Automatic Approximation of Head-Related Transfer Functions Using Parametric IIR Filters

Patrick Nowak[1,2], Véronique Zimpfer[1] and Udo Zölzer[2]
[1] French-German Research Institute of Saint-Louis, 68300 Saint-Louis, France, Email: veronique.zimpfer@isl.eu
[2] Helmut Schmidt University, 22043 Hamburg, Germany, Email: patrick.nowak@hsu-hh.de, zoelzer@hsu-hh.de

Abstract

Head-related transfer functions (HRTFs) are used in many applications for 3D spatial audio through headphones. Often, the HRTFs are stored as FIR filters. However, IIR filters give the opportunity to approximate the magnitude of these FIR filters with fewer coefficients. By using a cascade of parametric IIR filters such as shelving and peak filters, the amount of stored data can be reduced to three parameters (center frequency, gain and Q-factor) per filter stage. In the first step of the design process, the low- and high-frequency shelving filters are adjusted. Secondly, peak filters are added consecutively until the error is inside the given tolerance. After including a new peak filter, the cascade of IIR filters is post-optimized in order to yield the best approximation for the current number of peak filters. In this work, the minimum number of peak filters needed to approximate HRTFs within a given error tolerance is evaluated for different directions.

Introduction

When producing spatial audio through headphones, usually head-related transfer functions (HRTFs) are used, which are defined as transfer functions between an external sound source and the human ear. These HRTFs contain monaural spectral cues, like peaks and notches, for the vertical localization as well as interaural cues, like interaural time and level differences (ITD, ILD), for horizontal localization. Often, the HRTFs are stored as finite impulse response (FIR) filters. However, infinite impulse response (IIR) filters can be used to approximate the magnitude of these FIR filters with fewer coefficients.

In [1], Hasegawa et al. have shown that approximating HRTFs with a cascade of four to seven second-order IIR filters can lead to similar localization results in the horizontal plane as using the original FIR implementation. Furthermore, in [2], Zhixin and Cheung-Fat have modeled HRTFs by IIR filters via common factor decomposition analysis. Here, HRTFs from the same azimuth share common poles and HRTFs from the same elevation share common zeros. Moreover, Ramos and Cobos [3] have proposed to use a cascade of one second-order low-frequency shelving (LFS) filter and multiple peak filters in order to approximate HRTFs. The principle of the approximation is based on the loudspeaker equalization method from [4]. Here, the next peak filter is initialized by the biggest error area between the current approximation and the target. Afterwards, a random search based optimization is performed in order to find a parameter set close to the initial value that minimizes the error. This procedure continues for a given number of peak filters. Finally, the random search based optimization is used for post-processing triples of neighboring filters in order to improve their interaction. Furthermore, the cascade of second-order shelving and peak filters is transformed into a parallel structure of a low-pass and multiple band-pass filters in [5].

In addition to the approximation of HRTFs with IIR filters, parametric IIR filters are also used to tune given HRTFs in order to improve their performance. In [6], Yao and Chen proposed a method to adjust non-individual HRTFs based on peak filters within a user-interface in order to improve the localization capabilities. Additionally, Frank and Zotter [7] have used a supplementary high-frequency shelving (HFS) filter in order to reduce front-back confusions.

In this work, the minimum number of peak filters needed to approximate HRTFs within a given error tolerance is evaluated for different directions. At first, the LFS and HFS filters are adjusted. Then, peak filters are added consecutively with an immediate post-optimization of the whole cascade in order to yield the best approximation for the current number of peak filters. When the error is within the given tolerance, the approximation stops and the needed number of peak filters is evaluated.

In the following section, parametric IIR filters are introduced. Then, the implementation of the proposed HRTF approximation method is described. Afterwards, approximation results are shown and the needed number of peak filters is evaluated. Finally, conclusions are drawn.

