Investigation on Localization Performance Using Smoothed Individual Head-Related Transfer Functions

Janina Fels, Ramona Bomhardt, Florian Pausch
Institute of Technical Acoustics, Medical Acoustics Group, 52074 Aachen, Germany,
Email: janina.fels@akustik.rwth-aachen.de

Introduction

A characteristic feature of head-related transfer functions (HRTFs) are narrow spectral dips which are mainly caused by resonances of the concha. The resulting dynamic range may be problematic in practical applications like crosstalk cancellation systems where inverted HRTFs are used. To avoid excessive load on the loudspeakers, smoothing algorithms are often applied. Due to the fact that smoothing affects the localization performance, especially the monaural cues are modified [5], different spectral and spatial smoothing algorithms are investigated in a localization experiment.

Algorithms

Spectral Smoothing

In frequency domain, smoothing can be realized using fractional octave band filters. Real-valued discrete-time signals feature a complex spectrum with frequency bins in the range of $0 \leq k \leq (\mathrm{N}-1)$, where k is the frequency index. The smoothing operation can be realized as circular convolution of the magnitude spectrum $|H(k)|$ and a window $W(k)$, which yields a smoothed version $H_{sm}(k)$ of the HRTF magnitude spectrum while preserving the original phase spectrum [3]:

$$H_{sm}(k) = \sum_{i=0}^{\mathrm{N}-1} W(m(k), i) \cdot \left|H[(k-i) \bmod \mathrm{N}]\right| \cdot e^{\mathrm{j}\arg\{H(k)\}} \tag{1}$$

The smoothing index $m(k)$ is defined by the quality factor $Q(k)$, which is dependent on the octave fraction n:

$$m(k) = \left\lfloor \frac{1}{2} \frac{k}{Q(n)} \right\rfloor \tag{2}$$

and the quality factor $Q(n)$:

$$Q(n) = \frac{1}{2^{1/(2n)} - 0.5^{1/(2n)}}. \tag{3}$$

Spatial Smoothing

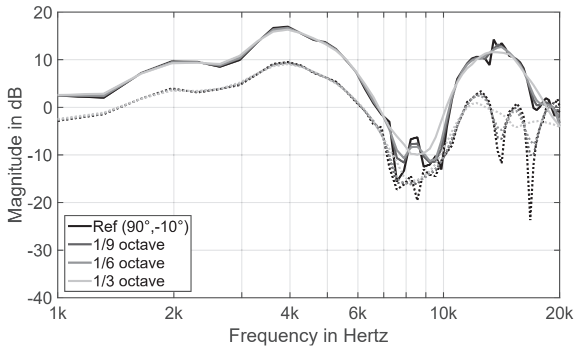
In the spatial domain, an HRTF set measured at discrete positions can be expanded into a finite set of spherical harmonics (SH). In matrix notation, this expansion can be written as $\mathbf{h}_\mathrm{L} = \mathbf{Y}_\mathrm{L}\mathbf{c}_\mathrm{L}$, where $\mathbf{h}_\mathrm{L}$ contains the Fourier-transformed decibel magnitude HRTF spectra, $\mathbf{Y}_\mathrm{L}$ are the SH basis functions up to order L, which is dependent on the measurement grid sampling, and $\mathbf{c}_\mathrm{L}$ are the SH coefficients [2]. Consequently, the SH coefficients are calculated by a matrix inversion

$$\hat{\mathbf{c}}_\mathrm{L} = \left(\mathbf{Y}_\mathrm{L}^\mathrm{H}\mathbf{W}\mathbf{Y}_\mathrm{L} + \kappa\mathbf{D}\right)^{-1}\mathbf{Y}_\mathrm{L}^\mathrm{H}\mathbf{W}\mathbf{h} \tag{4}$$

