EUROREGIO BNAM2022
Joint Acoustics Conference
9th-11th of May, Aalborg, Denmark

Data-driven Optimization of Parametric Filters for Simulating Head-Related Transfer Functions in real-time Rendering Systems

Fenja Schwark[1,*], **Marc R. Schädler**[1], **Giso Grimm**[1,2]

[1]Medizinische Physik, Universität Oldenburg, and Cluster of Excellence "Hearing4all", Oldenburg, Germany.

[2]Hörzentrum gGmbH, Oldenburg, Germany.

[*]fenja.schwark@uni-oldenburg.de

Abstract

In binaural virtual acoustics, the head-related directional properties of the receiver, i.e., the listener, are often modeled by convolution of the signal with measured head-related impulse responses (HRIRs). However, the computational cost of HRIR-convolution is rather high, especially when targeting interactive low-delay reproduction. Furthermore, depending on the spatial resolution of the HRIR data set, interpolation needs to be applied to simulate all source directions. In order to reduce the computational cost in low-delay real-time virtual acoustic rendering, this study uses a parametric digital filter model with delay lines to approximate the direction-dependent features of the head. A data-driven optimization method for the direction-dependent design function parameters is introduced that aims at matching the direction-dependent features of modeled and measured HRIRs using a spectral distance metric. Using an objective binaural speech intelligibility model, it was shown that the estimate of the speech intelligibility for the optimized model approaches the estimate for the measured HRIRs. This suggests that the parametric HRIR model may be sufficient to enable plausible spatial perception in virtual acoustic environments. With the parametric HRIR model a reduction of computational cost in the order of two magnitudes is possible for virtual acoustic environments with a small number of objects. Further work will include subjective testing of the model, compared against measured HRIRs.

Keywords: virtual acoustic environments, binaural rendering, parameterization.

1 Introduction

Virtual acoustics is used in a wide range of applications, including hearing aid research and audiology. Compared to conventional setups with only a few sound sources, the use of virtual acoustic environments (VAEs) allows more complex acoustic configurations to be tested while remaining in a predefined and reproducible environment. Thus, virtual acoustic environments form a kind of intermediate level between simple laboratory conditions and real-life environment [1].

Binaural reproduction of VAEs is most commonly realized by convolution with measured head-related impulse responses (HRIRs) [2, 3, 4, 5]. Virtual acoustic scene rendering with measured HRIRs results in an accurate reconstruction of the direction- and frequency-dependent filter effect of our head on impinging sound. At the same time, the limited number of discrete incidence directions for which measurements are available leads to spatial limitations or – when interpolation is applied – to undesired spectral artifacts. Furthermore, convolution-based systems are quite inefficient when it comes to low-delay real-time applications that require short block sizes.

Parametric models can resolve these limitations by resembling the main features of the head-related transfer functions (HRTFs) with low-order digital filters with a continuous filter design. A simple and structured modeling approach aimed at modeling physiologically motivated features such as shadow effects of head, pinna

and torso or interference effects arising from reflections within the pinna structure is structural decomposition [6, 7, 8]. A structured implementation of a small number of acoustic filters related to the geometric characteristics of the body and the physical properties of sound propagation makes such models quite understandable. In addition, model-based reproduction of VAEs is more efficient and less time-consuming than reproduction using convolution-based rendering methods for real-time applications.

An analysis of the directional magnitude characteristics revealed that such parametric models insufficiently resemble the characteristics of measured HRTFs. Moreover, the spatial fine structure of binaural intelligibility level differences cannot be reproduced correctly.

To overcome the limitations of computational performance and spatial fine structure in speech recognition threshold (SRT) predictions, we propose a framework for data-driven optimization of parametric models. The design parameters of the filters are optimized to achieve better congruence with the directional characteristics of measured HRTFs. The performance of this data-driven optimization framework is analyzed using an objective binaural speech intelligibility model [9]. Furthermore, computational performance is analyzed for a set of typical simulation parameters.

2 Methods

2.1. Parametric HRTF Models

In this study, a parametric HRTF model based on the spherical head model (SHM) introduced by Brown and Duda in 1998 [6] with additional filters modeling shadow effects and reflections [10] was used. This model consists of three stages – a *spherical head model*, a stage to reduce *front-back confusions*, and filters to reduce *up-down confusions*. The filter design functions for each stage are rotation-symmetric around a filter axis, and depend on an opening angle θ between this axis and the incident direction, see Figure 1 for an overview.

