PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

# Spatial upsampling of individual sparse head-related transfer function sets by directional equalization

Christoph PÖRSCHMANN[(1)†], Johannes M. AREND[(1)(2)], Fabian BRINKMANN[(2)]

[(1)]Institute of Communications Engineering, TH Köln, D-50679 Cologne, Germany,
[(2)]Audio Communication Group, TU Berlin, D-10587 Berlin, Germany
[†]Corresponding author, E-mail: christoph.poerschmann@th-koeln.de

**Abstract**

Determining full-spherical individual sets of head-related transfer functions (HRTFs) based on sparse measurements is a prerequisite for various applications in virtual acoustics. However, when applying HRTF interpolation in the spatially continuous spherical harmonics (SH) domain, the number of measured HRTFs limits the maximal accessible SH order. This results in a restricted spatial resolution and can cause perceptual artefacts like coloration or localization errors. In a previous publication we presented the SUpDEq method (Spatial Upsampling by Directional Equalization), which reduces these artifacts by a directional equalization based on a spherical head model prior to the SH transform. This removes direction-dependent temporal and spectral components and thus reduces the spatial complexity of the HRTF set enabling improved interpolation of HRTFs already at low SH orders. A subsequent de-equalization recovers energy in higher spatial orders that was discarded in the sparse HRTF set. In this study we analyze 96 individual HRTF sets and investigate to what extent the performance of SUpDEq, which we already analyzed for dummy heads, can be transferred to individual HRTF sets. The results show that the SUpDEq method clearly outperforms common SH interpolation of individual HRTFs with respect to the spectral structure and to modeled localization performance.

Keywords: Binaural hearing, Localization, Head-related transfer functions, Virtual acoustic environments

## 1 INTRODUCTION

A spatial presentation of sound sources is a fundamental element of virtual acoustic environments (VAEs). For this, monaural and binaural cues, which are mainly caused by the shape of the pinna and the head, need to be considered. While spectral information serves as main cue to determine elevation, differences between the signals reaching the left and the right ear allow lateral localization. These binaural cues manifest in interaural time differences (ITDs) and interaural level differences (ILDs). In many headphone-based VAEs, head-related transfer functions (HRTFs) are applied to describe the sound incidence from a source, which is typically in the far-field, to the left and right ear incorporating both, monaural and the binaural cues. Generally, the use of individual HRTFs is advantageous, for example regarding localization accuracy in the median plane [5]. However, a high number of HRTFs is required to adequately capture the relevant cues for all directions of incidence which makes the measurements time-consuming and tedious.

To allow an optimized interpolation between the measured directions, complete sets of HRTFs can be measured on a spherical grid and described in the spherical harmonics (SH) domain [14, 12]. In this case a decomposition into spherical base functions of different spatial orders $N$ is applied, where higher orders correspond to a higher spatial resolution. A subsequent inverse spatial Fourier transform at arbitrary angles can be used to recover a spatially upsampled HRTF set. However, describing sparse HRTF sets in the SH domain results in a limited spatial order and incorporates an incomplete description of the spatial properties resulting in spatial aliasing or truncation errors. To avoid spatial alisaing, an order $N \geq kr$ with $k = \omega/c$, and $r$ being the head radius is required [11, 4]. For the full audio bandwidth ($f \leq 20\,kHz$) this leads to $N = 32$ requiring at least 1089 measured directions when assuming $r = 8.75$ cm and $c = 343$ m/s.

Different studies analyzed the artifacts of sparsely measured HRTF sets or examined methods to reduce them

