Synthesis of Near-Field HRTFs by Directional Equalization of Far-Field Datasets

Johannes M. Arend[1,2], Christoph Pörschmann[1]

[1] *Institute of Communications Engineering, TH Köln, D-50679 Cologne, Germany*

[2] *Audio Communication Group, TU Berlin, D-10587 Berlin, Germany*

Email: johannes.arend@th-koeln.de

Introduction

Headphone-based spatial audio rendering usually relies on a high-quality set of head-related transfer functions (HRTFs). Such a set of HRTFs describes the sound reaching a listener from various directions and is typically measured in the far-field on a spherical sampling grid. As HRTFs are virtually independent of sound source distance in the far-field, the distance of a virtual sound source can be easily varied by adjusting its level according to the inverse-square law. However, in the near-field, i.e. at distances less than 1 m [1], HRTFs vary significantly as a function of distance. Thus, interaural level differences (ILDs) increase substantially as a sound source approaches the head, whereas interaural time differences (ITDs) vary only slightly [1, 2]. Furthermore, diffraction and head-shadowing effects lead to a low-pass filtering character of a nearby sound source approaching the head [1, 2]. Last, in the near-field, the angle between source and ear differs from the angle between source and head center, resulting in acoustic parallax effects changing systematically with source distance [1].

In view of all these effects, it is evident that near-field HRTFs are necessary for proper reproduction of nearby sound sources. However, measuring full-spherical HRTF sets at numerous near-field distances is a time consuming and also technically difficult task (for more details see e.g. [1, 2, 3]), which is why it seems desirable to synthesize near-field HRTFs from a well-measured set of far-field HRTFs. A common synthesis approach is to apply a so-called distance variation function (DVF) to far-field data [3]. Such a DVF basically describes the change of an HRTF when a sound source shifts in distance, and can be obtained (for a particular direction) using analytical solutions of a rigid sphere model [4]. However, to our knowledge, none of the DVF methods presented so far account for (high-frequency) parallax effects, thereby neglecting characteristics of nearby sound sources [3]. Moreover, even though the DVF method could be extended with a simple parallax model [3], difficulties may occur because DVF methods mostly work with HRTFs from discrete directions, i.e. the far-field HRTF required according to the parallax model might simply not be available for synthesis. Another approach to generate near-field HRTFs is range extrapolation in the spherical harmonics (SH) domain [5]. Here, a full-spherical HRTF set is first transformed to the SH domain by means of a spherical Fourier transform (also called SH transform) [6, p. 16], and the resulting SH coefficients are then multiplied with specific distance-dependent spherical Hankel functions to obtain near-field HRTFs at any desired range. However, as the spherical Hankel functions grow exponentially for higher orders and low frequencies, numerical instabilities and artifacts may arise when applying them without any regularization [5]. To avoid this, the spherical Hankel functions usually require amplitude limiting, which is a rather critical task and again may lead to artifacts [7].

In this work, we present a novel approach to synthesize near-field HRTFs by directional equalization of far-field datasets. The presented approach is a modified version of our recently proposed SUpDEq (Spatial Upsampling by Directional Equalization) method for spatial upsampling of sparse (individual) HRTF sets [8]. In brief, SUpDEq is based on a spectral equalization of the sparse HRTF set with directional rigid sphere transfer functions (STFs), spatial upsampling of this equalized dataset by an inverse SH transform on a dense grid, and a spectral de-equalization of the processed dataset with the same STFs. In the modified implementation presented here, near-field HRTFs are synthesized by applying STFs for a point source in the near field as the de-equalization function instead of STFs for a plane wave in the far field. Thus, as further outlined below, the method combines SH-based processing with the DVF method, thereby enabling near-field HRTF synthesis that is numerically stable and can take acoustic parallax effects into account.

