Time-domain interpolation of head-related transfer functions with correct reproduction of notch frequencies

Julian Bernhard, Gabriel Gomez, Bernhard Seeber

Audio Information Processing, TU München, Arcisstr. 21, 80333 Munich
Email: seeber@tum.de

Summary and Introduction

Realistic synthesis in virtual binaural displays requires interpolation of the complex frequency and phase responses of head-related transfer functions (HRTFs) to correctly reproduce the acoustical impact of small changes in sound source position. State-of-the-art interpolation approaches are classified into time and frequency domain interpolation. Frequency domain interpolation cannot reproduce the direction specific position of spectral notches in HRTFs leading to interpolation errors due to a flattening of filter notches, and it has difficulties reconstructing the phase. Time domain interpolation is prone to comb filter artefacts caused by incorrect temporal alignment.

A new HRTF interpolation approach was developed which overcomes these disadvantages by using a local delay to temporally align neighboring head-related-impulse-responses prior to time domain interpolation. The dynamic time warping algorithm is used for aligning the peaks, avoiding the need to estimate ITDs. A numerical analysis of interpolated versus measured HRTFs shows improved reconstruction of HRTF phase and the notch frequency and depth. Residual, infrequent interpolation errors at the contralateral side can be classified into wideband and narrowband deviations. Their effect on monaural and binaural perception was investigated in a listening experiment. Results indicate that large narrowband intensity deviations around notches have a larger perceptual impact than the smaller wideband errors. Overall, the new method achieves improved representation of sound direction particularly for difficult to interpolate contralateral directions and it significantly reduces localization errors due to wrong ITD estimation.

Methods

Warping Interpolation

We assume that spectral notch movements within HRTFs can be associated with temporal periodicity shifts in HRIRs. To detect the temporal resonance shifts between adjacent HRIRs we apply the Dynamic Time Warping (DTW) algorithm which is a well-known time series registration algorithm [1]. It detects correspondences between samples of adjacent HRIRs providing a sample-wise temporal delay. The local delay is used to accurately align adjacent HRIRs prior to time domain interpolation. The locally aligned filters are interpolated in the time domain and the resulting filter is locally displaced by the amount of the interpolated local delays, averaged from the local delays of adjacent filters. Bilinear Interpolation of local delays and aligned filters is based on barycentric coordinates. The above routine replaces the commonly applied global time shift which is critical for time domain interpolation and correct ITD estimation.

Numerical Evaluation

HRTF data of a human subject was split up into sets of computation and analysis HRTFs. The interpolation algorithm was applied to the computation data to estimate the HRTFs at the positions of the analysis data. The spacing of analysis HRTFs was between 6° and 10°. We compared frequency domain interpolation with minimum-phase time domain reconstruction and normal time domain interpolation to the warping interpolation algorithm. The times of arrival required for time and frequency domain interpolation were estimated with leading edge detection.

To evaluate spectral errors, the audio signal was split into bandlimited signals using Gammatone filters. The absolute averaged intensity level difference between interpolated and analysis filters was calculated in each band. The ITD interpolation error of an HRTF pair was estimated by applying first a linear-phase lowpass-filter with cutoff at 1.5 kHz to the analysis and interpolated HRTF pair. The ITD error was calculated from the normalized cross-correlation.

Perceptual Evaluation

The HRTFs with largest interpolation errors obtained with warping and frequency domain interpolation, respectively, were investigated in a same/different task. Subjects heard two stimuli and had to decide if they were the same or different.

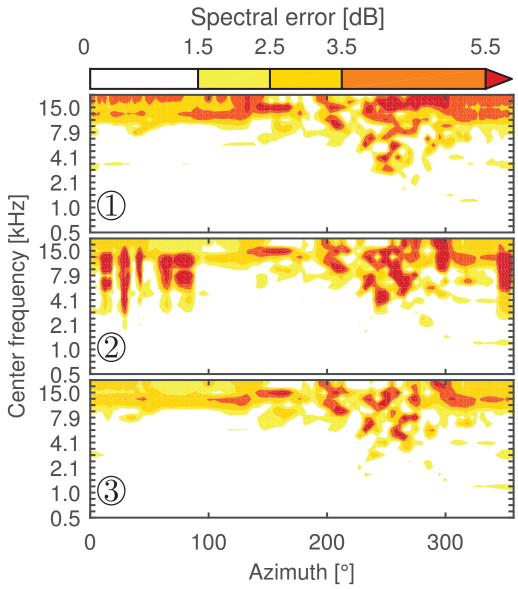

Figure 1: Contour plots of the right ear spectral interpolation errors for (1) warping interpolation, (2) time domain interpolation and (3) frequency domain interpolation. The error was calculated bandwise using a Gammatone filterbank.