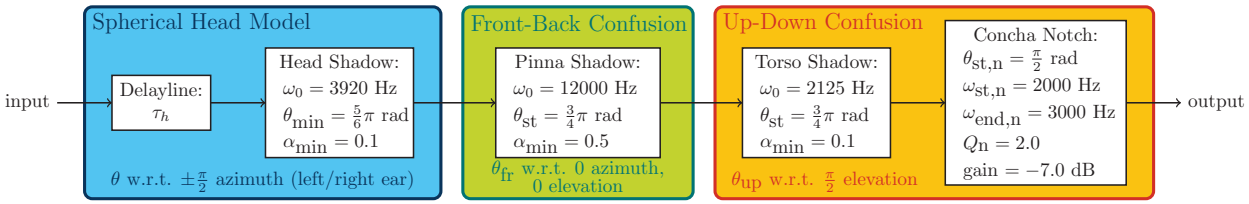


Figure 1: Structure of the spherical head model (blue box) and the extensions to reduce front-back confusions (green box) and up-down confusions (orange box), with an overview of the chosen design parameters.

Spherical Head Model

The SHM provides a simple model to describe the diffraction behavior of sound waves around the head by modeling the propagation of sound around a rigid sphere. The interaural time difference τ_h is calculated according to the frequency-independent formula of Woodworth and Schlosberg [11], which gives the time delay of a wave traveling to the observation point compared to a wave traveling to the center of the sphere in the free field:

$$\tau_h(\theta) = \begin{cases} -\frac{a}{c}\cos(\theta) & \text{if } 0 \le |\theta| < \frac{\pi}{2} \\ \frac{a}{c}(|\theta| - \frac{\pi}{2}) & \text{if } \frac{\pi}{2} \le |\theta| < \pi \end{cases} \tag{1}$$

Here θ is the angle of incidence relative to the axis of the ears at azimuth angles of $-\frac{\pi}{2}$ (left ear) or $\frac{\pi}{2}$ (right ear).

The shadow effect of the rigid sphere is approximated by the following stable single-pole, single-zero filter:

$$H_{\text{hs}}(z) = \frac{(\omega_0 + \alpha f_s) + (\omega_0 - \alpha f_s)z^{-1}}{(\omega_0 + f_s) + (\omega_0 - f_s)z^{-1}} \tag{2}$$

The frequency ω_0 is related to the head radius a and the speed of sound c by $\omega_0 = c\,a^{-1}$. The filter coefficient α is calculated as a function of the incidence angle θ according to the design function $\alpha(\theta)$:

$$\alpha(\theta) = \left(1 + \frac{\alpha_{\text{min}}}{2}\right) + \left(1 - \frac{\alpha_{\text{min}}}{2}\right) \cos\left(\frac{\theta}{\theta_{\text{min}}} \cdot \pi\right) \tag{3}$$

The original model parameters [6] are $\alpha_{\text{min}} = 0.1$ and $\theta_{\text{min}} = \frac{5}{6}\pi$, with a head radius of $a = 8.75$ cm. The directional magnitude characteristics of the SHM in the transverse and median plane is shown in Fig. 2a and 2e.

Extended Spherical Head Model

Ewert et al. [10] introduced extensions of the original SHM to account for pinna effects and torso shadow. First, a pre-warping method was added to better control the attenuation behavior for small azimuth angles (equal to equation (7) of [10] with $w_{exp} = 0.5$ and $w_{\text{a}} = 0.5$):

$$\theta_{\text{wrp}} = \frac{\theta}{3}\left(2 - \cos\left(\sqrt{\pi\theta}\right)\right) \tag{4}$$

θ was replaced by θ_{wrp} in Equation (1) and Equation (3).

To reduce *front-back confusions* a filter that attenuates high-frequency components of sources behind the head was added. This filter has the same structure as the filter for the head shadow effect, Equation (2). However, by adopting the design function, the filter operates only for a restricted range of incidence directions $\theta > \theta_{\text{st}}$:

$$\alpha(\theta) = \left(\frac{1 + \alpha_{\text{min}}}{2}\right) + \left(\frac{1 - \alpha_{\text{min}}}{2}\right) \cos\left(\frac{\theta - \theta_{\text{st}}}{\pi - \theta_{\text{st}}} \cdot \pi\right) \tag{5}$$

This is equal to equation (8) in [10] with $d = 1$. For this filter, the angle θ_{fr} is defined with respect to the frontal direction [azimuth 0, elevation 0] and the design function $\alpha(\theta_{\text{fr}})$ is evaluated. The original parameter choice was $\alpha_{\text{min}} = 0.5$, $\theta_{\text{st}} = \frac{3}{4}\pi$ and $\omega_0 = 12$ kHz.

