Examination of different HRTF interpolation methods

Jan Völkering, Eugen Rasumow, Matthias Blau
Institut für Hörtechnik und Audiologie, Jade Hochschule, 26121 Oldenburg, Deutschland,
Email: jan.voelkering@gmx.de

Introduction

For the presentation of virtual sound sources over headphones, the individual direction dependent head-related transfer functions (HRTF) are needed. In order to reduce the number of required HRTF measurements, it is common to measure a limited number of HRTFs and to interpolate the HRTFs for the directions in between. Different interpolation methods are analyzed in this study, both objectively and perceptually.

HRTF interpolation methods

Four different methods for spatial HRTF interpolation are implemented and evaluated. The most straightforward method is the linear interpolation of adjacent head-related impulse responses (hrir) in the time domain (method 1). It has to be considered that the initial delay of each hrir has to be matched to avoid errors by destructive interpolation. The temporally matched hrirs are interpolated by using linear and cubic spline interpolation functions. The second method is the linear interpolation of HRTF magnitudes by principal component analysis (PCA) and minimum-phase reconstruction (see [1]) (method 2). Based on the HRTF magnitudes, the basis functions and the corresponding weighting functions obtained by PCA are calculated. In this study, the maximum number of weighting functions are interpolated linearly using an order of 257. The obtained HRTFs were represented by their minimum-phase components. The third method is the separate linear and spline interpolation of HRTF magnitudes and group delays in the frequency domain (method 3). The fourth method is the linear and spline interpolation of complex-valued HRTFs in the frequency domain with a linearized phase for frequencies above 1500 Hz (method 4), which seems reasonable due to findings in [4] and [5]. For all methods the initial delays were estimated by calculating the maximum of the Hilbert envelope of the impulse responses in the time domain. These delays were interpolated separately to re-synthesize the interaural time delays (ITD).

Objective evaluation

In order to objectively evaluate the interpolation methods, a HRTF dataset of 72 directions in the horizontal plane (elevation $\phi = 0°$, azimuth $\theta = 0, ..., 355°$) with a grid resolution of $\Delta\phi = 5°$ measured with an artificial head (G.R.A.S. KEMAR) was used. The spectral distortion and the error of interaural phase differences (IPD) between the measured HRTF and the interpolated HRTF (centered between adjacent directions) are determined. The grid resolution of interpolation was 10°.

Spectral distortion

Regarding the psychoacoustic aspects of the auditory filter bands on the basilar membrane the spectra of the HRTFs were divided into 39 gammatone filter bands. The differences of log-magnitudes in each filter band between measured and interpolated HRTFs in the horizontal plane were calculated by

$$\epsilon_{i,\theta} = 20\log_{10}\left(\frac{|\mathrm{HRTF}_{\mathrm{meas},i,\theta}|}{|\mathrm{HRTF}_{\mathrm{int},i,\theta}|}\right)(\mathrm{dB}), \tag{1}$$

where $\mathrm{HRTF}_{\mathrm{meas},i,\theta}$ is the measured HRTF magnitude and $\mathrm{HRTF}_{\mathrm{int},i,\theta}$ is the interpolated HRTF magnitude at the i-th filter band and direction θ.

The absolute error $\epsilon_{i,\theta}$ of left and right earside are added and the mean across all filter bands is determined as

$$L_{\mathrm{crit},\theta} = \frac{1}{39}\sum_{i=1}^{39}(|\epsilon_{i,\theta,\mathrm{left}}| + |\epsilon_{i,\theta,\mathrm{right}}|). \tag{2}$$

$L_{\mathrm{crit},\theta}$ is related to the approach by Minnaar et al. [3]. Hence, a single measure for each direction is obtained to describe the spectral error by interpolation. The distribution of the results of all directions is depicted in the left graph of fig. 1. It can be seen that the interpolation in the time domain (method 1) induces large magnitude errors, mainly at contralateral directions. This causes the high variance of the absolute errors. The other methods yield more robust performances regarding the spectral error. Method 4, which is very similar to method 1, except that it is in the frequency domain and the phase for $f \geq 1500$ Hz is linearized, doesn't show such large errors. This can be explained by the fact that the phase linearization can avoid spectral distortion caused by phase differences between HRTFs of adjacent directions.

IPD error

The phase error is determined by comparing the IPDs of measured and interpolated HRTFs at different azimuth directions for frequencies 100 Hz $\geq f \geq$ 1300 Hz. The absolute IPD differences are reciprocally weighted with ITD-thresholds (see [4]) and summed up to

$$L_{\mathrm{IPD}} = \nu \cdot \sum_{f=100\ \mathrm{Hz}}^{1300\ \mathrm{Hz}} \frac{\Delta f}{\mathrm{ERB}(f)} \cdot \frac{|\mathrm{IPD}_{\mathrm{meas}}(f) - \mathrm{IPD}_{\mathrm{int}}(f)|}{\phi_{\mathrm{JND}}(f)} \tag{3}$$

$$\text{with} \quad \nu \cdot \sum_{f=100\ \mathrm{Hz}}^{1300\ \mathrm{Hz}} \frac{\Delta f}{\mathrm{ERB}(f)} = 1, \tag{4}$$

where Δf is the frequency resolution of the FFT. ERB is the equivalent rectangular bandwidth, $\mathrm{IPD}_{\mathrm{meas}}$ and $\mathrm{IPD}_{\mathrm{int}}$ are the measured and interpolated IPDs and

$\phi_{\mathrm{JND}}(f)$ is the interaural phase difference threshold from [2]. The results shown in the right box of fig. 1 illustrate the rather bad IPD re-synthesis using method 2 with the minimum-phase reconstruction, leading to the largest IPD errors. Method 1 and method 4 are identical in the frequency range below $f = 1500$ Hz, showing the same performance with the lowest IPD errors due to interpolation of the original phase in this frequency range.

