Spatial Interpolation of HRTFs Approximated by Parametric IIR Filters

Patrick Nowak, Udo Zölzer

Helmut Schmidt University, 22043 Hamburg, Germany, Email: patrick.nowak@hsu-hh.de, zoelzer@hsu-hh.de

Abstract

In binaural synthesis through headphones, head-related transfer functions (HRTFs) are used to position virtual sound sources in the three-dimensional space around the listener. In practice, only a finite number of measured HRTFs is available, resulting in a limited spatial resolution. One approach to increase the spatial resolution is the usage of spatial interpolation of the measured HRTFs. When the measured HRTFs are represented as finite impulse response (FIR) filters, bilinear rectangular or triangular interpolation can be used to compute the filter coefficients of an intermediate direction. However, when the measured HRTFs are represented as infinite impulse response (IIR) filters instead, the interpolation of the coefficients is not as straightforward as for FIR filters due to stability considerations. In this work, an interpolation algorithm is proposed targeting HRTFs represented as cascades of parametric IIR filters. This interpolation algorithm is based on the bilinear interpolation of the parameters of the individual filter stages (center frequency, gain, and Q-factor) together with an assignment of related peak filters. In order to evaluate the proposed interpolation algorithm, two listening tests are performed including static virtual sound sources as well as moving virtual sound sources. The results confirm the validity of the proposed interpolation algorithm.

Introduction

In order to reduce the number of saved parameters per HRTF during binaural synthesis through headphones, measured HRTFs can be approximated using cascades of parametric IIR filters [1, 2]. This cascade contains one first-order low-frequency shelving filter (LFS), $M-2$ second-order peak filters, and one high-frequency shelving filter (HFS) (see Fig. 1). First-order LFS and HFS are controlled by their cut-off frequencies ($f_{c,\mathrm{L}}$, $f_{c,\mathrm{H}}$) and gains (G_{L}, $G_{,\mathrm{H}}$) whereas second-order peak filters are controlled by center frequency $f_{c,i}$, gain G_i, and Q-factor Q_i. In this work, ten peak filters are used to approximate measured HRTFs ($M = 12$). Since measuring HRTFs with a high spatial resolution is time-consuming and expensive, only a finite number of measured HRTFs is available. Here, spatial interpolation of HRTFs can be used to increase the number of possible static virtual sound source directions during binaural synthesis. Additionally, a high spatial resolution of measured HRTFs is required for smooth transitions in moving virtual sound sources. Thus, the following section summarizes certain methods for spatial interpolation of HRTFs represented as FIR and IIR filters. Then, the proposed parametric interpolation is described and evaluated using two listening tests. Finally, conclusions are drawn and suggestions for further research are made.

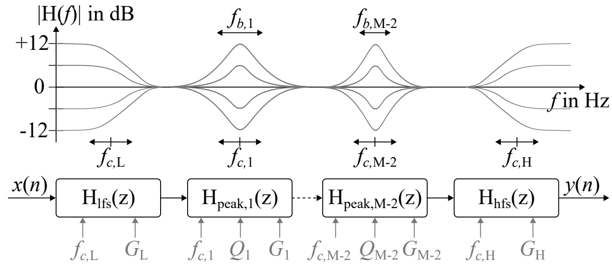

Figure 1: Block diagram of the parametric IIR filter cascade containing one LFS, $M-2$ peak filters, and one HFS. Additionally, exemplary magnitude responses are shown.