Analogously to the shadow effect of the pinna, the torso attenuates high-frequency components from signals impinging from the lower hemisphere, resulting in less *up-down confusion*. The torso shadow filter has the same structure and design function as the pinna effect filter, with a filter axis pointing upwards, i.e., the incident angle θ_{up} is defined relative to the upward direction and the design function $\alpha(\theta_{\text{up}})$ is evaluated. The filter is applied for incidence angles $\theta_{\text{up}} > \theta_{\text{st}}$ with an original value of $\theta_{\text{st}} = \frac{3}{4}\pi$. The other initial parameters of this filter were $\omega_0 = 2125$ Hz and $\alpha_{\text{min}} = 0.1$.

Another important cues for elevation perception are created by reflections in the pinna structure. The effect of such reflections can be modeled by a parametric filter that produces a notch whose gain and center frequency change linearly as a function of elevation angle. This notch was realized by a parametric equalizer. For the original filter, the applied gain in the transverse plane ($\theta_{\text{st,n}} = \frac{\pi}{2}$) was zero and a maximum attenuation of 7 dB was applied for the sounds from above. The notch frequency changed from 2 kHz in the transversal plane to 3 kHz at $\theta_{\text{up}} = 0$. The Q-factor of the notch was set to 2.

2.2. Optimization of the Parametric Model by Measured HRTFs

For the data-driven optimization of the parametric model, the OlHeaD-HRTF Database [12] was used. From this database, the ear drum microphones of the KEMAR head-and-torso simulator were selected as a reference. For a uniform weighting of all input directions, HRIRs measured for a directionally balanced subset of 47 incidence directions were extracted from the provided measurements. The measured HRIRs were Fourier-transformed and smoothed with a function that mimics auditory filters. The magnitude of the smoothed HRTFs was extracted at center frequencies spaced by 0.5 ERB in the range between 250 Hz and 16 kHz.

The parametric filters of the extended SHM do not account for the direction-independent components of the HRTFs, e.g. the ear canal. In order to model these direction-independent filter characteristics of the ear canal,

an FIR filter for diffuse field equalization was applied. For this purpose the measured diffuse field response of the KEMAR dummy head – which is included in the OlHeaD-HRTF Database [12] – was used.

The cost function to be minimized in the optimization process was the mean square difference between the logarithmic magnitude of the sampled and simulated HRTFs, for all frequencies and directions.

For the optimization, the Nelder-Mead simplex method [13] implemented in the predefined GNU Octave [14] function 'fminsearch' was used. Simplex methods provide a numerical solution to find the minimum of multidimensional functions. While the convergence speed of these methods is rather slow, the implementation is relatively easy and the system is robust. For this study, the function 'fminsearch' was modified to only allow values in parameter-dependent predefined intervals that make sense in the context of our model, e.g., an angle should fall in the interval of 0 and π. After each operation of the Nelder-Mead Method, it was verified whether the parameters exceed the set boundaries. If any vertex element exceeded the defined boundaries, a mechanism that mirrors all values beyond the boundaries periodically back into the defined interval was executed.

The simplex was initialized with the original design parameters (see Figure 1 for an overview). As the features are generated by the entirety of all filters, it is sensible to optimize all filter parameters in parallel. However, to achieve faster convergence, the parameters of each filter were pre-optimized separately before parallel optimization of all parameters begins, taking into account the interactions of the filters. First, the parameters of the head shadow model were optimized, then the Pinna model was optimized with these parameters, and finally the notch filter was optimized, keeping all other parameters fixed.

Since the OlHeaD-Database does not provide enough measurement data in the lower hemisphere, the filter parameters of the high-shelf filter that describes the shadow effect of the torso were not included in the optimization.

2.3. Speech Intelligibility Prediction

Speech intelligibility measures are important for the description of acoustic scenes as well as signal processing methods related to acoustic communication. A parametric HRTF model which replaces measured HRTFs for rendering VAEs, should result in comparable speech intelligibility like convolution-based methods. The simulation framework for auditory discrimination experiments (FADE) [9] is a framework to perform different kinds of acoustical discrimination experiments. FADE has been designed to work as an automatic speech recognizer specialized to the simple grammar and limited speech material of the Matrix-Tests [15]. The automatic speech recognizer is trained and tested with noisy signals at different signal-to-noise ratios (SNRs) and that SNR at which a threshold of speech intelligibility, i.e. SRT, of 50% is reached is determined.

