ (6). The left and right correction delays converting the array ITDs to the head-tracked HRTF model ITDs are

$$\begin{aligned}\Delta\tau_l(\phi) &= \tau(\phi) - \tau_m(\phi_0) + \tau_1 \\ \Delta\tau_r(\phi) &= \tau(-\phi) - \tau_m(-\phi_0) + \tau_1\,, \end{aligned} \tag{8}$$

where τ_1 is again an offset such that the delays are never negative.

3.3 Head-Tracking with Measured HRTF or BRIR

Consider a similar scenario as previously, measuring a sound system, such as shown in Figure 12, but not using a microphone array, but with a person with in-ear microphones or a dummy-head. These measured IRs (HRTFs or BRIRs) contain all needed cues and can directly be used for binaural audio.

Our model can be applied in a way to enable head-tracking. Consider DOA of a recorded HRTF (or BRIR) $(\alpha_0,\ \beta_0)$. The corresponding (1) and (2) are denoted ϕ_0 and ψ_0. For the DOA with head-tracking, $(\alpha,\ \beta)$, (1) and (2) are denoted ϕ and ψ.

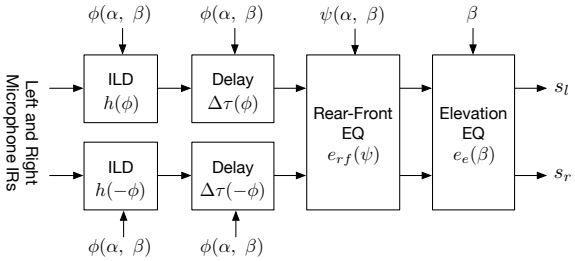


Fig. 13: HRTF model applied to left and right omnidirectional-microphone IRs.

ITD of the HRTF is converted to the ITD with head-tracking with (8).

Similarly, ILD correction filters are computed using the model and then applied to the measured HRTFs. That is, we are looking for a filter which converts $H(s,\phi_0)$ (4) to $H(s,\phi)$. The filter for converting the left HRTF ILD is

$$\Delta H(s,\phi,\phi_0) = \frac{H(s,\phi)}{H(s,\phi_0)}\,, \tag{9}$$

and the filter for right is $\Delta H(s,-\phi,-\phi_0)$. The corresponding time-domain filters are denoted $\Delta h_l(\phi)$ and $\Delta h_r(\phi)$.

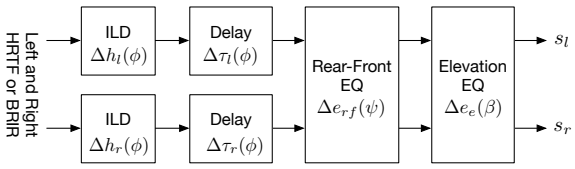


Fig. 14: HRTF model applied to left and right HRTFs or BRIRs.

For the rear-front cues, the model is used to remove the cues for ψ_0 and add cues for ψ,

$$\Delta e_{rf}(\psi) = e_{rf}^{-1}(\psi_0) e_{rf}(\psi)\,, \tag{10}$$

where e_{bf}^{-1} is as specified in Table 1 but the signs of the gains in dB are flipped.

Given elevation with and without head-tracking, β and β_0, respectively, the filter capturing the spectral cues difference of both is

$$\Delta e_e(\beta) = e_e^{-1}(\beta_0) e_e(\beta)\,. \tag{11}$$

4 Auralization of a Car

We used the model to auralize a car with head-tracking, based on microphone-array measurements and dummy-head measurements.



Fig. 15: Dummy-head with in-ear microphones (left) and omnidirectional-microphone array (right).

Each loudspeaker channel of a series-production car was measured with the linear tuning in the headunit activated. Figure 15 show dummy-head and microphone arrays placed within a car. Even though the microphone array featured 8 microphones, only two microphones were used. Left-right distance between a L-R microphone pair in the array is $D_m = 200$ mm.

The dummy-head and microphone-array measurements were used for auralization with head-tracking, as described in Sections 3.2 and 3.3, respectively. USB head-tracker [9] attached to Sennheiser HD 650 headphones and a soundcard connected to a PC laptop are used to run the auralization and HRTF model.

For the microphone-array IRs, low and high shelving filters are used to "calibrate the headphones" ad hoc by expert-listening[1]. Informal testing with several experienced listeners, sitting with the headphones in the car, gave promising results. All listeners found the experience faithful and externalized. We are planning more rigorous comparison of real-car, dummy-head based model, and microphone array based model in the near future.

5 Conclusions

An HRTF model was presented based on high shelving filters and delays for ILD and ITD modeling, respectively, and parametric equalizers for monaural cues. The parameters were determined by averaging HRTF sets, with the aim of modeling good none-individualized HRTFs.

As the model separates ILD, ITD, and monaural cues, it can be also applied to IRs of two omni-microphone arrays, by correcting ITD (which may differ depending on microphone distance), and fully applying ILD and the monaural cues.

Another usage of the model is to convert measured HRTFs (or BRIRs) from their given direction to a different direction. This is useful for head-tracking when only a few static HRTFs are given.

As a first real-world test we auralized a car based on dummy-head and microphone-array measurements on its seats. First informal results are promising.

[1] One could view this as imposing a free-field calibration on the headphones.

References

[1] Bauer, B. B., "Stereophonic earphones and binaural loudspeakers," *J. Audio Eng. Soc.*, 9, pp. 148–151, 1961.

[2] Huopaniemi, J., *Virtual Acoustics and 3D Sound in Multimedia Signal Processing*, Ph.D. thesis, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Finland, 1999, rep. 53.

[3] Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D., "Binaural technique: Do we need individual recordings?" *J. Audio Eng. Soc.*, 44(6), pp. 451–469, 1996.

[4] Wightman, F. L. and Kistler, D. J., "Individual differences in human sound localization behavior," *J. Acoust. Soc. Am.*, 99, pp. 2470–2500, 1996.

[5] Blauert, J., *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press, Cambridge, Massachusetts, USA, revised edition, 1997.

[6] Duda, R. O., "Modeling head related transfer functions," in *Proc. Twenty-Seventh Annual Asilomar Conference on Signals, Systems and Computers (Asilomar, CA)*, 1993.

[7] Faller, C. and Menzer, F., "Binaural Audio with Relative and Pseudo Head Tracking," in *Preprint 138th Conv. Aud. Eng. Soc.*, 2015.

[8] Warusfel, O., "Listen HRTF Database," `http://recherche.ircam.fr/equipes/salles/listen/`, 2002, accessed on January 30, 2024.

[9] Supper, B., "USB Head-Tracker," `https://supperware.co.uk/headtracker-overview`, 2024, accessed on March 8, 2024.