Spatial Interpolation

For HRTFs represented as FIR filters, the coefficients can be interpolated in time- or frequency-domain using bilinear rectangular or triangular interpolation [3, 4] of neighboring head-related impulse responses (HRIRs) or HRTFs, respectively. Figure 2 illustrates the principle of bilinear rectangular interpolation using a measurement grid with azimuthal resolution $\Delta\varphi$ and elevation resolution $\Delta\theta$. By calculating the weighted sum of neighboring HRIRs ($h_1(n)$, $h_2(n)$, $h_3(n)$, and $h_4(n)$), the HRIR of an intermediate direction $\bar{h}(n)$ can be calculated as

$$\begin{aligned}\bar{h}(n) = {} & \overbrace{(1-c_\varphi)(1-c_\theta)}^{c_1}\, h_1(n) + \overbrace{c_\varphi(1-c_\theta)}^{c_2}\, h_2(n) \\ & + \underbrace{c_\varphi c_\theta}_{c_3}\, h_3(n) + \underbrace{(1-c_\varphi)c_\theta}_{c_4}\, h_4(n)\end{aligned} \tag{1}$$

with

$$c_\varphi = \frac{C_\varphi}{\Delta\varphi} = \frac{\varphi \bmod \Delta\varphi}{\Delta\varphi}, \quad c_\theta = \frac{C_\theta}{\Delta\theta} = \frac{\theta \bmod \Delta\theta}{\Delta\theta}. \tag{2}$$

In order to avoid audible artifacts due to different time of arrivals inside neighboring HRIRs, minimum-phase approximated HRIRs should be used [5]. When the measured HRTFs are represented as IIR filters instead, the interpolation of the coefficients is not as straightforward as for FIR filters due to stability considerations [6]. Thus, different interpolation methods have been developed for IIR filters, e.g. interpolation of pole-zero models [6, 7]. Additionally, in [8], the possibility of interpolating the parameters of the individual filter stages in the cascade is mentioned. In this work, an extension of this parametric interpolation is proposed when using HRTFs approximated by parametric IIR filters.

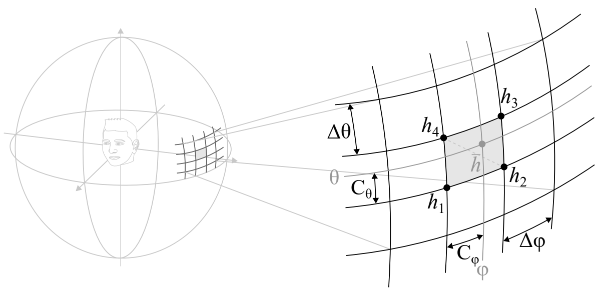

Figure 2: Principle of bilinear rectangular interpolation in order to calculate an intermediate HRIR $\bar{h}(n)$ at relative position $(C_\varphi,\, C_\theta)$ inside a given area defined by measured HRIRs $(h_1(n),\, h_2(n),\, h_3(n)$, and $h_4(n))$.

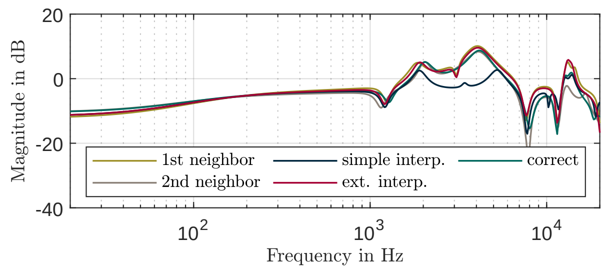

Figure 3: Interpolated left ear magnitude response of *Subject_065* from CIPIC database ($\varphi = -10°$, $\theta = 0°$) using a resolution of $\Delta\varphi = 15°$ and $\Delta\theta = 11.25°$, and first ($\varphi = -15°$, $\theta = 0°$) and second neighbor ($\varphi = 0°$, $\theta = 0°$).

Parametric Interpolation

Simple Parametric Interpolation

Simple parametric interpolation uses bilinear rectangular interpolation as given in Eq. (1) in order to calculate the parameter matrix of an intermediate direction

$$\bar{\mathbf{P}} = c_1\mathbf{P}_1 + c_2\mathbf{P}_2 + c_3\mathbf{P}_3 + c_4\mathbf{P}_4 \tag{3}$$

from the parameter matrices of neighboring directions

$$\mathbf{P}_i = \begin{bmatrix} f_{c,\mathrm{L},i} & G_{\mathrm{L},i} & 1 \\ f_{c,1,i} & G_{1,i} & Q_{1,i} \\ \vdots & \vdots & \vdots \\ f_{c,\mathrm{M\text{-}2},i} & G_{\mathrm{M\text{-}2},i} & Q_{\mathrm{M\text{-}2},i} \\ f_{c,\mathrm{H},i} & G_{\mathrm{H},i} & 1 \end{bmatrix}. \tag{4}$$