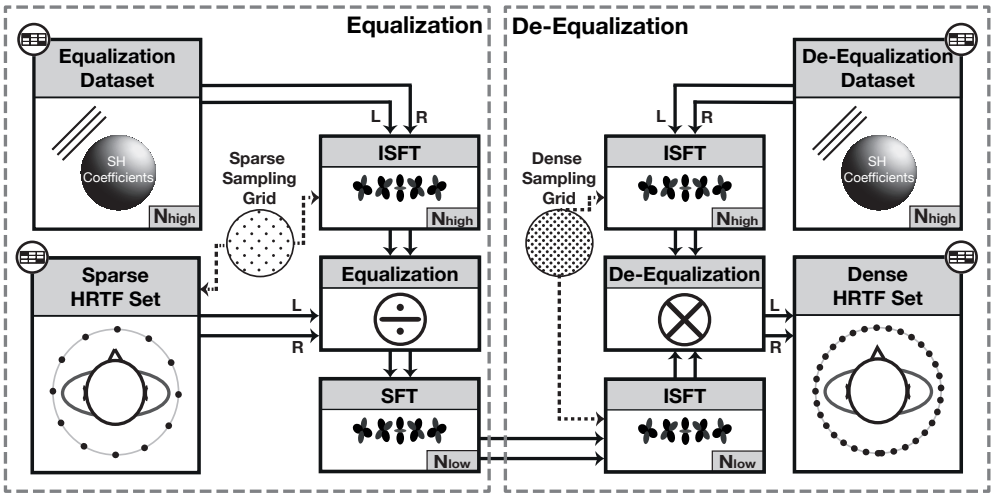


Figure 1. Block diagram of the SUpDEq method. Left panel: A sparse HRTF set is equalized on the corresponding sparse sampling grid before transformed to the SH domain with $N = N_{low}$. Right panel: The equalized set is de-equalized on a dense sampling grid. If required, the resulting dense HRTF set can again be transformed to the SH domain with $N = N_{high}$.

(e.g. [4, 3, 15, 7]). In this scope we recently introduced the SUpDEq (Spatial Upsampling by Directional Equalization) method [10], which removes frequency-dependent ITDs and ILDs as well as head-related elevation-dependent spectral features from the HRTFs. SUpDEq applies a spectral division (equalization) of the HRTF with a corresponding equalization function prior to the SH transform. A directional rigid sphere transfer function can be used here as equalization function, resulting in a significantly reduced spatial order $N$. After spatial upsampling, a de-equalization by means of a spectral multiplication with the same equalization function recovers a spatially upsampled HRTF set. In this paper we analyze the SUpDEq method for a large number of measured and simulated datasets.

## 2 METHOD

The SUpDEq method has been described in detail in [10]. In the following we thus briefly outline the basic concept. The corresponding block diagram is given in Fig. 1. First, the sparse HRTF set $H_{HRTF}$ measured at $S$ sampling points $\Omega_s = \{(\phi_1, \theta_1), \ldots, (\phi_S, \theta_S)\}$ is spatially equalized with an appropriate equalization dataset $H_{EQ}$

$$H_{HRTF,EQ}(\omega, \Omega_s) = \frac{H_{HRTF}(\omega, \Omega_s)}{H_{EQ}(\omega, \Omega_s)}. \tag{1}$$

While generally different equalization datasets can be applied, in this study a rigid sphere transfer function is used [14, p. 227]. The radius of the sphere corresponds to the physical dimensions of a human head, as ear position $\phi = \pm 90^\circ$ and $\theta = 0^\circ$ is considered. The rigid sphere transfer function can thus be regarded as a simplified HRTF set featuring basic temporal and spectral components, but leaving out information on the shape of the outer ears or the fine structure of the head. Thus, by the equalization a time-alignment of the HRTFs is performed and direction-dependent influences of the spherical shape of the head are compensated. As a consequence, the equalization with the rigid sphere transfer function considerably reduces the directional complexity of $H_{HRTF,EQ}$ and thus the required order for the SH transform. As the equalization dataset $H_{EQ}$ can be calculated based on an analytical description, it can be determined at a freely chosen maximal order, typically $N_{high} \geq 35$. The SH coefficients for the equalized sparse HRTF set are obtained by applying the SH transform

on the equalized HRTFs up to an appropriate low maximal order $N_{low}$ which corresponds to the maximal order that can be resolved by $\Omega_s$. Then an upsampled HRTF set $\widehat{H}_{HRTF,EQ}$ is calculated on a dense sampling grid $\Omega_d = \{(\phi_1, \theta_1), \ldots, (\phi_D, \theta_D)\}$, with $D \gg S$ by using the inverse SH transform. Finally, HRTFs are reconstructed by a subsequent de-equalization by means of spectral multiplication with a de-equalization dataset $H_{DEQ}$

$$\widehat{H}_{HRTF,DEQ}(\omega, \Omega_d) = \widehat{H}_{HRTF,EQ}(\omega, \Omega_d) \cdot H_{DEQ}(\omega, \Omega_d). \tag{2}$$