Parametric IIR Filters

Shelving and peak filters amplify or attenuate frequencies in a certain band and let frequencies outside of this band pass. The transfer function of a first-order LFS filter is given in [8] as

$$H_{\mathrm{LFS}}(z) = 1 + \frac{H_0}{2}[1 + A_1(z)], \tag{1}$$

where $A_1(z)$ is a first-order all-pass filter given by

$$A_1(z) = \frac{a + z^{-1}}{1 + az^{-1}} \tag{2}$$

and $H_0 = 10^{G/20} - 1$ with G as the gain in dB. The coefficient a controls the cut-off frequency of the shelving filter. In [8] it is shown that the coefficient a has to be calculated differently for boost and cut case in order to achieve a symmetric shelving filter. Thus, the coefficient is given by

$$a_{\mathrm{B}} = \frac{\tan(\pi f_c/f_s) - 1}{\tan(\pi f_c/f_s) + 1} \tag{3}$$

for the boost case ($G \geq 0\,\text{dB}$) and

$$a_C = \frac{\tan(\pi f_c/f_s) - V_0}{\tan(\pi f_c/f_s) + V_0} \tag{4}$$

for the cut case ($G < 0\,\text{dB}$). Here, f_c denotes the cut-off frequency of the shelving filter, f_s is the sampling frequency and $V_0 = 10^{G/20}$. In order to achieve a symmetric first-order HFS filter, Eq. (1) is changed to

$$H_{\text{HFS}}(z) = 1 + \frac{H_0}{2}[1 - A_1(z)]. \tag{5}$$

Here, the same coefficient a_B from Eq. (3) can be used for the boost case, but the coefficient for the cut case has to be modified to

$$a_C = \frac{V_0 \cdot \tan(\pi f_c/f_s) - 1}{V_0 \cdot \tan(\pi f_c/f_s) + 1}. \tag{6}$$

In the same form, the transfer function of a second-order peak filter is expressed as

$$H_{\text{Peak}}(z) = 1 + \frac{H_0}{2}[1 - A_2(z)], \tag{7}$$

where $A_2(z)$ is a second-order all-pass filter given by

$$A_2(z) = \frac{-a + d(1-a)z^{-1} + z^{-2}}{1 + d(1-a)z^{-1} - az^{-2}} \tag{8}$$

with

$$d = -\cos(2\pi f_c/f_s) \tag{9}$$

controlling the center frequency, and a_B and a_C being the same as the ones from the LFS given in Eq. (3) and Eq. (4), respectively. However, in this context, f_c has to be replaced by the bandwidth f_b. The Q-factor of a peak filter is defined as

$$Q = \frac{f_c}{f_b} \quad \Leftrightarrow \quad f_b = \frac{f_c}{Q}. \tag{10}$$

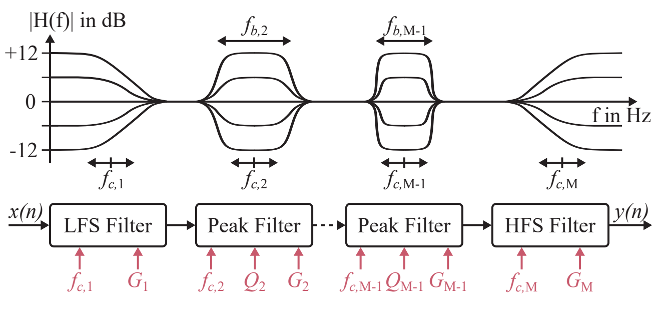

Figure 1: IIR filter cascade with M filter stages: one LFS and one HFS controlled by the cut-off frequencies ($f_{c,1}$, $f_{c,\text{M}}$) and the gains (G_1, G_M), and $M-2$ peak filters controlled by the center frequencies ($f_{c,2}$, ..., $f_{c,\text{M-1}}$), the Q-factors (Q_2, ..., $Q_{\text{M-1}}$) and the gains (G_2, ..., $G_{\text{M-1}}$).

In order to control the whole frequency range, a cascade of M parametric IIR filters can be used (see Fig. 1). This cascade consists of one LFS, $M-2$ peak filters and one HFS. All filter stages can be adapted individually by tuning their parameters $f_{c,m}$, G_m and Q_m.