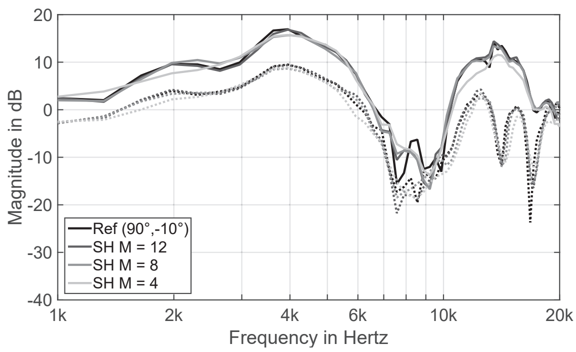
with the weighting matrix $\mathbf{W}$ containing the Voronoi weights associated with each measurement point, the decomposition order-dependent regularization matrix $\mathbf{D}$ and the regularization coefficient κ. To obtain spatially smoothed HRTFs, the first $(\mathrm{M}+1)^2$ SH from $\mathbf{Y}_\mathrm{L}$ and $\hat{\mathbf{c}}_\mathrm{L}$ are taken where M is the SH synthesis order, and multiplied, using the original HRTF phase spectra:

$$\hat{\mathbf{h}} = \mathbf{Y}_\mathrm{M}\hat{\mathbf{c}}_\mathrm{M}. \tag{5}$$

Figure 1 shows the magnitude spectra of individually measured HRTFs at $\vartheta = 90°$ (zenith), $\varphi = -10°$ (azimuth, mathematically positively defined horizontal plane) filtered by different smoothing algorithms for the left ear (solid line) and right ear (dashed line).



(a) Spectral smoothing with $n = 9, 6, 3$.



(b) Spatial smoothing with $\mathrm{M} = 12, 8, 4$.

Figure 1: Influence of different smoothing algorithms on the magnitude spectra of individually measured HRTF data sets at $\vartheta = 90°$ (zenith angle) and $\varphi = -10°$ (azimuth angle).

Listening Experiment

The influence of smoothing on localization performance is subjectively evaluated in a listening experiment using the proximal pointing method where the subject points with a stick in the direction of a virtual sound source (VSS) with reference to the head center [3].
The proximal pointing method is realized using an optical tracking system which tracks the position of a rigid body mounted on headphones and of another rigid body mounted on the pointing device which is held in the right hand, as depicted in Figure 2. The subjects confirm the perceived sound direction by pressing a button on a presenter in the left hand.

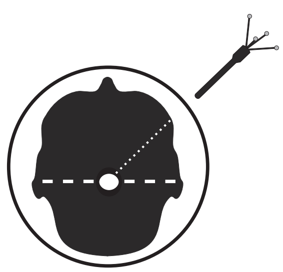


Figure 2: Proximal pointing method, realized using an optical tracking system.

The VSSs are generated binaurally over headphones using individually measured HRTFs, after filtering the playback by an inverse averaged headphone transfer function of all participants [4]. Four directions in the horizontal plane at -10°, -30°, -60° and -80° are tested eight times per direction in seven conditions: (1) no smoothing, (2)-(4) spectral smoothing with octave fraction n = 9, 6, 3, and (5)-(7) spatial smoothing with SH synthesis order M = 12, 8, 4 with $\kappa = 0.01$. In total, 12 male and 4 female subjects at the age of 30 ± 4 (mean ± std) participated in the experiment.

Results and Discussion

For the evaluation of the influence of the smoothing on the localization performance in the horizontal plane, the signed error is evaluated (see Figure 3). In general, the error of the reference results (no smoothing applied on HRTF data sets) lies between 2° and 6° which is in line with the investigations of Bahu et al. [3].
The evaluation of the error for the smoothed HRTF dataset shows that the larger the filter bandwidth and the smaller the SH synthesis order, the more flat are the notches, cf. Figure 1, which results in a significantly worse localization performance for frontal directions due to monaural cue dependency [5], as depicted in Figure 3. In case of spectral smoothing, it can also be seen that the zenith error rises for a larger filter bandwidth for frontal directions. Due to the large interquartile ranges (gray area), particularly for a third-octave band filter, it is obvious that most of the subjects have problems to localize accurately; compare directions $\varphi = -10°$ and $-80°$). This can be traced back to influences of spectral smoothing with a large octave fraction n.
The results for spatial smoothing show larger fluctuations for frontal directions compared to the results for spectral smoothing whereas the median of the zenith error is similar. However, for a low SH synthesis order M, the zenith error increases especially for the azimuth directions of $-30°$ and $-60°$.