For de-equalization, again the rigid sphere transfer function is used in the present study. This last step recovers energy at higher spatial orders that was transformed to lower orders within the equalization. Again, $H_{HRTF} = \widehat{H}_{HRTF,DEQ}$ holds if $N_{low}$ and $N_{high}$ are chosen appropriately. Energy which, after the equalization, still is apparent at high modal orders $N > N_{low}$ results in spatial aliasing and truncation errors as it is irreversibly mirrored to lower orders $N \leq N_{low}$ [4]. Thus we obtain $H_{HRTF} \approx \widehat{H}_{HRTF,DEQ}$. The following section analyzes the influence of these deviations for individual datasets and investigates which advantage the SUpDEq method provides compared to common (order-limited) SH interpolation without any pre- or postprocessing.

# 3 EVALUATION

In previous publications [10, 9] we investigated the performance of SUpDEq for different artificial heads. However, one of the target applications of the SUpDEq method is the reduction of the measurement effort of individual HRTF sets. Thus, in this study we analyze the performance of the SUpDEq method for the HUTUBS database which is online available on `http://dx.doi.org/10.14279/depositonce-8487`. The database contains of 96 acoustically measured and 96 numerically simulated datasets of full-spherical HRTFs (94 subjects plus 2 repeated measurements of a human subject and an artificial head). For more detailed information on the database please refer to [6]. We apply the HRTF sets to compare the performance of the SUpDEq method (de-equalized HRTFs) to HRTFs obtained with with strictly order limited SH interpolation, i.e., without any pre- or post-processing before or after the SH transform. For this we generated SH coefficients from 15 sparse sampling grids equaling (limited) orders of $N = 1 - 15$. Thus, both order-limited (OL) and de-equalized (DEQ) sets are based on the same respective sparse grid. To generate various sparse HRTF sets which we used as input data for the evaluation, we simply spatially subsampled each individual reference set in the SH domain by means of the inverse SH transform at the required directions. We calculated the optimal radius for the rigid sphere model for each of the sets according to Algazi et al. [1] based on the individual anthropometry resulting in an average value over the complete set of $r = 9.1\,\text{cm}$ ($SD = 0.23\,\text{cm}$).

## 3.1 Spectral differences

First we analyze the spectral deviations to the reference set as a function of $N$ on various test sampling grids with $T$ sampling points $\Omega_t = \{(\phi_1, \theta_1), \ldots, (\phi_T, \theta_T)\}$. For this the frequency-dependent spectral differences per sampling point were calculated in dB as

$$\Delta g(\omega, \Omega_t) = 20lg \frac{|\, H_{HRTF,REF}(\omega, \Omega_t)\,|}{|\, H_{HRTF,TEST}(\omega, \Omega_t)\,|}, \tag{3}$$

where $H_{HRTF,REF}$ is the left ear HRTF extracted from the reference set and $H_{HRTF,TEST}$ the one extracted from the order-limited or the de-equalized datasets at the sampling point $\Omega_t$. Then, the absolute value of $\Delta g(\omega, \Omega_t)$ was averaged across the temporal frequency $\omega$ to obtain one value $\Delta G_{sp}(\Omega_t)$ (in dB) per sampling point

$$\Delta G_{sp}(\Omega_t) = \frac{1}{n_\omega} \sum_{\omega=1}^{n_\omega} |\, \Delta g(\omega, \Omega_t)\,|, \tag{4}$$

across all sampling points $\Omega_t$ to obtain the frequency-dependent measure $\Delta G_f(\omega)$ (in dB)

$$\Delta G_f(\omega) = \frac{1}{n_{\Omega_t}} \sum_{\Omega_t=1}^{n_{\Omega_t}} |\, \Delta g(\omega, \Omega_t)\,|, \tag{5}$$