Method

A spherical dataset $H(\omega, \Omega_g)$ can be described in the SH domain by the SH coefficients $f_{nm}(\omega)$ that are obtained via the SH transform [6, p. 16]

$$f_{nm}(\omega) = \sum_{g=1}^{G} H(\omega, \Omega_g) Y_n^m(\Omega_g)^* \beta_g \,, \tag{1}$$

with ω the temporal frequency, β_g the sampling weights, and the G discrete HRTF-angles $\Omega_g = \{(\phi_1, \theta_1), \ldots, (\phi_G, \theta_G)\}$ at azimuth ϕ, and elevation θ. The notation $(\cdot)^*$ denotes complex conjugation, and Y_n^m the spherical harmonics of order n and mode/degree m

$$Y_n^m(\theta, \phi) = \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(cos\theta) e^{im\phi} \,, \tag{2}$$

with the associated Legendre functions P_n^m and $i = \sqrt{-1}$ the imaginary unit. The inverse SH transform can be used to recover H at arbitrary angles

$$\widehat{H}(\omega, \Omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} f_{nm}(\omega) Y_n^m(\Omega) \,, \tag{3}$$

This work was funded by the German Federal Ministry of Education and Research (BMBF 03FH014IX5-NarDasS)

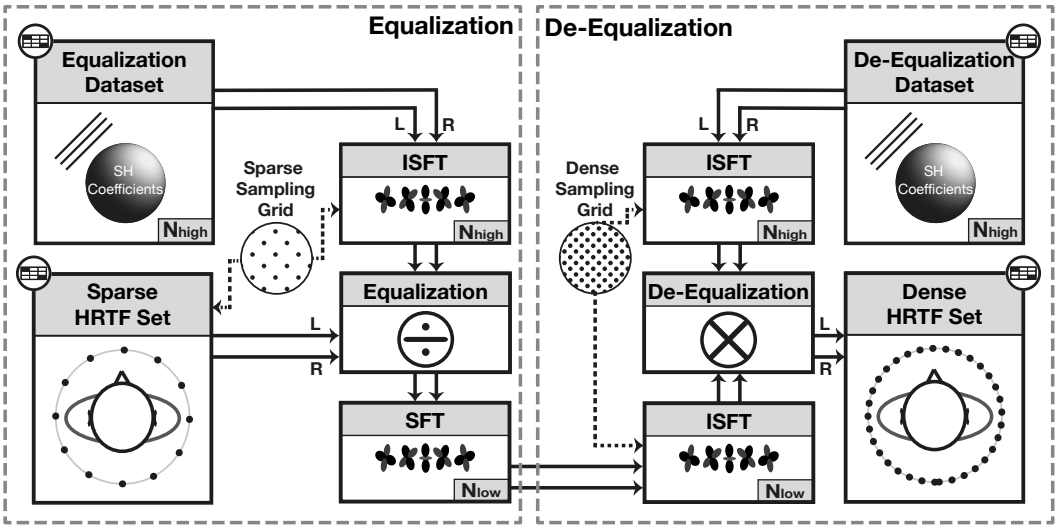

Figure 1: Block diagram of the SUpDEq method. Left panel: A sparse HRTF set is equalized on the corresponding sparse sampling grid. The equalized set is then transformed to the SH domain with $N = N_{low}$. Right panel: The equalized set is de-equalized on a dense sampling grid, resulting in a dense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

where N is the maximal spatial order. If H is order-limited, a sufficient choice of N results in $H = \widehat{H}$.
Depending on the spherical sampling grid Ω_g, the coefficients f_{nm} can be calculated up to a maximum order N. The number of measured directions G directly corresponds to the maximum order N by $G \approx (N+1)^2$. In case the order of H exceeds N, spatial aliasing occurs [6, ch. 3.7]. In this context, an appropriate pre-processing that reduces the spatial complexity of H will directly relax the requirement on G.
The following paragraphs first briefly describe the SUpDEq method according to Fig. 1 (please refer to [8] for more details). After this, the modifications necessary to synthesize parallax-adjusted near-field HRTFs are introduced. Since the processing for the left and right ear signals is identical for the most parts, corresponding subscripts were omitted for ease of display.
In a first step, the sparse HRTF set H_{HRTF} measured at S sampling points $\Omega_s = \{(\phi_1, \theta_1), \ldots, (\phi_S, \theta_S)\}$ is equalized with an appropriate equalization dataset H_{EQ}

$$H_{HRTF,EQ}(\omega, \Omega_s) = \frac{H_{HRTF}(\omega, \Omega_s)}{H_{EQ}(\omega, \Omega_s)}\,. \tag{4}$$