FADE was used to predict speech intelligibility of speech signals from the German Matrix Test presented in the *icra5-250* masking noise [16]. As the long term spectra of target and noise signal equal each other this noise masks the speech signal well. However, the pauses in the noise signal enable for "listening in the dips". Simulations were performed with FADE (version 2.1.1) using contralateral inhibition (FADE-kain).

FADE performs speech intelligibility predictions on the sound signals at the ear drum. In the convolution-based rendering technique, the HRIRs measured at the ear drum can be used to directly obtain the desired signals. For the parametric model, the direction-independent diffuse field response provided by the OlHeaD-HRTF Database [12] was used to obtain the simulated ear-drum audio signals.

It was possible to reproduce empirical data such as the results of a study by Beutelmann and Brand [17], in which the SRT was measured for a speech signal impinging from the front and noise presented at different azimuth angles – the so-called S_0N_φ condition – by conducting speech intelligibility prediction using FADE. This particular condition was therefore used to compare the speech intelligibility obtained with the parametric models and the measured HRTFs.

2.4. Computational performance

For a comparison of computational performance, the time required to simulate a virtual acoustic environment of 60 seconds duration was measured. The simulation was performed using the Toolbox for Acoustic Scene Creation and Rendering (TASCAR) [18, 19], version 0.223.1. This toolbox, including an implementation of the parametric HRTF model with optimized parameters, is written in C++ and optimized for real-time interactive processing with low delay.

Six different rendering methods were compared: A direction-independent method called 'omni' served as a baseline indicating the overhead of the simulation toolbox, e.g. due to the position calculation. The second method tested was the parametric HRTF simulation with the proposed optimization ('hrtf'). The non-parametric spatial rendering was implemented as 2-dimensional or 3-dimensional VBAP [20]. The 2-dimensional VBAP was rendered on eight virtual loudspeakers arranged on a circle. The 3-dimensional VBAP was rendered on 47 speakers arranged on a sphere, as used for optimization. Both methods were tested without and with convolution using the corresponding impulse responses from the HRIR database [12]. The labels for the nonparametric methods were 'vbap2d', 'vbap2dconv', 'vbap3d', and 'vbap3dconv'.

To estimate the factors influencing the required computing time, the block size P (64, 128, 256, 512, 1024) and the number of virtual sound sources N (1, 10, 100, 1000) were varied. The sampling rate was set to 48 kHz. The air absorption model as well as the time-varying delay line were disabled to isolate the effect of the rendering method as much as possible.

3 Results

Optimized Model

Starting from the original parameters of the extended SHM model as given in Table 1, the data-driven optimization was performed and the final parameters were computed.

Table 1: Original and final filter parameters of the data-driven optimization.

(a) Head Shadow

Parameter	original	final
ω_0 / Hz	3920	3100
θ_{min} / rad	$\frac{5}{6}\pi$	$\frac{8}{9}\pi$
α_{min}	0.1	0.14

