A Method for Spatial Upsampling of Directivity Patterns of Human Speakers by Directional Equalization

Christoph Pörschmann[1], Johannes M. Arend[1,2]
[1] TH Köln, Institute of Communications Engineering, Cologne, Germany
[2] TU Berlin, Audio Communication Group, Berlin, Germany
Email: christoph.poerschmann@th-koeln.de

Introduction

In many everyday situations we experience the influence of the directivity of human speakers. We perceive loudness and timbre significantly different when one faces us or when one turns away from us. An important specific aspect of human speech production is its dynamic directivity, i.e. time-variant alterations of the directivity which occur while speaking (e.g. [1, 2]) or while singing (e.g. [3]). To determine the directivity pattern of a speaker or singer, the sound radiation into an appropriately large number of directions needs to be captured. For a reproduction in virtual acoustic environments, measurements on a spherical sampling grid are advantageous. Generally, these measurements can be performed sequentially for an arbitrary number of directions. Alternatively, with a surrounding microphone array [1, 4, 5] the measurements can be performed simultaneously. By this, loudness or articulation-depending influences on the directivity of natural speakers can be investigated. However, as the setup of such surrounding arrays is restricted to a limited number of microphones, the spatial resolution is low. Applying the principle of reciprocity allows to regard the radiation from a distinct point on the sphere in the same way as a sound wave reaching the sphere. By this, upsampling sparse directivity sets which are obtained from measurements with a low spatial resolution, can be handled comparably to measured sets of head-related transfer functions (HRTFs). For HRTFs many interpolation methods have been elaborated. HRTF sets can be described in the spherical harmonics (SH) domain [6]. In this case, the HRTFs are decomposed into spherical base functions of different orders N, where higher orders correspond to a higher spatial frequency. Spatial upsampling can be applied by evaluating the SH functions at the corresponding directions.

Recently, we presented the SUpDEq (Spatial Upsampling by Directional Equalization) method [7] which removes directional components from HRTF sets before SH transform. The directivity set is spatially equalized by a division with a corresponding rigid sphere transfer function (STF), which can be regarded as a simplified directivity set only comprising basic temporal and spectral features. After spatial upsampling (SH interpolation), a de-equalization by means of a spectral multiplication with the same STFs is performed and thus a spatially upsampled directivity set is recovered. The proposed SUpDEq method and its application to human speaker directivities is described in greater detail in this paper.

Method

A directivity pattern can be described by the frequency-depending pressure function $p_{DIR}(\omega, \Omega_g)$ measured at G discrete angles $\Omega_g = \{(\phi_1, \theta_1), \ldots, (\phi_G, \theta_G)\}$ at azimuth ϕ, and elevation θ at a predefined distance in the far-field. The corresponding SH coefficients $f_{nm}(\omega)$ are obtained via the SH transform, often referred to as spatial (or spherical) Fourier transform [6, p. 2]

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

sampling weights β_g depend on the type of the grid. The complex conjugation is denoted by $(\cdot)^*$ and the spherical harmonis of order n and mode/degree m by

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

with the associated Legendre functions P_n^m, and $i = \sqrt{-1}$ the imaginary unit. The inverse spatial Fourier transform

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

can be used to recover p_{DIR} at arbitrary angles. N denotes the maximal order. If p_{DIR} is strictly order-limited, a sufficient choice of N results in $p_{DIR} = \widehat{p}_{DIR}$. In case the order of p_{DIR} exceeds N, order-limitation errors and spatial aliasing occurs. Depending on the spatial sampling grid Ω_g, the coefficients f_{nm} can be calculated up to a maximum order N without suffering of spatial aliasing. In this case, the number of measured directions G directly corresponds to the maximum order N by $G \propto (N+1)^2$. Consequently, sparse directivity patterns result in a limited SH order. To avoid spatial aliasing for the full audio bandwidth, a maximum order $N \geq kr$ with $k = {}^{\omega}/_{c}$, and r being the head radius is required. Thus, an appropriate pre-processing that reduces the spatial complexity of p_{DIR} will ease the requirement on G. The following section describes the SUpDEq method applied to directivity patterns of human speakers. According to the principle of reciprocity the SUpDEq method [7] can be applied to the sound radiation from the human mouth as shown in Fig. 1. First, the sparse directivity set p_{DIR} 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}