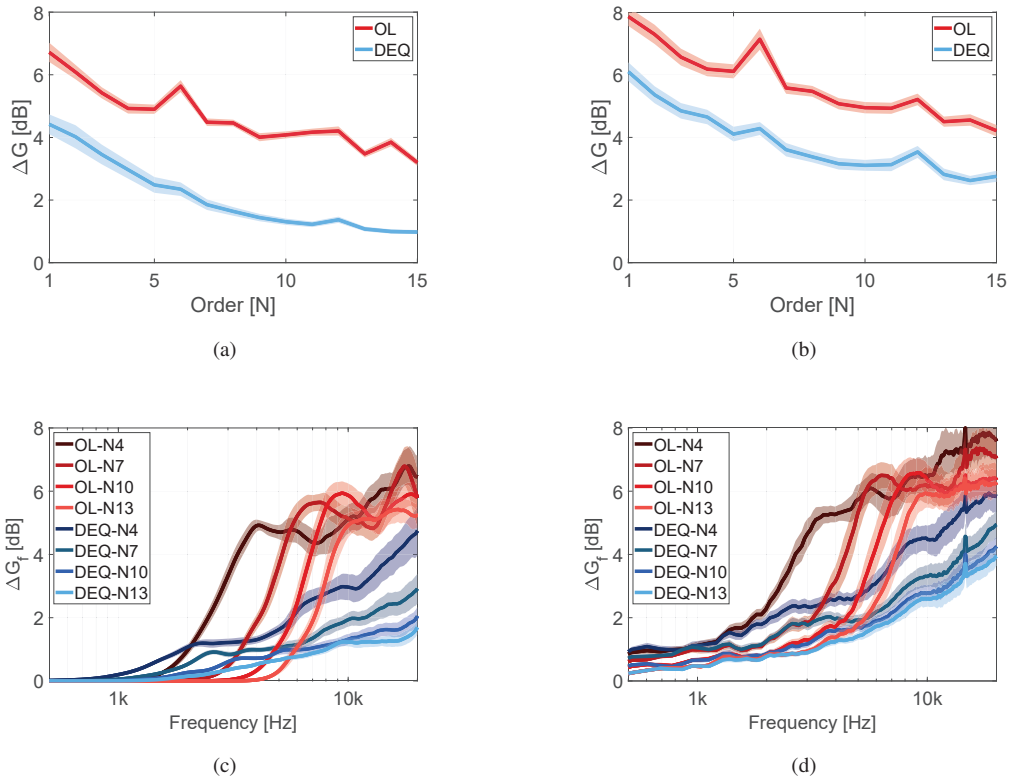


Figure 2. Spectral differences in dB (left ear) between the reference HRTF sets and the order-limited (OL) or de-equalized HRTF sets (DEQ), both based on the respective sparse set, averaged over all 96 datasets. Additionally the standard deviations are plotted (shaded). The left row (a,c) illustrates the results for the simulated datasets, in (b,d) the ones for the measured datasets are given. In (a,b) the spectral differences $\Delta G$ averaged over the full audio bandwidth across $N$ for order-limited datasets (red) and the de-equalized datasets (blue) are given, in (c,d) the frequency-dependent spectral differences $\Delta G_f(\omega)$ for $N = 4, 7, 10, 13$ (color saturation).

and across $\omega$ and $\Omega_t$, resulting in a single value $\Delta G$ (in dB) describing the spectral difference

$$\Delta G = \frac{1}{n_{\Omega_t}} \frac{1}{n_\omega} \sum_{\Omega_t=1}^{n_{\Omega_t}} \sum_{\omega=1}^{n_\omega} | \Delta g(\omega, \Omega_t) | . \tag{6}$$

Finally, the average values and standard deviations over all 96 datasets were calculated for the simulated and the measured datasets.