(b) Pinna Shadow

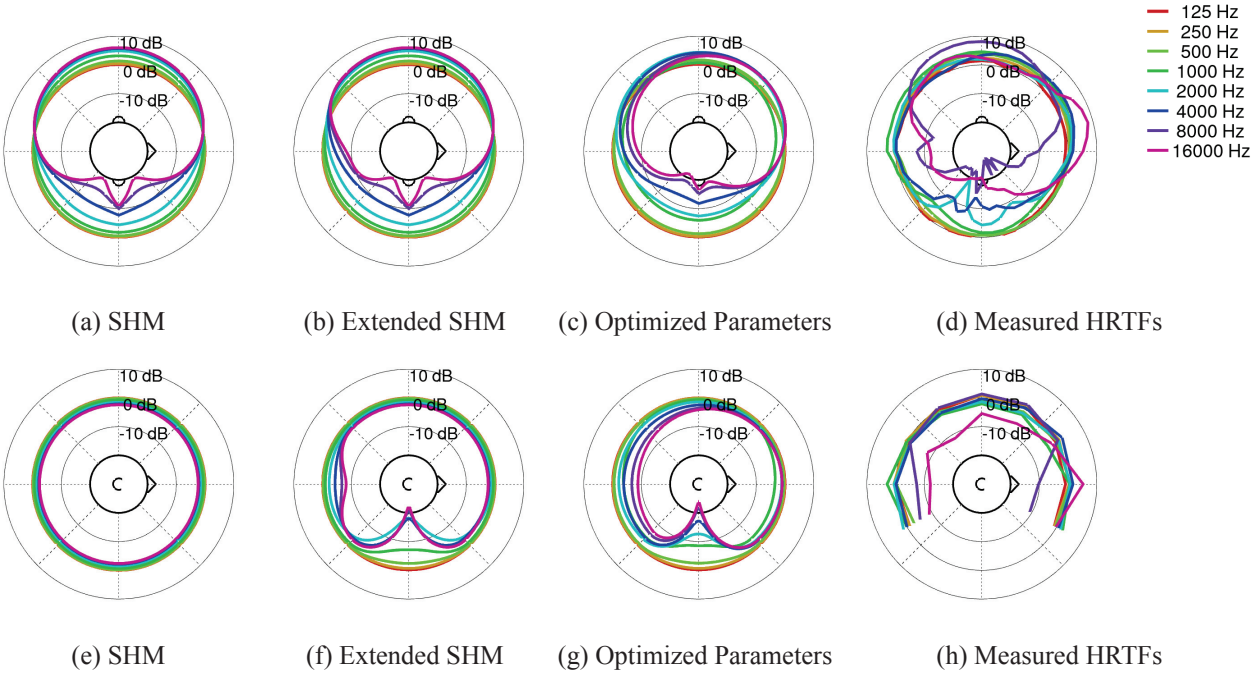
Parameter	original	final
ω_0 / kHz	12.0	11.2
θ_{st} / rad	$\frac{3}{4}\pi$	0
α_{min}	0.50	0.39

(c) Concha Notch

Parameter	original	final
$\theta_{\text{st,n}}$ / rad	$\frac{\pi}{2}$	$0.5\overline{6}\pi$
$\omega_{\text{st,n}}$ / Hz	2000	1300
$\omega_{\text{end,n}}$ / Hz	3000	650
Q_{n}	2.0	2.3
gain / dB	-7.0	-5.4

The directional magnitude characteristics of the model with the final parameters found by the data-driven optimization are shown in Fig. 2c and 2g.

The directional magnitude characteristics in the transverse plane obtained with the optimized parameters (Fig. 2c) compared to those obtained with the original parameters (Fig. 2b) show an attenuation behavior on the contralateral side that is in better agreemnet with the feature observed in the measured data (Fig. 2d). This difference is mainly related to the change in the parameters of the head shadow filter. Increasing the angle θ_{min} results in a less pronounced back lobe at high frequencies on the contralateral side. In addition, shifting the frequency ω_0 to a lower frequency leads to a shift of the transition region of the high-shelf filter to lower frequencies, i.e. attenuation of lower frequency components.



(a) SHM (b) Extended SHM (c) Optimized Parameters (d) Measured HRTFs

(e) SHM (f) Extended SHM (g) Optimized Parameters (h) Measured HRTFs

Figure 2: Directional magnitude characteristics of the left ear in the transversal plane (upper panel) and median plane (lower panel) for the SHM, the original extended SHM, the extended SHM with optimized parameters and measured HRTFs from the OlHeaD-HRTF Database [12] (diffuse field equalized measurements at the ear drum of the KEMAR head-and-torso simulator).

After optimization, the pinna shadow filter to limit *front-back confusion* operates in a completely different range than originally intended. This can be seen in the extreme change of the stating angle θ_{st}. The obtained filter leads to a general attenuation of high-frequency components in the posterior hemisphere, which is particularly visible in the characteristics in the median plane (see Fig. 2g). In addition, a general directionality of the high frequencies is observed in the transversal plane, which is slightly rotated compared to the low frequencies (Fig. 2c). These effects are in very good agreement with the characteristics of the measured HRIRs.

The features modeled by the parametric equalizer are barely visible in the directional magnitude characteristics. However, it can be seen that after the optimization, the notch is modeled in a lower frequency region than before. A detailed analysis of the transfer functions reveals that this notch obtained with the optimized parameters models a notch observed in the measured HRTFs.

An implementation of the proposed extension of the model including the optimized parameter set for the OlHeaD-Database is part of the Toolbox for Acoustic Scene Creation and Rendering (TASCAR) [18] and available as open source [19].