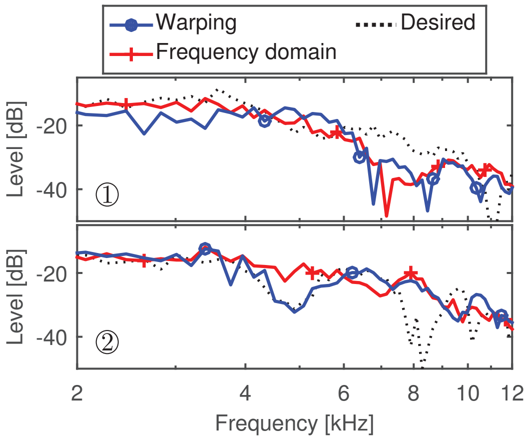

Figure 2: Contralateral spectra with the largest spectral deviation found: (1) for warping interpolation at 100° azimuth for the left ear, (2) for frequency domain interpolation at 270° azimuth for the right ear.

For the monaural test, diotic stimuli were generated by filtering pink noise (100 Hz - 15 kHz, duration 700 ms) with the interpolated or analysis HRTF of the left or right ear. Stimuli for the investigation of binaural perception were pulses of pink noise (100 Hz - 15 kHz, 3 bursts of 250 ms duration with 100 ms pause in-between) which were filtered with the interpolated or analysis HRTF pair to obtain a dichotic signal.

Four test conditions were evaluated, two with “same” stimuli in both stimulus intervals and two with “different” stimuli. These were all possible combinations of analysis and interpolated HRTFs, being AB, BA, AA and BB. In each trial one of these conditions was presented to the subject, randomized over all tested positions and the two algorithms. Levels were identically roved in both intervals from a base level of 60 dB SPL in 5 steps (0, ±2, ±4 dB). Preliminary results of two subjects with normal hearing (<20 dB HL) are presented.

Results and Discussion

Figure 1 shows the spectral interpolation errors. By using a local delay for alignment, comb-filter artefacts were significantly reduced compared to standard time domain interpolation. Particularly at the ipsilateral side, the errors obtained with warping interpolation were reduced to below 1.5 dB. Deviations above 12 kHz are due to spatial aliasing and are perceptually less salient, as filter levels in this frequency region show a relatively low level. At the contralateral side (200° < azimuth < 300°), interpolation errors of frequency domain interpolation and warping interpolation are roughly balanced. At the ipsilateral side both algorithms performed equally well.

For two contralateral HRTFs showing worst interpolation results with warping or frequency domain interpolation, respectively, we investigated the impact of the interpolation errors on binaural perception. Figure 2 shows the interpolated spectra of these contralateral HRTFs. The interpolated HRTF with largest error of warping interpolation was obtained at position 1 where a broadband deviation around 5 dB between 3-6 kHz was observed. The HRTF with largest error of frequency domain interpolation resulted at position 2. The resulting HRTF differs at 5 kHz from the analysis HRTF by about 15 dB. The warping algorithm instead estimated the filter notch well.

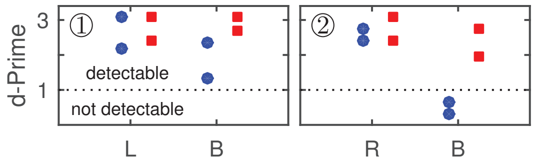

Figure 3: Preliminary results of the listening test: Detection of any perceivable difference is expressed with the sensitivity measure d-Prime for the warping (blue, circles) and frequency domain interpolation (red, squares). The number in each panel (1 or 2) corresponds to the tested contralateral HRTF in the monaural case (L or R) shown in figure 2. Binaural results are labeled with B.

Figure 3 shows the detectability of differences based on the sensitivity measure d-Prime. For position 1, errors from both algorithms were detectable both monaurally and binaurally. For position 2, errors were only monaurally detectable with both algorithms. Due to the correct notch detection with the warping algorithm, errors were rendered inaudible in the binaural case. This indicates that the inaccurate notch interpolation at the contralateral side lead to audible artefacts with higher relevance for binaural perception than the wideband errors occurring with frequency domain interpolation. We therefore assume that binaural perception is improved when warping interpolation is used in virtual auditory displays.

Table 1 compares the resulting ITD errors. For leading edge detection with time domain interpolation and minimum phase reconstruction, the average ITD deviation was above the perceivable low-frequency group delay limit of 30 µs [2] for some spatial regions. In contrast, at all spatial regions warping interpolation achieved an average ITD error below the detectable low-frequency group delay limit. Thus, localization errors due to wrong ITD estimation should be significantly reduced with our approach.

Table 1: ITD errors for warping (W), time (T) or minimum phase (MP) algorithms. Deviations were averaged over four spatial regions. Each region comprised a range of 90° azimuth.

	Front			Back		
	W	T	MP	W	T	MP
ITD Error [µs]	18.0	24.5	66.8	20.6	47.4	45.6

	Right			Left		
	W	T	MP	W	T	MP
ITD Error [µs]	22.0	14.9	48.4	24.3	38.4	75.9

Literature

[1] H. Sakoe *et al*, “Dynamic programming algorithm optimization for spoken word recognition,” *Acoustics, Speech and Signal Processing, IEEE Transactions on,* vol. 26, no. 1, pp. 43-49, Feb, 1978.

[2] J. Plogsties *et al.*, "Audibility of All-Pass Components in Head-Related Transfer Functions," *Audio Engineering Society Convention 108*, 2000.