Fig. 2 (a,b) show the spectral differences $\Delta G$ across $N$ for order-limited interpolation and the SUpDEq method (de-equalized datasets) over the full audio bandwidth using the reference Lebedev$_{2702}$ grid as test sampling grid $\Omega_t$. The SUpDEq method clearly outperforms the order-limited interpolation both for the simulated and the mesaured HRTF sets. The spectral differences are about $2-3$ dB lower than for order-limited interpolation. Fig. 2 (c,d) show the frequency-dependent spectral differences $\Delta G_f(\omega)$ at $N = 4, 7, 10, 13$. Generally, the spectral differences are quite small at low frequencies. For order-limited interpolation they suddenly rise within one octave from about 2 dB up to about 5 dB or more above a specific alias frequency. For the SUpDEq method, however, the spectral differences show a much more gentle rise. The differences exceed 2 dB for frequencies

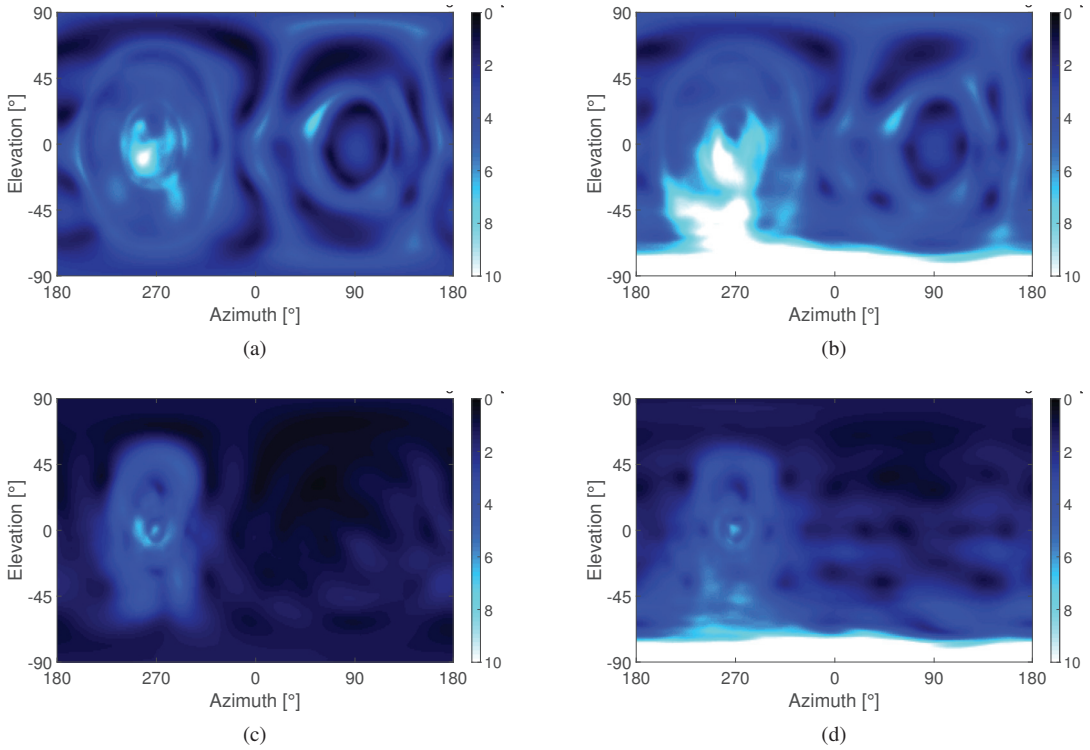


Figure 3. Spectral differences $\Delta G_{sp}(\Omega_t)$ per sampling point for order-limited interpolation (a,b) and for the SUpDEq method (c,d) at $N = 4$ and $f \leq 10\,\text{kHz}$ averaged over all 96 datasets. The left row (a,c) shows the results for the simulated datasets, the right row (b,d) the results for the measured ones.

above 3 kHz for $N = 4$, while differences stay below 2 dB for orders of $N \geq 10$ up to 10 kHz (DEQ).

Fig. 3 concludes the spectral analysis and shows the spectral differences $\Delta G_{sp}(\Omega_t)$ per sampling point at $N = 4$, $f \leq 10\,\text{kHz}$, and a full spherical test sampling grid $\Omega_t$ with a resolution of $1^\circ$ in azimuth and elevation. As depicted in Fig. 3 (a,b), the order-limited interpolation results in distinct spectral differences spread over the entire angular range. On the contrary, Fig. 3 (c,d) shows that for the SUpDEq method the spectral differences are mainly located at contralateral directions. At frontal directions, where order-limited interpolation typically performs badly, the SUpDEq method shows good results. The same can be observed for various ipsilateral directions. The spectral differences are generally higher for order-limited interpolation, with a maximum of about $\Delta G_{sp}(\Omega_t) = 10.4\,\text{dB}$ at $\phi = 262^\circ$ and $\theta = -10^\circ$ averaged over all subjects for the simulated datasets. For these datasets applying the SUpDEq method results in a maximal spectral difference $\Delta G_{sp}(\Omega_t)$ of 6.6 dB at $\phi = 257^\circ$ and $\theta = 2^\circ$. Finally, Fig. 3 (b,d) show the same trend for the measured datasets, but reveal large deviations for the downward directions. This is caused by the acoustic shadowing of the measurement equipment and is described in detail in [6].