The equalization dataset is intended to remove the directional dependency in H_{HRTF} to a certain degree with the goal to minimize the required order for the SH transform. Different equalization datasets can be applied – throughout this study a rigid sphere transfer function for an incident plane wave is used [6, p. 44]:

$$\begin{aligned} &H_{STF,PW}(\omega, \Omega_g) \\ &= P 4\pi \sum_{n=0}^{N_{high}} \sum_{m=-n}^{n} i^n j_n(kr) Y_n^m(\Omega_e) Y_n^m(\Omega_g)^*, \end{aligned} \tag{5}$$

where k is the wavenumber, j_n the spherical Bessel function of the first kind, and Ω_e the ear position at $\phi = \pm 90°$ and $\theta = 0°$. P denotes an arbitrary sound pressure. The radius r of the sphere should match the physical dimensions of a human head. Except for an angular shift in azimuth, the set is identical for the left and the right ear. As the equalization dataset is based on an analytical description, it can be determined at a freely chosen maximal order, typically, a high order $N_{high} \geq 35$.
In a second step, SH coefficients $f_{EQ,nm}$ for the equalized sparse HRTF set are obtained by applying the SH transform according to Eq. (1) on $H_{HRTF,EQ}(\omega, \Omega_s)$ up to an appropriate low maximal order N_{low}.
In a third step, 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 described by Eq. (3).
In a fourth and final step, HRTFs are reconstructed by a subsequent de-equalization by means of spectral multiplication with a de-equalization dataset H_{DEQ}

$$\widehat{H}_{HRTF,DEQ}(\omega, \Omega_d) = \widehat{H}_{HRTF,EQ}(\omega, \Omega_d) \cdot H_{DEQ}(\omega, \Omega_d)\,, \tag{6}$$

where again $H_{HRTF} = \widehat{H}_{HRTF,DEQ}$ holds if N_{low} and N_{high} are chosen appropriately. In the initial implementation [8], STFs for an incident plane wave as given in Eq. (5) are used for de-equalization. Instead, STFs for an incident spherical wave (point source) at a specific distance r' with $r' > r$ can be used though [6, p. 45]:

$$\begin{aligned} &H_{STF,SW}(\omega, \Omega_g, r') \\ &= P 4\pi i k \sum_{n=0}^{N_{high}} \sum_{m=-n}^{n} j_n(kr) h_n(kr') Y_n^m(\Omega_e) Y_n^m(\Omega_g)^*, \end{aligned} \tag{7}$$

with h_n the spherical Hankel function of the first kind. Applying STFs for a point source at a distance r' (in the near-field) not only recovers energies at higher spatial orders that were transformed to lower orders in the first step, but also results in a distance shift, thereby allowing to synthesize near-field HRTF sets based on (sparse) far-field datasets. Furthermore, to take parallax effects into

account, a simple parallax model adjusting the azimuth and elevation angles of Ω_d with respect to the sound source distance r' individually for each ear position of Ω_e can be applied, leading to Ω_{dp}. Actually, this results in one specific sampling grid Ω_{dp} for each ear, as the adapted azimuth and elevation values now describe the angle between the source and the left and right ear respectively, and not between the source and the center of the head. However, for ease of display, the processing for the left and right ear signals is still combined in the following description.