Similarly, mean magnitude responses $\mu_{\mathrm{H}_i,\mathrm{dB}}$ that are subtracted before approximation and extracted interaural time differences ITD_i are interpolated according to

$$\mu_{\bar{\mathrm{H}},\mathrm{dB}} = \sum_{i=1}^{4} c_i \mu_{\mathrm{H}_i,\mathrm{dB}}, \tag{5}$$

$$\mathrm{I\bar{T}D} = \sum_{i=1}^{4} c_i \mathrm{ITD}_i. \tag{6}$$

In Fig. 3, an interpolation result is shown for *Subject_065* from the CIPIC database [9] and an intermediate direction ($\varphi = -10°$, $\theta = 0°$). Here, a measurement grid with azimuthal resolution $\Delta\varphi = 15°$ and elevation resolution $\Delta\theta = 11.25°$ is used. Since the intermediate direction lies on an edge of the interpolation rectangle ($c_\theta = 0$), only two neighboring directions are used for interpolation, namely first ($\varphi = -15°$, $\theta = 0°$) and second neighbor ($\varphi = 0°$, $\theta = 0°$). As can be seen from the magnitude responses, the simple interpolation leads to an attenuation between 2 and 5 kHz. In order to explain this attenuation, Fig. 4 illustrates the magnitude responses of the individual filter stages for the interpolated direction as well as the neighboring directions. Although both neighboring directions contain a similar peak filter at 4 kHz, this peak filter does not arise in the interpolated direction, resulting in the attenuation visible in Fig. 3. The reason for the missing peak filter is given by an incorrect assignment of peak filters inside neighboring directions.

Since the peak filter at 4 kHz is numbered fourth for first neighbor and third for second neighbor, the similar peak filters at 4 kHz are interpolated with different peak filters rather than with each other, leading to a missing amplification between 2 and 5 kHz.

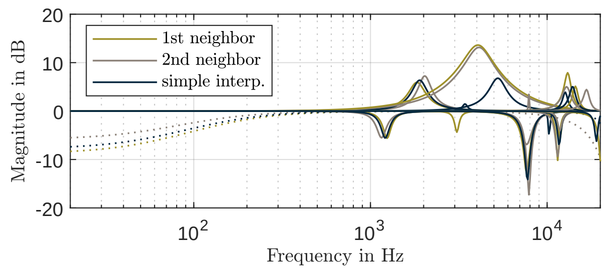

Figure 4: Magnitude responses of the individual filter stages for the simple parametric interpolation as well as the neighboring directions from Fig. 3.

Extended Parametric Interpolation

Since simple parametric interpolation suffers from an incorrect assignment, extended parametric interpolation targets on assigning the peak filters of neighboring directions appropriately before interpolation. This assignment process is given by the following steps:

1. Find closest neighboring direction

$$i_{\mathrm{ref}} = \arg\max_i c_i \text{ for } i \in \{1, 2, 3, 4\}.$$

2. Take $\mathbf{P}_{i_{\mathrm{ref}}}$ of closest direction $(\varphi_{i_{\mathrm{ref}}},\, \theta_{i_{\mathrm{ref}}})$ as reference parameter matrix.

3. For every peak filter $p_{i_{\mathrm{ref}}}$ of the reference direction, check the other neighboring directions ($i \neq i_{\mathrm{ref}}$) for peak filters p_i within a threshold of

$$\Delta f_c = \left| 20 \log_{10}\left(\frac{f_{c,\mathrm{p},i_{\mathrm{ref}}}}{f_{c,\mathrm{p},i}}\right)\right| \leq 2\,\mathrm{dB}.$$

4a. If a peak filter p_i is found, compare $\mathrm{sgn}(G_{\mathrm{p},i})$ and $\mathrm{sgn}(G_{\mathrm{p},i_{\mathrm{ref}}})$ in order to interpolate only peak filters that either boost or cut the given frequency range. If more than one peak filter is found, choose the most similar one. Additionally, check whether one of the next two reference peak filters is more similar.