HRTF Approximation

For the approximation of the HRTFs with IIR filters, the cascade shown in Fig. 1 is used. The flow chart of the approximation procedure can be seen in Fig. 2. The procedure is similar to the methods used in [9] for equalizing loudspeakers and [10] for modeling guitar amplifiers.

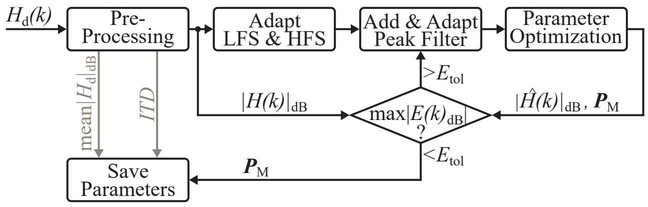

Figure 2: Flow chart of the approximation procedure with the original HRTF $H_\text{d}(k)$, the target magnitude response $|H(k)|_\text{dB}$, the approximated magnitude response $|\hat{H}(k)|_\text{dB}$, the parameter matrix $\mathbf{P}_\text{M}$, the approximation error $E(k)_\text{dB}$, and the error tolerance E_tol.

Before approximating the magnitude of the HRTF $H_\text{d}(k)$, the transfer function is pre-processed in order to yield the target magnitude response $|H(k)|_\text{dB}$. This pre-processing consists of the calculation of the magnitude in dB, the extraction of the ITD, $1/12^\text{th}$-octave smoothing, and the subtraction of the mean magnitude in dB. Here, the evaluated frequencies are exponentially spaced with frequency bins located at

$$f_k = 1000\,\text{Hz} \cdot 2^{\frac{-17K+k}{3K}}, \tag{11}$$

where $k \in \{0, 1, ..., 30K\}$ and $K = 16$ are chosen in order to achieve a $1/48^\text{th}$-octave resolution in the audible frequency range between 20 Hz and 20 kHz.

Once the pre-processing is done, an LFS and an HFS filter are adapted to approximate the behavior of the target magnitude response $|H(k)|_\text{dB}$ in the low and high frequencies. At first, the gains of the shelving filters are fixed to $G_1 = |H(0)|_\text{dB}$ for the LFS and $G_\text{M} = |H(30K)|_\text{dB}$ for the HFS. Afterwards, the cut-off frequencies $f_{c,1}$ and $f_{c,\text{M}}$ of the filters are linearly increased in order to find the parameters that minimize the error E_dB. Here, E_dB is defined as the log-spectral distance (LSD) between the target magnitude response $|H(k)|_\text{dB}$ and the approximated magnitude response $|\hat{H}(k)|_\text{dB}$. Thus, the error can be calculated by

$$E_\text{dB} = \sum_{k=0}^{30K} (E(k)_\text{dB})^2, \tag{12}$$

where

$$E(k)_\text{dB} = |H(k)|_\text{dB} - |\hat{H}(k)|_\text{dB} \tag{13}$$

contains the error for every frequency bin f_k. For the LFS, 500 different cut-off frequency values are tested between 20 Hz and 2 kHz. The one that minimizes the error E_dB is taken as the initial cut-off frequency $f_{c,1}$ for the LFS. If none of the tested frequencies reduces the initial error achieved without using an LFS, the gain of the LFS is set to $G_1 = 0\,\text{dB}$. Similarly, the initial cut-off frequency $f_{c,\text{M}}$ of the HFS is found in the range of 5 kHz