To finally generate parallax-adjusted near-field HRTFs, the third and fourth steps are carried out involving Ω_{dp} and $H_{STF,SW}$. Thus, an upsampled HRTF set $\widehat{H}_{HRTF,EQP}$ is calculated on a dense parallax-adjusted sampling grid Ω_{dp}, and finally distance-shifted HRTFs are reconstructed by spectral multiplication with a de-equalization dataset H_{DEQ} corresponding to $H_{STF,SW}$:

$$\begin{aligned}&\widehat{H}_{HRTF,DEQP}(\omega,\Omega_d)\\&=\widehat{H}_{HRTF,EQP}(\omega,\Omega_{dp})\cdot H_{DEQ}(\omega,\Omega_d)\,.\end{aligned}\tag{8}$$

where again N_{low} and N_{high} should be chosen appropriately. In fact, Eq. (8) already shows how parallax effects can be divided into so-called low- and high-frequency components [3], i.e. the radial portion dependent on r' and Ω_d induced by the rigid sphere model as well as the pinna effects dependent on r' and Ω_{dp}.

Evaluation

For evaluation, we used a dense HRTF set as input data, even though SUpDEq initially aims at spatial upsampling of sparse HRTF sets, with the option introduced in this paper to vary the distance of the upsampled dataset. However, as we already analyzed the performance of the upsampling in great detail in our recent publication [8], this study solely deals with synthesis of dense near-field HRTF sets based on a dense far-field dataset. Nevertheless, the described method can of course be applied to sparse far-field datasets to first upsample to any dense sampling grid before applying distance variations.

The evaluation is based on HRTFs of a Neumann KU100 dummy head, measured on a Lebedev grid with 2702 nodes (abbreviated Lebedev$_{2702}$ in the following) as well as on a circular grid with steps of 1° in the horizontal plane at distances of 1.50 m, 1.00 m, 0.75 m, 0.50 m, and 0.25 m. The closest distance of 0.25 m was only measured on the circular grid though (please refer to [2] for more details). In the following, the HRTF set for a distance of 1.50 m represents the dense far-field HRTF set used as input data, whereas the other HRTFs serve as a reference at the respective distance in the near-field.

For synthesis, the far-field HRTF set was first equalized on the corresponding dense Lebedev$_{2702}$ grid Ω_d according to Eq. (4), applying an equalization dataset based on STFs for an incident plane wave, as given in Eq. (5). The radius for the rigid sphere model was calculated according to Algazi et al. [9] based on the dimensions of the Neumann KU100 dummy head, leading to r = 9.19 cm. Next, the equalized set was transformed to the SH domain with the maximal accessible spatial order $N = 44$. To accomplish the distance variation, the equalized set was then de-equalized on the respective parallax-adjusted Lebedev$_{2702}$ grid Ω_{dp} as described in Eq. (8), applying a de-equalization dataset based on STFs for an incident spherical wave according to Eq. (7) for the respective distances $r' = 1.00$ m, 0.75 m, 0.50 m, and 0.25 m and sphere radius r as above. At this point it should be mentioned again that the spatially continuous description of the equalized set in the SH domain allows the reconstruction of transfer functions by means of SH interpolation for any required direction according to the parallax model. Thus, parallax effects can be integrated spatially very precisely with this processing. In a next step, the four resulting near-field datasets were windowed as well as peak-normalized and shifted in time to compensate for the changes in level and propagation time caused by the STFs for a point source. Finally, the synthesized near-field HRTF sets were again transformed to the SH domain with $N = 44$, so that HRTFs can be extracted for any desired direction for further analysis.