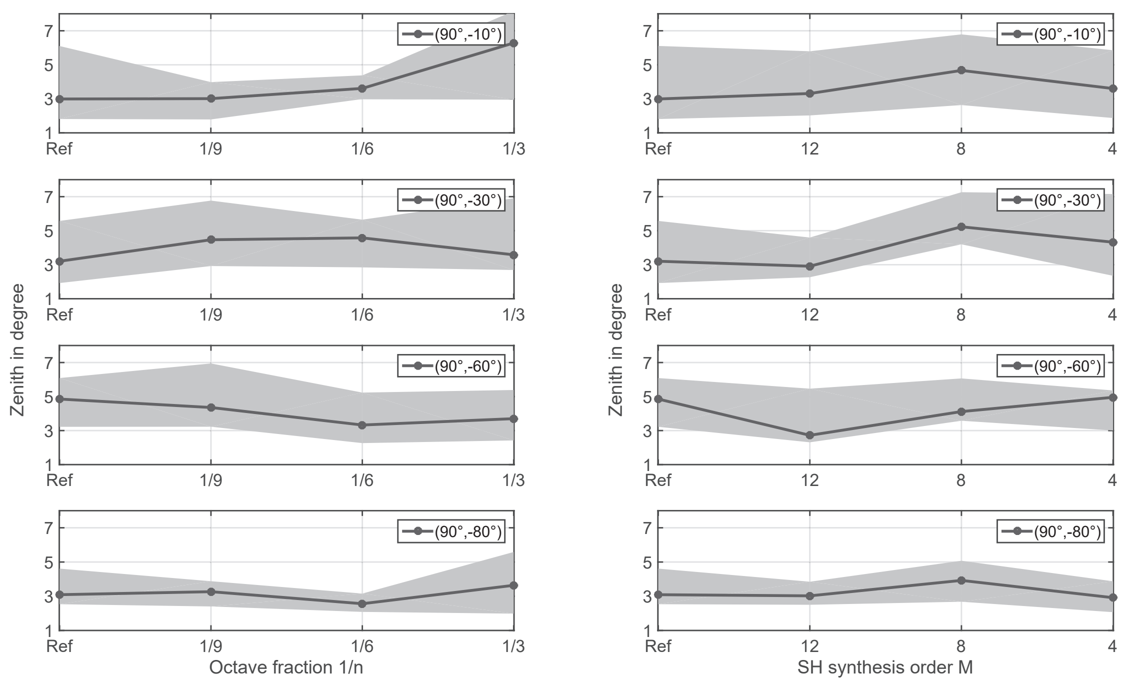


Figure 3: Influence of spectral (left) and spatial (right) smoothing on perceived elevation close to the median plane. The medians (solid line) show the localization precision for the tested directions $\varphi = -10°, -30°, -60°$ and $-80°$. The precision is defined as the standard deviation of the perceived directions. The gray areas display the 25^{th} and 75^{th} quartile.

Additionally, the interquartile range for the zenith error for lateral directions in Figure 3 is much smaller than for frontal directions. This indicates that the localization of virtual sound sources located at lateral directions was easier and less affected by the smoothing algorithms than all other directions.

Conclusion

To conclude, all smoothing methods which are investigated here, can lead to significant changes in the perceived location of a virtual sound source. However, the perceived azimuth directions are less affected than the zenith directions. Furthermore, a lot of subjects reported in-head localizations for small M or large n, respectively. As only four gradations for the spectral and spatial smoothing are assessed, no general tendency with respect to localization accuracy can be observed. Therefore, further investigations are necessary.

References

[1] Hatziantoniou, P. D., Mourjopoulos, J. N.: Generalized Fractional-octave Smoothing of Audio and Acoustic Responses, Journal of the Audio Engineering Society (2000), 48, 259-280.

[2] Pollow, M. et al.: Calculation of Head-Related Transfer Functions for Arbitrary Field Points Using Spherical Harmonics Decomposition, Acta Acustica united with Acustica (2012), 98, 72-82.

[3] Bahu, H., Carpentier, T., Noisternig, M., Warusfel, O.: Comparison of Different Egocentric Pointing Methods for 3D Sound Localization Experiments, Acta Acustica united with Acustica (2016), 102, 107-118.

[4] Masiero, B., Fels, J.: Perceptually Robust Headphone Equalization for Binaural Reproduction, AES Convention 130, London (2011).

[5] Asano, F., Suzuki, Y., Sone, T.: Role of spectral cues in median plane localization. In: The Journal of the Acoustical Society of America 88 (1)(1990), S. 159–168.