### 3.2 Localization performance

To compare the localization performance of order-limited HRTFs and de-equalized HRTFs in the median sagittal plane, we used the model from Baumgartner et al. [2] which compares the spectral structure of a reference HRTF set to a set of test HRTFs. Based on a probabilistic estimate of the perceived sound source location, the model determines the polar RMS error which describes the expected angular error between the actual and

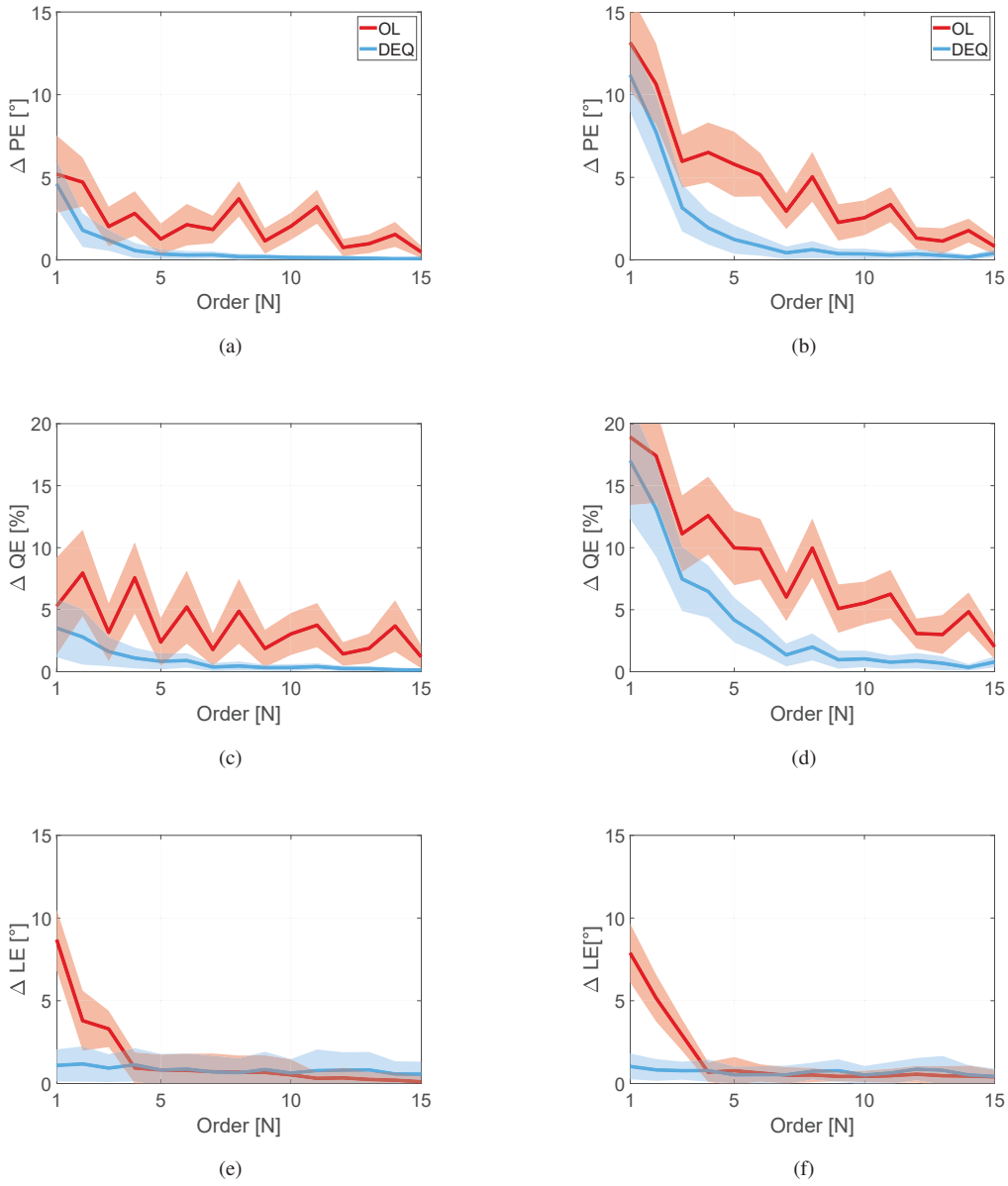


Figure 4. Absolute polar error difference $\Delta PE$ (a,b), quadrant error difference $\Delta QE$ (c,d), and lateral error difference $\Delta LE$ (e,f) over SH order $N$ for order-limited interpolation (red) and the SUpDEq method (blue) averaged over the 96 individual datasets. Additionally, the standard deviations are shown (shaded). In the left row (a,c,e) the results for the simulated HRTF sets are shown, in the right row (b,d,f) the results for the measured HRTF sets.

perceived source positions. Additionally, it determines the quadrant error rate which specifies the front-back and up-down confusions. Regarding the localization performance in the horizontal plane, we used the model from May et al. [8] which weighs the frequency-dependent binaural cues (ILDs, ITDs) to estimate the azimuthal

position of a sound source. A lateral error can be calculated by comparing the intended and the estimated source position. For the analysis of both models we used the Auditory Modeling Toolbox (AMT) [13]. The procedure for determining the errors has been described in detail in [10] and can be outlined as follows. To estimate median sagittal plane localization performance, we used a test sampling grid $\Omega_t$ with $\phi = \{0°, 180°\}$ and $-30° \leq \theta \leq 90°$ in steps of 1°, and assumed a median listener sensitivity of $S = 0.76$ (according to Baumgartner et al. [2]). For the horizontal plane localization performance, we used a test sampling grid with $\phi = \pm 90°$ in steps of 5°. We determined the absolute polar error difference (PE in degree)