As an exemplary comparison between synthesized and measured HRTFs, Fig. 2 shows the magnitude and impulse responses of reference HRTFs and de-equalized HRTFs for the frontal direction ($\phi = 0°$, $\theta = 0°$) at $r' = 1.00$ m and $r' = 0.25$ m. The reference HRTFs were obtained from the circular grid measurements [2], whereas the de-equalized HRTFs were extracted from the respective dataset in SH domain by means of inverse SH transform. Concerning the parameters, we decided for these two distances as they represent the two most distant points of the measured dataset, and for the frontal direction as strong parallax effects usually occur here. The plots for $r' = 1.00$ m (see Fig. 2 (a) and (b)) show only marginal differences between the reference and the synthesized counterpart. The onset points in time domain lie exactly on top of each other and the magnitude response of the de-equalized HRTF covers the one

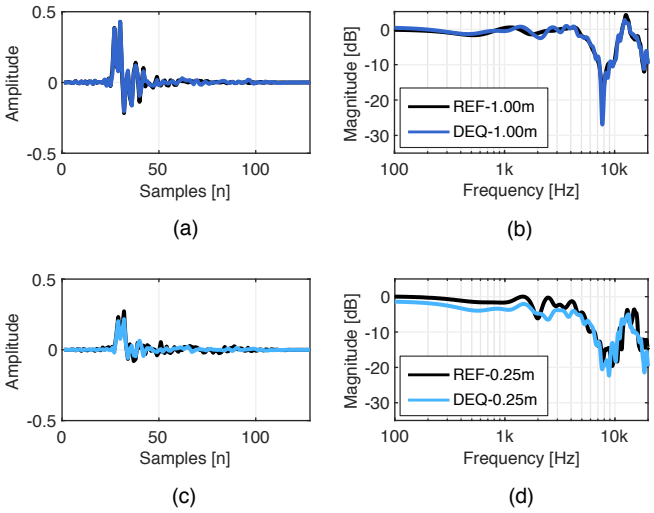

Figure 2: Left ear impulse (left) and magnitude (right) responses of reference HRTFs (black) and de-equalized HRTFs (blue) for the frontal direction ($\phi = 0°$, $\theta = 0°$) at sound source distances $r' = 1.00$ m (a), (b) and $r' = 0.25$ m (c),(d).

from the reference nearly exactly. The differences for $r' = 0.25\,\mathrm{m}$ (see Fig. 2 (c) and (d)) are somewhat larger, even if the responses are still very similar. Again, the onset points in time domain are equal, but the magnitude response of the de-equalized HRTF shows some small deviations. The ripples in the magnitude response of the reference HRTF however are due to reflections between the dummy head and the measurement loudspeaker occurring at this distance [2], which is of course not reproduced by the synthesis. Thus, an exact comparison is somewhat complicated by measurement inaccuracies.
Furthermore, we compared ILDs and ITDs of the synthesized HRTFs with the ones of the circular grid measurements for the two exemplary distances. Again, the de-equalized HRTFs were extracted from the respective dataset in the SH domain by inverse SH transform for the directions of the circular grid ($\theta = 0°$, azimuth resolution of 1°). The broadband ILDs were calculated as the ratio between the energy of the left and right ear head-related impulse response (HRIR, the time-domain equivalent of an HRTF). The ITDs were calculated by means of a threshold-based onset detection on the ten times upsampled and low-pass filtered HRIRs (10th order Butterworth low-pass filter at 3 kHz). Fig. 3 shows the ILDs and ITDs of the reference and of the synthesized dataset at $r' = 1.00\,\mathrm{m}$ and $r' = 0.25\,\mathrm{m}$. As expected for nearby sound sources, both datasets show a significant increase in ILD as a function of distance (see Fig. 3 (a)). Thus, the synthesis adequately reproduces the perhaps most prominent feature of near-field HRTFs. For $r' = 1.00\,\mathrm{m}$, there are slight differences between the reference and the synthesized dataset, with an overall maximum deviation of about 2.5 dB. For $r' = 0.25\,\mathrm{m}$ though, the ILDs of both datasets are pretty similar, showing only slight deviations for rearward directions with an overall maximum difference of about 1.5 dB. In accordance with literature, the ITDs change only marginally with respect to distance (see Fig. 3 (b)). However, when compared to the reference, the synthesis tends to lead to slightly larger ITDs especially for rearward directions, with an overall maximum difference of about 0.06 ms for $r' = 1.00\,\mathrm{m}$ and about 0.04 ms for $r' = 0.25\,\mathrm{m}$.