to 20 kHz. Finally, the approximation with only shelving filters $|\hat{H}_0(k)|_{\mathrm{dB}}$ is subtracted from the target response $|H(k)|_{\mathrm{dB}}$ to get the approximation error per frequency bin $E(k)_{\mathrm{dB}}$, which defines the new target response for the first peak filter.
After the LFS and HFS are set, a first peak filter is added and initialized. During the initialization the center frequency of the peak filter is fixed to the frequency bin with the highest error ($f_{c,2} = f_{k_{\max}}$) and the gain is fixed to the current error at this frequency ($G_2 = E(k_{\max})_{\mathrm{dB}}$), where $k_{\max}$ defines the index with the maximum absolute error $k_{\max} = \arg\max_k |E(k)_{\mathrm{dB}}|$. Analogous to finding the cut-off frequency for the shelving filters, the Q-factor of the peak filter is searched in the range of $Q = 1$ to $Q = 100$ and the Q-factor that minimizes the approximation error E_{dB}, is taken as parameter Q_2.
Afterwards, the parameter optimization block tunes the whole cascade in order to minimize the approximation error E_{dB} for the current number of peak filters by optimizing the interaction between the filter stages. For the optimization, the Levenberg-Marquardt algorithm is used. Here, the target magnitude response $|H(k)|_{\mathrm{dB}}$ and the parameter matrix $\mathbf{P}_{\mathrm{M}}$ are used as inputs. The parameter matrix is defined as $\mathbf{P}_{\mathrm{M}} = [\mathbf{f_c}, \mathbf{G}, \mathbf{Q}]$, where $\mathbf{f_c} = [f_{c,1}, ..., f_{c,\mathrm{M}}]^T$, $\mathbf{G} = [G_1, ..., G_{\mathrm{M}}]^T$, and $\mathbf{Q} = [Q_1, ..., Q_{\mathrm{M}}]^T$. Since the shelving filters are implemented as first-order filters, Q_1 and Q_{M} are not used and therefore set to zero. The outputs of the parameter optimization are updated values for the approximated magnitude response $|\hat{H}(k)|_{\mathrm{dB}}$ and the parameter matrix $\mathbf{P}_{\mathrm{M}}$, where $M - 2$ defines the number of used peak filters.
When the optimization is done for the current number of peak filters, it is checked whether the maximum absolute approximation error $\max |E(k)_{\mathrm{dB}}|$ is lower than the given error tolerance E_{tol}. If this is the case or the second termination condition of 30 peak filters is reached, the approximation is finished and the parameters are saved. Otherwise, the approximation continues by adding a new peak filter and optimizing all parameters afterwards.

Evaluation

In order to evaluate the minimum number of peak filters needed to achieve a given error tolerance, HRTFs taken from the CIPIC Database [11] are approximated. The database contains HRTFs of 45 subjects measured for 1250 directions in the interaural-polar coordinate system. These directions consist of 25 azimuthal angles $\phi = [-80°, -65°, -55°, -45°, -40°, ..., 45°, 55°, 65°, 80°]$ and 50 elevation angles uniformly distributed between $\theta = -45°$ and $\theta = 225°$ with steps of 5.625°.
In this work, the approximations are performed for an azimuthal resolution of 15° inside $|\phi| \leq 45°$ and all azimuths outside of this range. Moreover, a resolution of 45° was chosen for the elevation, maintaining a total of 91 evaluated directions. Additionally, the results for both ears are combined, leading to an altered interpretation of the azimuthal directions. Here, positive azimuths φ denote ipsilateral directions and negative ones contralateral directions. This data augmentation leads to 90 approximated magnitude responses per direction. Furthermore, an error tolerance of $E_{\mathrm{tol}} = 2\,\mathrm{dB}$ was chosen. Fig. 3 shows an exemplary approximation of the left ear's magnitude response of 'subject_008' from the CIPIC database for the frontal direction ($\varphi = 0°$, $\theta = 0°$). It can be seen, that the error $E(k)_{\mathrm{dB}}$ falls within the given tolerance $E_{\mathrm{tol}} = 2\,\mathrm{dB}$.

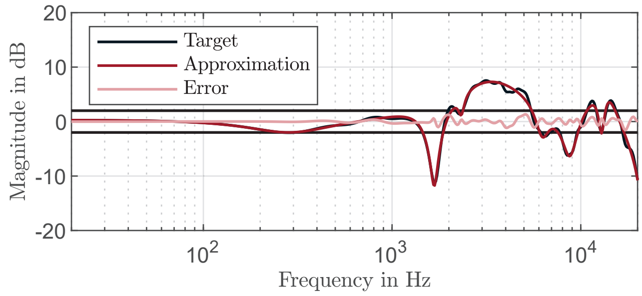

Figure 3: Magnitude responses of the target $|H(k)|_{\mathrm{dB}}$ and the approximation $|\hat{H}(k)|_{\mathrm{dB}}$ for the left ear of 'subject_008' and the frontal direction ($\varphi = 0°$, $\theta = 0°$). Additionally, the error $E(k)_{\mathrm{dB}}$ and the tolerance $E_{\mathrm{tol}} = 2\,\mathrm{dB}$ are plotted.