$$\Delta PE = \mid PE_{REF} - PE_{TEST} \mid, \tag{7}$$

the absolute quadrant error difference (QE in percent)

$$\Delta QE = \mid QE_{REF} - QE_{TEST} \mid, \tag{8}$$

as well as the absolute lateral error difference (LE in degree)

$$\Delta LE = \frac{1}{T} \sum_{t=1}^{T} \mid LE_{REF}(\Omega_t) - LE_{TEST}(\Omega_t) \mid, \tag{9}$$

for each order $N$ with the subscripts *REF* describing the referene dataset and *TEST* the dataset under test. Again we calculated the averages and standard deviations over all datasets separated for the simulated and the measured sets.

As plotted in Fig. 4 (a–d), in the median sagittal plane the order-limited interpolation leads both for the simulated and the measured datasets to higher errors than the SUpDEq method. High-frequency deviations of the order-limited HRTFs affect spectral cues which are relevant for sagittal plane localization. For the de-equalized datasets, $\Delta PE$ decreases with increasing order $N$, $\Delta PE \leq 2°$ holds for $N \geq 4$. Thus the spectral cues seem to be mostly unimpaired here. The extent of the quadrant error $\Delta QE$ varies greatly between the measured and the simulated sets and lies for the order-limited sets between 4 % (simulated) and 10 % (measured) at $N \geq 7$. However, for the de-equalized datasets, $\Delta QE$ is below 2 % at $N \geq 7$. Generally, in the median sagittal plane the average errors are much higher for the measured datasets than for the simulated ones. This is probably a result of the measurement inaccuracies for downward directions, which as well have been observed in Sec. 3. In Fig. 4 (e–f) the localization performance in the horizontal plane is shown. Here the order-limited interpolation performs quite well, even though lateral errors are distinctly amplified at orders $N \leq 3$. This might be caused by strong pre-ringing artifacts causing wrong ITDs, as already discussed in [10]. The SUpDEq method leads to hardly any increase in lateral error over the entire tested range of $N$.