In this case, exclude the assigned peak filter from the current interpolation. If an assigned peak filter is found, calculate interpolated parameters of the current peak filter $p_{i_{\mathrm{ref}}}$ as

$$\left[\bar{f}_{c,\mathrm{p}_{i_{\mathrm{ref}}}} \bar{G}_{\mathrm{p}_{i_{\mathrm{ref}}}} \bar{Q}_{\mathrm{p}_{i_{\mathrm{ref}}}}\right] = \sum_{i=1}^{4} c_i \cdot \left[f_{c,\mathrm{p},i} \;\; G_{\mathrm{p},i} \;\; Q_{\mathrm{p},i}\right].$$

4b. If no peak filter p_i is assigned to the reference peak filter $p_{i_{\mathrm{ref}}}$, exclude this neighbor from interpolation and modify weights to

$$c_{i,\mathrm{p}} = \frac{c_i}{1 - c_{i_{\mathrm{ex}}}} \qquad \text{for } i \neq i_{\mathrm{ex}}$$

for that specific peak filter $p_{i_{\mathrm{ref}}}$ and calculate interpolated parameters of current peak filter $p_{i_{\mathrm{ref}}}$ as

$$\left[\bar{f}_{c,\mathrm{p}_{i_{\mathrm{ref}}}} \bar{G}_{\mathrm{p}_{i_{\mathrm{ref}}}} \bar{Q}_{\mathrm{p}_{i_{\mathrm{ref}}}}\right] = \sum_{\substack{i=1 \\ i \neq i_{\mathrm{ex}}}}^{4} c_{i,\mathrm{p}} \cdot \left[f_{c,\mathrm{p},i} \;\; G_{\mathrm{p},i} \;\; Q_{\mathrm{p},i}\right].$$

5. Continue assignment of peak filters with next peak filter $p_{i_{\mathrm{ref}}}+1$. Start comparison with $\{p_i+1\}^{\mathrm{th}}$ peak filter for every neighboring direction i, where p_i is defined by the previously assigned peak filter.

6. Interpolate mean magnitude values $\mu_{\mathrm{H}_i,\mathrm{dB}}$ and ITDs according to Eqs. (5) and (6), respectively.

The influence of extended parametric interpolation on the resulting interpolated magnitude response is shown in Fig. 3. As can be seen, the attenuation between 2 and 5 kHz disappears for extended parametric interpolation.

Extensions for Moving Virtual Sound Sources

For generating moving virtual sound sources, smooth transitions between the magnitude responses of intermediate directions are needed. Since especially time-variant IIR filter implementations suffer from audible clicks due to the mismatch between updated coefficients and internal states of the recursive parts, two IIR filter cascades are connected in parallel using a combination of cross-fading [10] and input-switching [11] as shown in Fig. 5.

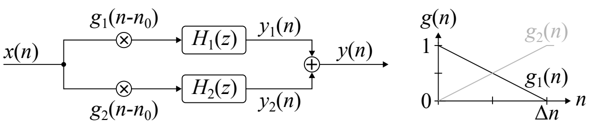

Figure 5: Combining two IIR filter cascades in parallel using cross-fading input-switching in order to avoid audible clicks during the update of IIR filter coefficients while generating moving virtual sound sources.