Speech intelligibility prediction

In addition to the directional magnitude characteristics as a physical measure, speech intelligibility is analyzed as a perceptual measure. In Fig. 3, the predicted SRT is plotted as a function of azimuth angle φ, equivalent to the direction of incidence of the noise signal. Since contralateral inhibition was used for the SRT prediction, the SRT function is (almost) axisymmetric with respect to $\varphi = 0$. It can be seen that the predicted SRTs of the measured HRTFs were better reproduced by the parametric model with the optimized parameters than by the model with the original parameters.

Differences in performance between the original and optimized parametric model are most visible in the region near full lateralization, i.e. $\pm\frac{\pi}{2}$. For these incidence directions, the SRTs predicted from the measured HRTFs

show two minima and a maximum. This behavior is not reproduced by the SRTs predicted from the model with the original parameter set, but the minimum is quite blurred. The deviation from the SRT predicted for the measured HRTFs is quite high, with up to 6 dB. In contrast, for the SRTs predicted for the model with optimized parameters, two minima and one maximum are clearly visible. The depth of the minima is less pronounced than for the SRTs obtained by the measured HRTFs. The maximum deviation from the predicted SRTs for the measured HRTFs is about 3.5 dB.

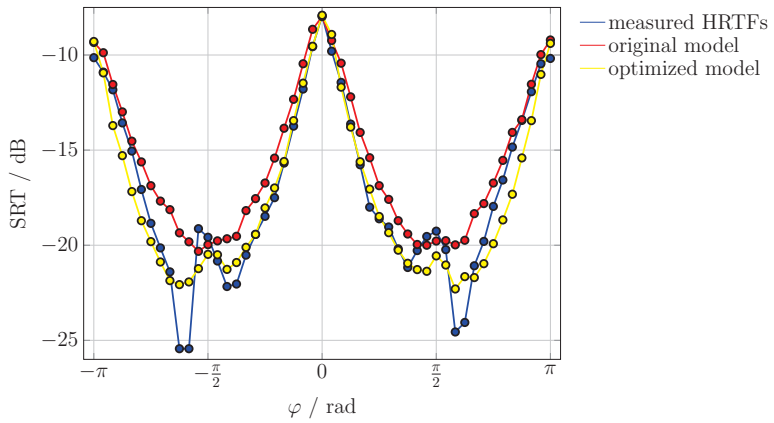
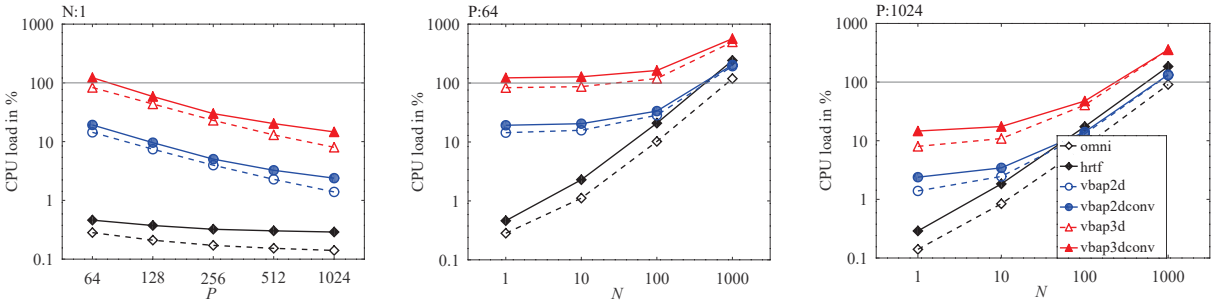


Figure 3: Predicted SRTs for the S_0N_φ condition for measured HRTFs as well as the parametric model with original and optimized parameters as a function of φ at $\frac{\pi}{24}$ resolution.

Computational Performance



(a) Influence of P (b) Influence of N, small P (c) Influence of N, large P

Figure 4: CPU load needed while rendering a virtual acoustic scene, for different rendering methods and parameters.

Figure 4a shows the CPU load, i.e., the time needed to compute an audio segment divided by the duration of that segment, as a function of block size P, for the simulation from one sound source. The CPU load of the parametric model is largely independent of the block size, while for the other rendering methods the CPU load is roughly inversely proportional to the block size. It can be seen that the parametric model has a performance advantage over 3-dimensional simulation via convolution of about a factor of 100 for small block sizes. This advantage decreases as the block size increases.