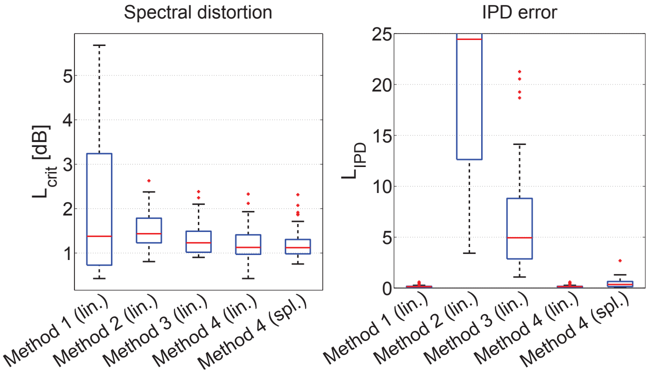

Figure 1: Right box: Distribution of spectral distortion for all directions. Left box: Distribution of the IPD error for all directions.

Listening test

For the listening test, the individual HRTFs of each of the eight participants were measured at 15 azimuth directions (five loudspeakers with a spatial resolution of 2.5° at each of $\theta = 0°$, 90° and 225° directions). The HRTFs were measured simultaneously to avoid influence by head movements. Pulsed white noise, bandpass limited for 200 Hz $\geq f \geq$ 16000 Hz was filtered with measured (reference signal) and interpolated HRTFs (test signal) and presented at $\theta = 0°$, 90° and 225° over headphones (Sennheiser HD 800). The task of the subjects was to evaluate the test signal with respect to the reference signal regarding the perception of overall quality on a category scale. The results of the perceptual evaluation of the overall quality for the grid resolution of 10° (in fig. 2) show the best performance for method 1 and method 4 for all directions. The Wilcoxon signed-rank test indicates no significant difference between these two methods in all conditions ($p \geq 0.0732$, $\alpha = 0.05$), except for $\theta = 90°$ and 10° grid resolution. In this condition method 4 is significantly different from method 1 ($p = 0.006$). The overall quality of the PCA with minimum-phase reconstruction was significantly worse than the other methods in most conditions. This is in line with the results of the introduced IPD errors. However, at $\theta = 0°$ all methods show similar performances.

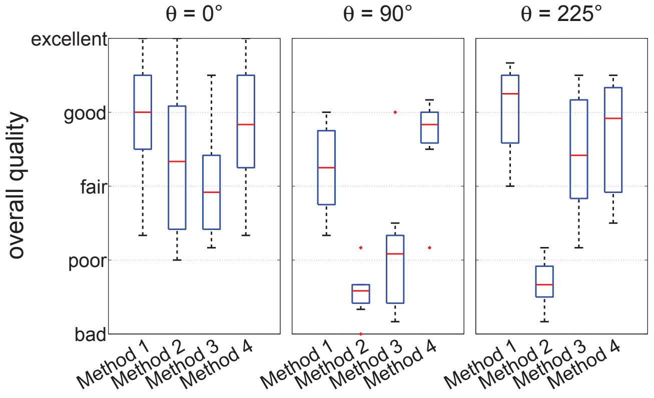

Figure 2: Subjective evaluation of overall quality for different azimuth directions ($\theta = 0°$, 90° and 225°). All methods shown here use linear interpolation.

Conclusions

The comparison between method 1 and 4 indicates that the phase linearization in method 4 may reduce destructive interferences caused by noisy phases of adjacent HRTFs bins, which is evident for method 1. The listening experiment indirectly confirms no significant audible artifacts of phase linearization for frequencies above 1500 Hz. Further, the results imply that the HRTF interpolation by PCA works well for the magnitudes, but the minimum-phase reconstruction introduces large IPD errors, which may explain the poor perceptual overall quality of this method. The results of the objective evaluation and the listening experiment indicate that the overall quality is correlated to an appropriate re-synthesis of original IPDs at lower frequencies. This cue seems to be very important for an authentic perception of virtual sound sources over headphones. However, the results in fig. 1 are based on KEMAR HRTFs. For a more exact examination of the influence by spectral distortion, individual HRTFs need to be considered.

References

[1] D. J. Kistler, F. L. Wightman, 1992, *A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction*, J. Acoust. Soc. Am. 91 (3), 1637-1647

[2] R. G. Klumpp, H. R. Eady, 1956, *Some measurements of interaural time difference thresholds*, J. Acoust. Soc. Am. 28 (5), 859-860

[3] P. Minnaar, J. Plogsties, F. Christensen, 2005, *Directional resolution of head-related transfer functions required in binaural synthesis*, The Journal of the Audio Engineering Society, 53(10), 919-929

[4] E. Rasumow, M. Blau, M. Hansen, S. van de Par, S. Doclo, V. Mellert, and D. Pueschel, 2014, *Smoothing individual head- related transfer functions in the frequency and spatial domains*, J. Acoust. Soc. Am. 135 (4), 2012 - 2025

[5] E. Rasumow, M. Blau, M. Hansen, S. van de Par, S. Doclo, V. Mellert, and D. Pueschel, 2012, *Smoothing head-related transfer functions for a virtual artificial head*, Proceedings of the Acoustics, Nantes Conference, 1025-1030