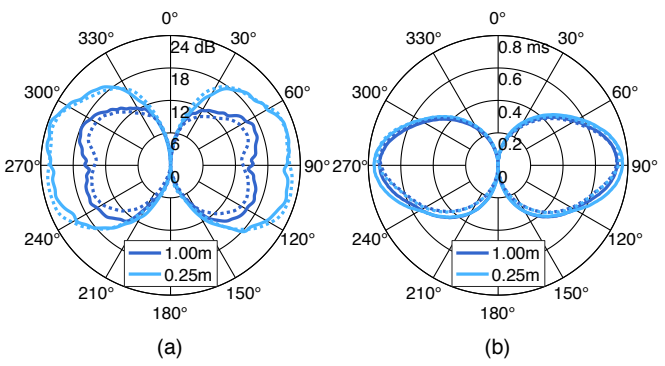

Figure 3: ILDs (a) and ITDs (b) in the horizontal plane for the reference HRTF set (dashed line) and for the de-equalized HRTF set (solid line) at sound source distances $r' = 1.00\,\mathrm{m}$ and $r' = 0.25\,\mathrm{m}$. The radius describes the magnitude of the level differences (in dB) or time differences (in ms).

Conclusion

This paper presented a novel approach to synthesize near-field HRTFs by directional equalization of far-field datasets. The so-called SUpDEq method is based on a (de-)equalization with directional STFs for distance variation, which is quite similar to the common DVF method, and SH interpolation to additionally account for acoustical parallax effects with high spatial accuracy. The evaluation of the proposed method revealed good agreement between synthesized and measured datasets concerning spectral and temporal features as well as regarding ILDs and ITDs. Whereas the SUpDEq method initially only aims at spatial upsampling of sparse HRTF sets, the presented modification of the approach allows to combine upsampling with near-field HRTF synthesis.
In general, SUpDEq provides the most accurate results if the (de-)equalization functions exactly match the properties of the dummy (or human) head. Thus, especially ILDs and ITDs of distance-shifted (and upsampled) near-field HRTF sets could be improved with respect to a particular reference, for example, using an optimized sphere radius or shifted ear positions according to the reference. In this context, applying an ellipsoid to describe the head geometry instead of a rigid sphere model would further enhance the results, as an ellipsoid approximates a human head much better. Combining all this in a freely adjustable head model would furthermore allow to individualize HRTFs in the de-equalization step based on anthropometric head measures.

References

[1] Brungart, D. S. and Rabinowitz, W. M., "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.*, 106(3), pp. 1465–1479, 1999.

[2] Arend, J. M., Neidhardt, A., and Pörschmann, C., "Measurement and Perceptual Evaluation of a Spherical Near-Field HRTF Set," in *Proc. 29th VDT Int. Con.*, pp. 356–363, 2016.

[3] Kan, A., Jin, C., and van Schaik, A., "A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function," *J. Acoust. Soc. Am.*, 125(4), pp. 2233–2242, 2009.

[4] Duda, R. O. and Martens, W. L., "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, 104(5), pp. 3048–3058, 1998.

[5] Duraiswami, R., Zotkin, D. N., and Gumerov, N. A., "Interpolation and range extrapolation of HRTFs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. IV45–IV48, 2004.

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

[7] Rettberg, T. and Spors, S., "Time-Domain Behaviour of Spherical Microphone Arrays at High Orders," in *Proc. 40th DAGA*, pp. 598–599, 2014.

[8] Pörschmann, C., Arend, J. M., and Brinkmann, F., "Directional Equalization of Sparse Head-Related Transfer Function Sets for Spatial Upsampling," *IEEE Transaction on Audio, Speech, and Language Processing*, (in press), 2019.

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