Figures 4b and 4c show the CPU load as a function of the number of sources N, each for one block size (4b for 64 samples, 4c for 1024). Here it can be seen that for the parametric model, the computational load increases linearly with the number of sources, largely independent of the block size. For the speaker-based methods, there is a lower bound independent of the number of sources.

Under all conditions tested, the parametric model was more efficient than 3D VBAP with convolution, which

corresponds to the training data set.

4 Discussion

Model Optimization

Through the proposed optimization, the directional characteristics of the parametric model approximate those of the measured HRIRs. However, the influence on the different filter stages is different: In the median plane, optimization leads to a better reproduction of the back lobe at high frequencies on the contralateral side. In addition, the directional characteristics of the high-frequency components are slightly rotated relative to those of the low-frequency components, which is consistent with the measured data. In the median plane, the general attenuation of the high-frequency components also leads to a higher agreement with the measured data.

Due to the limited number of available HRTF data points in the OlHeaD-HRTF Database [12] in the lower hemisphere, it is not possible to optimize the torso shadow modeled by the third high-shelf filter to measurement data. However, the relatively strong attenuation of high frequencies by this high-shelf filter seems appropriate because – compared to the pinna – the torso is a large object around which the sound is diffracted. This leads to an increased attenuation of high frequencies as well as to an attenuation of a wider frequency range compared to the high-shelf filter modeling the effect of the pinna. In addition, noise sources from below are rare under real conditions and therefore not very relevant.

Speech intelligibility prediction

Speech intelligibility prediction analysis shows that the proposed optimization reduces the difference between predicted SRT of the parametric model and predicted SRT with measured HRTFs. The most pronounced difference between the predicted SRTs is found for angles close to full lateralization. While the change of SRT in this region is negligible using the original filter parameters, the optimized model reproduces the two minima and the maximum in SRT as found with the measured HRTFs. However, these are more pronounced with the measured HRTFs. The prediction of SRTs shows that the agreement between parametric model and measured data is improved by the optimization, both in terms of directional characteristics and speech intelligibility.

Computational Performance

Computational performance analysis reveals that the parametric HRTF model outperforms the convolution with HRIRs for all 47 directions used in the optimization under all conditions tested. It is found that the CPU load of the parametric model is mostly independent of the block size P and increases linearly with the number of sources N. This can be explained by the implementation as a time-domain filter: the filter design function contains only few operations and is called only at the block boundaries. The computationally more intensive filter stages operate on every audio sample, thus the CPU load scales linearly with the number of audio samples, i.e., with the number of sources. In the case of the convolution, the situation is more complex: here, the audio samples are first panned according to the incidence direction. The result is a multi-channel audio, with one channel per direction. In a next step, each channel is convolved with the HRIR. If the number of sources is lower than one third of the number of channels, this implementation is less efficient than an explicit convolution for each source. However, the complexity depends on the interpolation method – here VBAP is chosen, which requires three channels for each source in the 3-dimensional case, but for Higher Order Ambisonics all channels are required even for a single source.

5 Conclusion

The physiologically motivated parametric HRTF model [10] reproduces the fundamental characteristics of binaural hearing. However, without adjustment of its parameterization, the directional characteristics of individual HRTF databases, e.g., the OlHeaD-HRTF Database [12], can be modeled only poorly. This can be demonstrated with directional SRT predictions: Prediction differences between convolution-based rendering methods

and parametric HRTF modeling indicate an effect at perceptual and even behavioral levels.

By applying the proposed data driven optimization method, the mean square deviation between the parametric HRTF model and the measured HRTF data is reduced by 1 dB from 4.1 to 3.1 dB, averaged across all directions. Likewise, the deviation of the predicted SRT based on the parametric model to the convolution-based rendering was reduced from a maximum of 6 dB to 3.5 dB without changing the structure of the model, thus maintaining the computational efficiency.

The computational performance of the parametric HRTF model is about two magnitudes higher than that of a convolution-based approach, as used for training, when the number of sources is small and the block sizes are short. As the number of sources and block size increases, this advantage decreases.

Acknowledgements

Stephan Ewert provided a Matlab implementation of an earlier version of the model described in [10]. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 352015383 – SFB 1330.