The magnitude responses of the eight peak filters used to yield the approximation shown in Fig. 3 are visualized in Fig. 4. Additionally, the magnitude responses of the shelving filters are plotted.

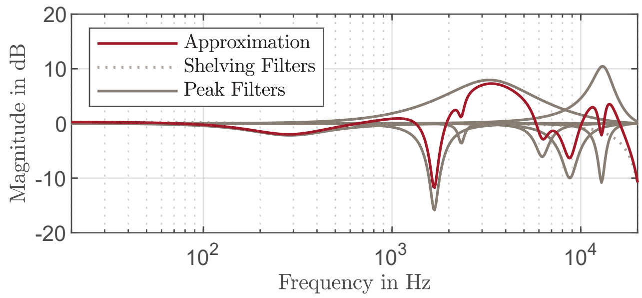

Figure 4: Magnitude responses of the individual filter stages of the approximation $|\hat{H}(k)|_{\mathrm{dB}}$ seen in Fig. 3.

It is visible in Fig. 3 and Fig. 4 that eight peak filters are enough to approximate the target magnitude response with a given error tolerance of $E_{\mathrm{tol}} = 2\,\mathrm{dB}$. However, this number is only valid for the given target response ('subject_008', left ear, $\varphi = 0°$, $\theta = 0°$). Therefore, Fig. 5 shows the relative and cumulative frequency of the needed number of peak filters for all 8190 approximated directions.

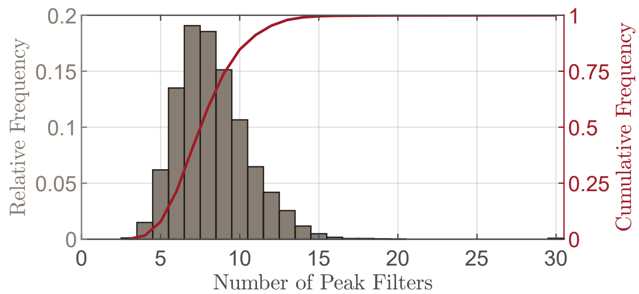

Figure 5: Relative and cumulative frequency of the needed number of peak filters over all 8190 approximations.

It can be seen, that most of the approximations need seven (19.1 %) or eight (18.6 %) peak filters to fulfill the given error tolerance of $E_{\mathrm{tol}} = 2\,\mathrm{dB}$. Additionally, the cumulative frequency shows that, when using ten peak filters, 84.7 % of the approximations are within the given error tolerance. Although, most of the approximations need at most ten peak filters, values of up to 20 or even 30 are required for some approximations. In the following, a relation between the needed number of peak filters and the direction is analyzed. Fig. 6 shows a heatmap of the average number of needed peak filters per direction. In general, a rise in the average number of needed peak filters is seen from ipsilateral ($\varphi \geq 0°$) to contralateral directions ($\varphi < 0°$). Ipsilateral directions $\varphi \geq 45°$ require 6.4 to 7.8 peak filters and contralateral directions $\varphi \leq -45°$ 7.8 to 11.6. Especially, the magnitude responses originating from frontal contralateral directions show average numbers of needed peak filters above ten.

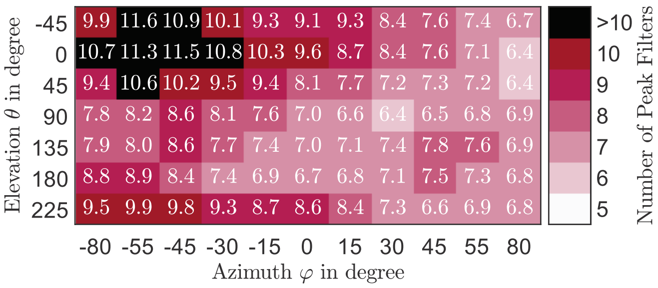

Figure 6: Heatmap of the average number of needed peak filters per direction for 90 subjects. Here, positive values of φ define ipsilateral directions.