Although cross-fading input-switching avoids audible clicks while updating the coefficients of the IIR filter cascades, peak filters contained only in a single neighboring direction can lead to strong changes in magnitude response when moving across the center of interpolation rectangles, resulting in audible coloration. In order to prevent this coloration, a normalization of the interpolation weight $c_{i_{\mathrm{ref}},\mathrm{p}}$ for gain $G_{\mathrm{p},i_{\mathrm{ref}}}$ of the p^{th} peak filter of the closest neighboring direction i_{ref} between 0 and 1 can be used, if only this direction contains a peak filter in the given frequency region. For an interpolation in the horizontal plane with only two neighboring directions, this normalization can be given as

$$\tilde{c}_{i_{\mathrm{ref}},\mathrm{p}} = \frac{2c_{i_{\mathrm{ref}},\mathrm{p}} - 1}{c_{i_{\mathrm{ref}},\mathrm{p}}}. \tag{7}$$

Listening Tests

In order to evaluate the proposed parametric interpolation, two listening tests are performed. The first listening test compares the localization accuracy of measured and interpolated static virtual sound sources whereas the second listening test evaluates audio quality ratings of moving virtual sound sources achieved using the proposed parametric interpolation and conventional bilinear rectangular interpolation of FIR filters.

Listening Test I: Localization Accuracy

In the fist listening test, both individual and non-individual dummy-head HRIRs are used. Here, measurements are taken in the horizontal plane with a resolution of $\Delta\varphi = 30°$ to form the measurement grid ($-150°$, $-120°$, ..., $180°$) required for interpolation. In between of these directions, interpolated directions ($-165°$, $-135°$, ..., $165°$) are calculated using extended parametric interpolation as well as bilinear rectangular interpolation of FIR filters. Additionally, HRIR measurements are taken for these interpolated directions, too, in order to compare the localization accuracy of measured and interpolated directions. As stimulus, a snap is used that contains a broad frequency range. During the listening test, the subjects are asked to listen to 168 different stimuli and giving the perceived azimuthal sound source direction by rotating an arrow to this direction. Overall, eight subjects participated in the listening test. All subjects are male research assistants in the Department of Signal Processing and Communication in the age of 26 to 37 years with an average age of 30.9 years. Every subject performed the listening test at home using a different over-ear headphone. In order to increase the number of results per direction, every subject performed the listening test twice with a duration of 30 to 40 minutes per session. The results are summarized in Tab. 1. As can be seen, measured and interpolated filters show similar localization results in mean angular error $\bar{\varphi}_{\mathrm{error}}$ and front/back confusion rate ρ_{fb} for bilinear rectangular interpolation of minimum-phase approximated FIR filters as well as parametric interpolation of parametric IIR filter cascades.

Listening Test II: Moving Virtual Sound Sources

In the second listening test, audio quality ratings of moving virtual sound sources are evaluated containing stimuli generated by bilinear rectangular interpolation of HRIRs (*FIR*) and minimum-phase approximated HRIRs (*minPh.*), a single parametric IIR filter cascade (*IIR*),

Table 1: Results of the first listening test summarizing the mean angular error $\bar{\varphi}_{\text{error}}$ and front/back confusion rate ρ_{fb} achieved using different filter types containing FIR filter implementations of HRIRs and parametric IIR filter approximations.

Filter type	FIR			IIR	
	Meas.	minPh.	Interp.	Meas.	Interp.
$\bar{\varphi}_{\text{error}}$ in degree	15.8 / 18.3	17.0 / -	16.8 / 17.0	16.3 / 18.6	16.2 / 18.2
ρ_{fb} in percent	33.6 / 33.3	30.5 / -	37.6 / 35.6	36.1 / 48.3	28.0 / 45.7

individual / non-individual (dummy-head)

two parallel IIR filter cascades with input-switching without fading (*IIR2*), two parallel IIR filter cascades with cross-fading input-switching (*IIR2f*), and two parallel IIR filter cascades with cross-fading input-switching and additional smoothing according to Eq. (7) (*IIR2f*). Here, non-individual HRIRs of *Subject_065* from the CIPIC database [9] are used. The listening test includes three stimuli (speech, music, noise) and three different scenarios ($-40° \rightarrow 40°$ with $\Delta\varphi = \{5°, 15°\}$, $-160° \rightarrow -140°$ with $\Delta\varphi = 15°$). During the listening test, all six filter types are evaluated in 9 comparisons, with the subject rating each stimulus with an audio quality between 0 (bad quality) and 100 (very good quality). Every comparison is performed twice. The same eight subjects participated in the second listening test than in the first one.