## 4 CONCLUSION

In this paper we analyzed the performance of the SUpDEq method for spatial upsampling of individual sparse HRTF sets. Regarding the spectral structure, the deviations from the reference HRTF set are significantly smaller for the SUpDEq method than for order-limited interpolation. The average difference is about 2 dB, both for the simulated and the measured datasets. Furthermore, the analysis of the spectral diffences showed for the SUpDEq methods a much more gentle rise over frequency than for the order-limited interpolation. Finally, the spectral differences induced by the SUpDEq method are mainly at contralateral directions, while the differences due to order-limited interpolation spread over the entire angular range, with distinct clusters at frontal and contralateral directions. Regarding the modeled localization performance the SUpDEq method performed better in both planes because spectral and binaural cues are less impaired in comparison to the order-limited interpolation.

Generally, the evaluation showed that the results found for dummy heads in [10] can be generalized to individually measured or simulated datasets. Thus, the SUpDEq approach can help closing the gap between a practical and fast measurement procedure and sufficient accuracy of the upsampled HRTF set. However, for such a simplified procedure other influencing factors like e.g. the elimination of room reflections [9] or the compensation of small displacements of the human head during the measurement need to be considered.

The research presented in this paper has been funded by the German Federal Ministry of Education and Research. Support Code: BMBF 03FH014IX5-NarDasS. A Matlab-based implementation of the SUpDEq method is available on `https://github.com/AudioGroupCologne/SUpDEq`.

## REFERENCES

[1] V. Algazi, C. Avendano, and R. O. Duda. Estimation of a Spherical-Head Model from Anthropometry. *J. Audio Eng. Soc.*, 49(6):472 – 479, 2001.

[2] R. Baumgartner, P. Majdak, and B. Laback. Modeling sound-source localization in sagittal planes for human listeners. *J. Acous. Soc. Am.*, 136(2):791–802, 2014.

[3] Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, and B. Rafaely. Spectral equalization in binaural signals represented by order-truncated spherical harmonics. *The Journal of the Acoustical Society of America*, 141(6):4087–4096, 2017.

[4] B. Bernschütz, A. Vázquez Giner, C. Pörschmann, and J. M. Arend. Binaural reproduction of plane waves with reduced modal order. *Acta Acustica united with Acustica*, 100(5):972–983, 2014.

[5] J. Blauert. *Spatial Hearing - The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA, revised edition, 1996.

[6] F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl. A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses. *Journal of the Audio Engineering Society, in press*, 2019.

[7] F. Brinkmann and S. Weinzierl. Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition. In *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality*, pages 1–10, 2018.

[8] T. May, S. Van De Par, and A. Kohlrausch. A probabilistic model for robust localization based on a binaural auditory front-end. *IEEE Trans. Audio, Speech, Lang. Process.*, 19(1):1–13, 2011.

[9] C. Pörschmann and J. M. Arend. Obtaining Dense HRTF Sets from Sparse Measurements in Reverberant Environments. In *Proceedings of the AES Conference on Immersive and Interactive Audio*, 2019.

[10] C. Pörschmann, J. M. Arend, and F. Brinkmann. Directional Equalization of Sparse Head-Related Transfer Function Sets for Spatial Upsampling. *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, 27(6):1060 – 1071, 2019.

[11] B. Rafaely. Analysis and Design of Spherical Microphone Arrays. *IEEE Transaction on Speech and Audio Processing*, 13(1):135–143, 2005.

[12] B. Rafaely. *Fundamentals of Spherical Array Processing*. Springer-Verlag, Berlin Heidelberg, 2015.

[13] P. Søndergaard and P. Majdak. The Auditory Modeling Toolbox. In J. Blauert, editor, *The Technology of Binaural Listening*, pages 33–56. Springer-Verlag, Berlin Heidelberg, 2013.

[14] E. G. Williams. *Fourier Acoustics - Sound Radiation and Nearfield Acoustical Holography*. Academic Press, London, UK, 1999.

[15] M. Zaunschirm, C. Schoerkhuber, and R. Hoeldrich. Binaural rendering of Ambisonic signals by HRIR time alignment and a diffuseness constraint. *J. Acous. Soc. Am.*, 143(6):3616 – 3627, 2018.