Since outliers can strongly increase the average value, also other statistical values like the minimum, first quartile, median, third quartile, and maximum are evaluated. In Fig. 7, these statistical values are combined into a box plot. Here, frontal and rear directions with an azimuthal spacing of around 30° are analyzed. The box plot confirms the trends seen in Fig. 6, but adds information about the spread of the needed number of peak filters between approximations for the same direction. The maximum interquartile range is 3 peak filters.

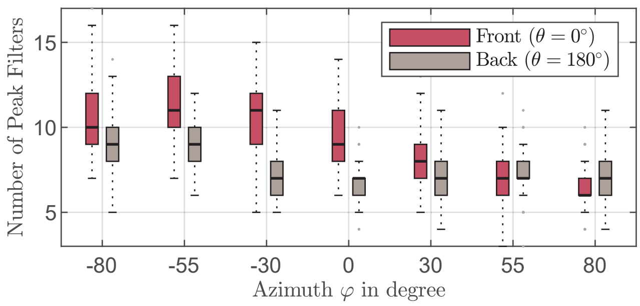

Figure 7: Box plot of the needed number of peak filters for frontal and rear directions for 90 subjects with minimum, first quartile, median, third quartile, maximum, and outliers.

Conclusions

In the present work the magnitudes of HRTFs originating from different directions are approximated with parametric IIR filters and the minimum number of needed peak filters in order to fulfill a given error tolerance is evaluated. In 84.7 % of the cases, at most ten peak filters are required to produce an approximation error that falls within the 2 dB tolerance. However, the amount of required peak filters increases when moving from ipsilateral to contralateral directions.

In future investigations, a listening test will be realized in order to confirm the usability of the approximated magnitude responses. Additionally, the interpolation of parameters will be investigated to incorporate directions in between.

References

[1] H. Hasegawa, M. Kasuga, S. Matsumoto, and A. Koike: Simply realization of sound localization using HRTF approximated by IIR filter. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Science (2000), vol. E83-A, no. 6, 973-978

[2] W. Zhixin and C.-F. Chan: Continuous function modeling of head-related impulse response. IEEE Signal Processing Letters (2015), vol. 22, no. 3, 283-287

[3] G. Ramos and M. Cobos: Parametric head-related transfer function modeling and interpolation for cost-efficient binaural sound applications. The Journal of the Acoustical Society of America (2013), vol. 134, no. 3, 1735-1738

[4] G. Ramos and J. J. López: Filter design method for loudspeaker equalization based on IIR parametric filters. Journal of the Audio Engineering Society (2006), vol. 54, no. 12, 1162-1178

[5] G. Ramos, M. Cobos, B. Bank, and J. A. Belloch: A parallel approach to HRTF approximation and interpolation based on a parametric filter model. IEEE Signal Processing Letters (2017), vol. 24, no. 10, 1507-1511

[6] S.-N. Yao and L. Chen: HRTF adjustments with audio quality assessments. Archives of Acoustics (2013), vol. 38, no. 1, 55-62

[7] M. Frank and F. Zotter: Simple reduction of front-back confusion in static binaural rendering. Fortschritte der Akustik: DAGA (2018)

[8] U. Zölzer: Digital audio signal processing. John Wiley & Sons Ltd, Chichester, 2nd edition, 2008

[9] H. Behrends, A. von dem Knesebeck, W. Bradinal, P. Neumann, and U. Zölzer: Automatic equalization using parametric IIR filters. Journal of the Audio Engineering Society (2011), vol. 59, no. 3, 102-109

[10] F. Eichas and U. Zölzer: Gray-box modeling of guitar amplifiers. Journal of the Audio Engineering Society (2018), vol. 66, no. 12, 1006-1015

[11] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano: The CIPIC HRTF database. IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (2001)