References

[1] Volker Hohmann, Richard Paluch, Melanie Krueger, Markus Meis, and Giso Grimm. The Virtual Reality Lab: Realization and Application of Virtual Sound Environments. *Ear & Hearing*, 41(Supplement 1):31S–38S, nov 2020. ISSN 0196-0202. doi: 10.1097/AUD.0000000000000945. URL https://journals.lww.com/10.1097/AUD.0000000000000945.

[2] Florian Denk, Birger Kollmeier, and Stephan D. Ewert. Removing reflections in semianechoic impulse responses by frequency-dependent truncation. *AES: Journal of the Audio Engineering Society*, 66(3): 146–153, 2018. ISSN 15494950. doi: 10.17743/jaes.2018.0002.

[3] Joachim Thiemann and Steven van de Par. A multiple model high-resolution head-related impulse response database for aided and unaided ears. *EURASIP Journal on Advances in Signal Processing*, 2019(1):9, dec 2019. ISSN 1687-6180. doi: 10.1186/s13634-019-0604-x. URL https://asp-eurasipjournals.springeropen.com/articles/10.1186/s13634-019-0604-x.

[4] Benjamin Bernschütz. A spherical far field hrir/hrtf compilation of the neumann ku 100. In *Proceedings of the 40th Italian (AIA) annual conference on acoustics and the 39th German annual conference on acoustics (DAGA) conference on acoustics*, page 29. AIA/DAGA Merano, 2013.

[5] Fabian Brinkmann, Alexander Lindau, and Stefan Weinzierl. A high resolution head-related transfer function database including different orientations of head above the torso. *Proceedings of the AIA- ...*, pages 596–599, 2013.

[6] C. Phillip Brown and Richard O. Duda. A structural model for binaural sound synthesis. *IEEE Transactions on Speech and Audio Processing*, 6(5):476–488, 1998. ISSN 10636676. doi: 10.1109/89.709673. URL http://ieeexplore.ieee.org/document/709673/.

[7] V Ralph Algazi, Richard O Duda, Reed P Morrison, and Dennis M Thompson. Structural composition and decomposition of hrtfs. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*, pages 103–106. IEEE, 2001.

[8] V Ralph Algazi, Richard O Duda, and Patrick Satarzadeh. Physical and filter pinna models based on anthropometry. In *Audio Engineering Society Convention 122*. Audio Engineering Society, 2007.

[9] Marc René Schädler, Anna Warzybok, Stephan D Ewert, and Birger Kollmeier. A simulation framework for auditory discrimination experiments: Revealing the importance of across-frequency processing in speech perception. *The journal of the acoustical society of America*, 139(5):2708–2722, 2016.

[10] Stephan D. Ewert, Oliver Buttler, and Hongmei Hu. Computationally Efficient Parametric Filter Approximations for Sound-Source Directivity and Head-Related Impulse Responses. In *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pages 1–6. IEEE, sep 2021. ISBN 978-1-6654-0998-8. doi: 10.1109/I3DA48870.2021.9610923. URL https://ieeexplore.ieee.org/document/9610923/.

[11] Jens Blauert. *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.

[12] Florian Denk, Stephan MA Ernst, Stephan D Ewert, and Birger Kollmeier. Adapting hearing devices to the individual ear acoustics: Database and target response correction functions for various device styles. *Trends in hearing*, 22:2331216518779313, 2018.

[13] John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7 (4):308–313, 1965.

[14] John W. Eaton et al. Gnu octave. https://www.gnu.org/software/octave/, 1998-2022.

[15] K Wagener, T Brand, and B Kollmeier. Development and evaluation of a german sentence test part iii: Evaluation of the oldenburg sentence test. *Zeitschrift Fur Audiologie*, 38:86–95, 1999.

[16] Kirsten Carola Wagener, Thomas Brand, and Birger Kollmeier. The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners. *International Journal of Audiology*, 45(1): 26–33, 2006.

[17] Rainer Beutelmann and Thomas Brand. Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 120(1):331–342, 2006.

[18] Giso Grimm, Joanna Luberadzka, and Volker Hohmann. A toolbox for rendering virtual acoustic environments in the context of audiology. *Acta acustica united with acustica*, 105(3):566–578, 2019.

[19] Giso Grimm, Tobias Herzke, Fenja Schwark, and Merle Gerken. Toolbox for Acoustic Scene Creation and Rendering (TASCAR). http://tascar.org/, https://github.com/gisogrimm/tascar, 2022.

[20] Ville Pulkki. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *J. Audio Eng. Soc*, 45(6):456–466, 1997.