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

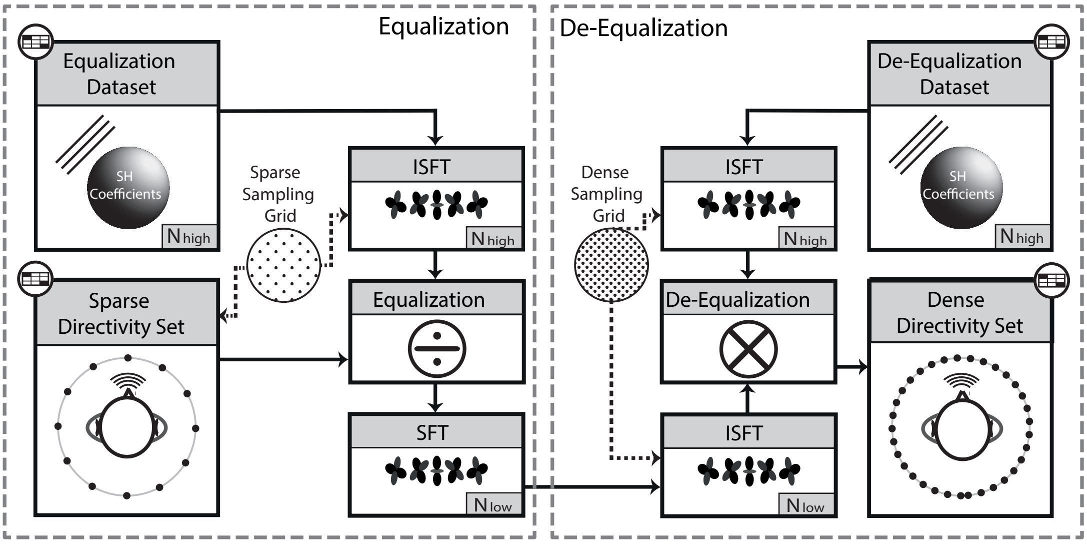

Figure 1: Block diagram of spatial upsampling of a speaker directivity set with the SUpDEq method. Left panel: A directivity set is equalized on the corresponding sparse sampling grid. The set is then transformed to SH domain with $N = N_{low}$. Right panel: The equalized set is de-equalized on a dense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

The equalization dataset reduces the directional dependency in p_{DIR} to a certain degree and can be regarded as a simplified directivity set which features basic temporal and spectral components but does not carry information on the fine structure of the head. In this study a rigid sphere transfer function is used [6, p. 227]:

$$H_{STF}(\omega, \Omega_g) = P4\pi \sum_{n=0}^{N_{high}} \sum_{m=-n}^{n} i^n j_n(kr) Y_n^m(\Omega_e) Y_n^m(\Omega_g)^*, \quad (5)$$

with P denoting an arbitrary sound pressure, j_n the spherical Bessel function of the first kind, and the mouth position Ω_e. We determined in informal tests an optimal Ω_e of $\phi = 0°$ and $\theta = -25°$ which is in-line with other studies [8]. The radius r should match the physical dimensions of a human head. As the equalization dataset is based on an analytic 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 directivity set are obtained by applying the SH transform according to Eq. (1) on the equalized directivity dataset given by Eq. (4) up to an appropriate low maximal order N_{low}. Then, in a third step an upsampled directivity set $\widehat{p}_{DIR,EQ}$ 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). Finally, the directivity is resored by a subsequent de-equalization by means of spectral multiplication with a de-equalization dataset H_{DEQ}:

$$\widehat{p}_{DIR,DEQ}(\omega, \Omega_d) = \widehat{p}_{DIR,EQ}(\omega, \Omega_d) \cdot H_{DEQ}(\omega, \Omega_d)\,. \quad (6)$$

This last step recovers energies at higher spatial orders that were transformed to lower ones in the first step. For de-equalization the STF as given in Eq. (5) is used. Again, $p_{DIR} = \widehat{p}_{DIR,DEQ}$ holds if N_{low} and N_{high} are chosen appropriately. Otherwise, deviations will be caused by signal energy which, after the equalization, still is apparent at high modal orders $N > N_{low}$. Due to spatial aliasing, this signal energy is irreversibly mirrored to lower orders, and we obtain $p_{DIR} \approx \widehat{p}_{DIR,DEQ}$. In the following section these influences are analyzed and the SUpDEq method is compared to SH interpolation without any pre- or postprocessing.

Materials

For evaluation we measured the speaker directivity of a HEAD acoustics HMS II.3 head and mouth simulator on a dense grid in the anechoic chamber at TH Köln. We calculated the head radius according to [9] resulting in $r = 8.78$ cm. We applied a sequential measurement and used the VariSphear measurement system [10] for precise positioning of the dummy head at the spatial sampling positions and for capturing the impulse responses. As microphone we used a Microtech Gefell M296S and measured at a distance of 2 m. An RME Babyface was used as AD / DA converter and microphone preamp. The excitation signal for all measurements was an emphasized sine sweep with 2^{18} samples at 48 kHz sampling rate. Generally, the measurement procedure and the subsequent postprocessing is comparable to the HRTF measurements described in [11]. We measured a reference set on a Lebedev full spherical grid with 2702 points, which was transformed to the SH domain with $N = 35$. To generate various sparse sets, we simply spatially subsampled the reference set in SH domain by means of the inverse SH transform on a Lebedev grid Ω_s ($N_{low} = 4$, 7, 10, 13; corresponding to 38, 86, 170, 266 positions).