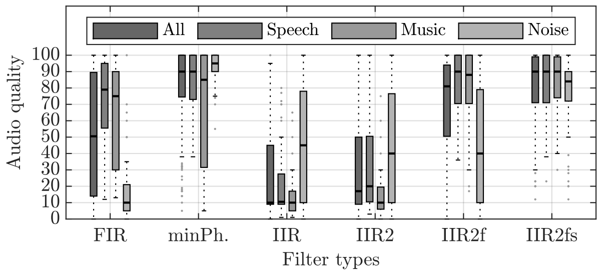

Figure 6: Box-and-whisker plots representing the audio quality for different filter types containing lower whisker, first quartile, median, third quartile, upper whisker, and outliers.

As can be seen in Fig. 6, the usage of minimum-phase approximated HRIRs improves the audio quality of FIR filter implementations due to the cancellation of audible comb filtering effects. Using two parallel IIR filter cascades with cross-fading input-switching shows similar audio quality ratings than FIR filter implementations of minimum-phase approximated HRIRs, with the exception of audible coloration for the noise stimulus in *IIR2f*. On the contrary, a single IIR filter cascade or two parallel IIR filter cascades without fading show worse audio quality ratings for all stimuli due to audible clicks.

Conclusion

In this work, an interpolation algorithm for parametric IIR filter cascades is proposed, enabling intermediate static virtual sound sources and smooth transitions inside moving virtual sound sources. The interpolation algorithm is based on bilinear interpolation of the parameters of the individual filter stages together with an assignment of related peak filters. Two listening tests confirm the validity of the proposed interpolation algorithm by evaluating the localization accuracy of interpolated static virtual sound sources and the audio quality of moving virtual sound sources, respectively.

References

[1] P. Nowak, V. Zimpfer, and U. Zölzer, "Automatic approximation of head-related transfer functions using parametric IIR filters", *in Fortschritte der Akustik: DAGA 2020*, 2020.

[2] P. Bhattacharya, P. Nowak, and U. Zölzer, "Optimization of cascaded parametric peak and shelving filters with backpropagation algorithm", *in Proceedings of the 23rd International Conference on Digital Audio Effects (DAFx-20)*, 2020.

[3] J.-M. Jot, V. Larcher, and O. Warusfel, "Digital signal processing issues in the context of binaural and transaural stereophony", *in Audio Engineering Society Convention 98*, 1995.

[4] F. P. Freeland, L. W. P. Biscainho, and P. S. R. Diniz, "Efficient HRTF interpolation in 3D moving sound", *in 22nd AES International Conference on Virtual, Synthetic, and Entertainment Audio*, 2002.

[5] J. Huopaniemi, "Virtual acoustics and 3-D sound in multimedia signal processing", *PhD thesis, Helsinki University of Technology*, 1999.

[6] M. Queiroz and G. H. M. Sousa, "Structured IIR models for HRTF interpolation", *in Proceedings of the 2010 International Computer Music Conference (ICMC)*, 2010.

[7] L. Wang, F. Yin, and Z. Chen, "An indirect interpolation method for head-related transfer function pole-zero models", *Acoustical Science and Technology*, vol. 29, no. 5, pp. 329-331, 2008.

[8] 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*, vol. 134, no. 3, pp. 1735-1738, 2013.

[9] V. R. Algazi, R. O. Duda, D. M. Thompson and C. Avendano, "The CIPIC HRTF database", *in Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001.

[10] V. Välimäki, "Discrete-time modeling of acoustic tubes using fractional delay filters", *PhD thesis, Helsinki University of Technology*, 1995.

[11] W. Verhelst and P. Nilens, "A modified-superposition speech synthesizer and its applications", *in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1986.