Evaluation

In the following we compare the high density reference set ($N_{high} = 35$) to sparse sets processed with SUpDEq and to order-limited (OL) directivities, obtained by means of SH interpolation without pre- or post-processing. We

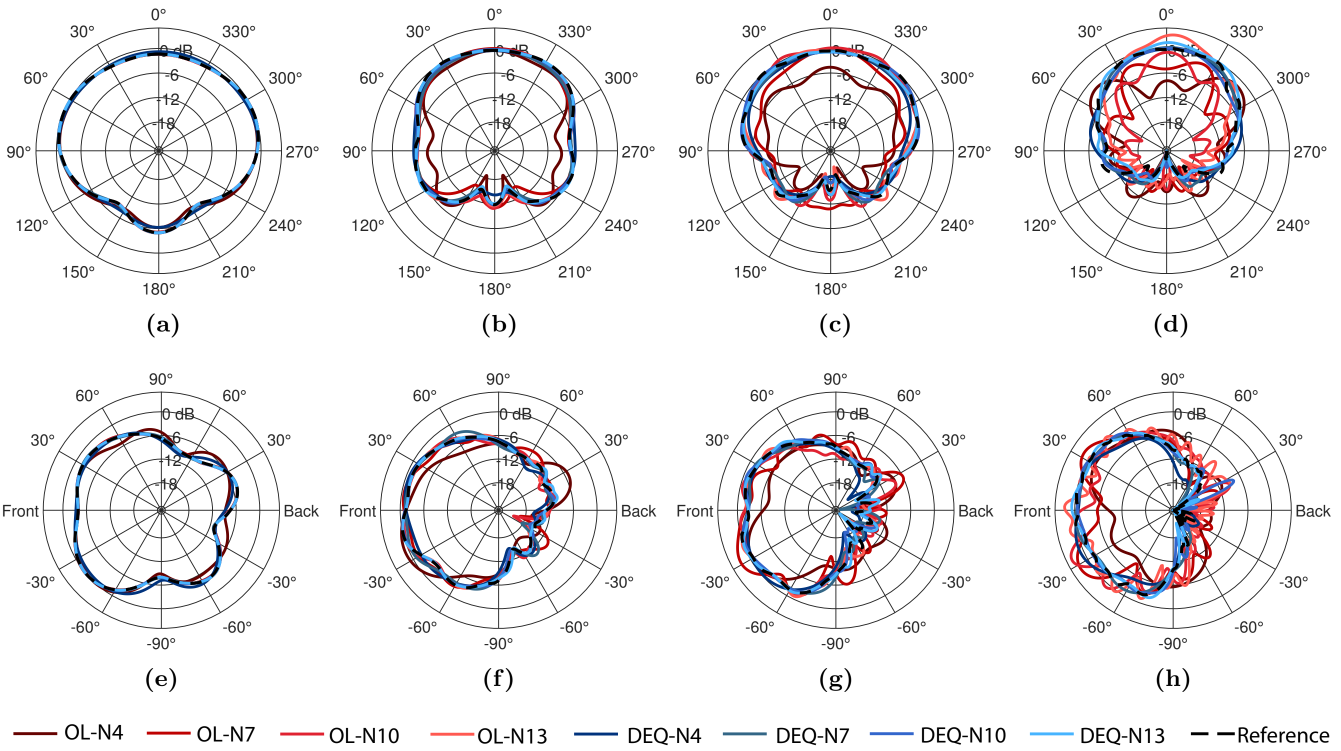

Figure 2: Radiation pattern in the horizontal plane (a – d) and the vertical plane (e – h) for the order-limited sets (red) and SUpDEq-processed sets (blue). Plotted are $N_{low} = 4, 7, 10, 13$ and the reference directivity set (black, dashed) in dB. (a) and (e): $1\,kHz$ band, (b) and (f): $2\,kHz$ band, (c) and (g): $4\,kHz$ band, and in (d) and (h): $8\,kHz$ band.

used a Matlab-based implementation of SUpDEq, which has been presented in [7].

In a first step, the directivity pattern was analyzed in octave bands in the horizontal and vertical plane. As shown in Fig. 2 in the octave bands around 1 kHz and 2 kHz only slight differences between the different sparse grids and the reference occur and are only relevant for $N_{low} = 4, 7$ for rear directions. At 4 kHz for the order-limited set deviations arise for $N_{low} = 7$ and spread at $N_{low} = 4$ over various radiation directions both in the horizontal and median plane. For the de-equalized (DEQ) set, which is the spatially upsampled dataset after the SUpDEq processing, differences to the reference are mainly observable for rearward directions. While for the order-limited radiation patterns at $N_{low} = 4, 7$ in the 8 kHz octave band deviations of 10 dB and more occur for various directions, the deviations are for the de-equalized set limited to directions to the rear and the top. It can be summarized that differences to the reference are much larger for the order-limited patterns than for the de-equalized ones.

To determine the deviations to a reference set over all T measured directions $\Omega_t = \{(\phi_1, \theta_1), \ldots, (\phi_T, \theta_T)\}$ of a test sampling grid, we calculated the spectral differences averaged across all 2702 measured directions:

$$\Delta G_f(\omega) = \frac{1}{N_{\Omega_t}} \sum_{\Omega_t} |20lg \frac{|p_{DIR,REF}(\omega, \Omega_t)|}{|p_{DIR,TEST}(\omega, \Omega_t)|}|, \quad (7)$$

with $p_{DIR,REF}$ the directivity extracted from $f_{REF,nm}$ and $p_{DIR,TEST}$ the respectively processed sparse directivity. Figure 3 illustrates the frequency-dependent spectral differences $\Delta G_f(\omega)$ at $N_{low} = 4, 7, 10, 13$ for the SUpDEq method and for an strictly order-limited interpolation. It can be observed that the spectral differences are significantly smaller for the SUpDEq method than for order-limited interpolation. Furthermore, for order-limited interpolation the spectral differences increase distinctively above 2 dB at aliasing frequencies between 2 and 6 kHz, depending on N. For the SUpDEq method, the spectral differences are generally lower and show a much more gentle rise. Here, for orders $N_{low} = 7, 10, 13$, differences are below or about 2 dB for frequencies up to 10 kHz.

Finally, the spatial distribution of the differences was in-

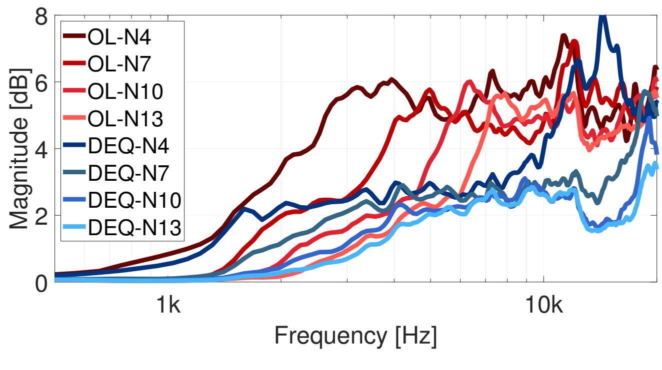

Figure 3: Spectral differences $\Delta G_f(\omega)$ in dB between reference directivity set and the order-limited sets (red) and SUpDEq-processed sets (blue) for $N_{low} = 4, 7, 10, 13$.

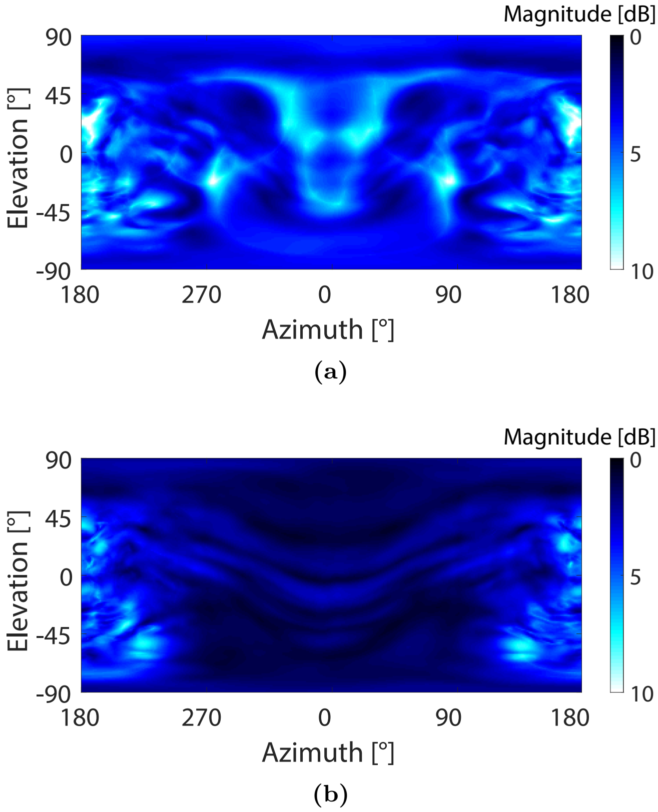

Figure 4: Spectral differences $\Delta G_{sp}(\Omega_t)$ per sampling point to the reference directivity set for order-limited interpolation (a) and for the SUpDEq method (b) at $N = 4$ ($f \leq 8\,\text{kHz}$).

vestigated and the directional deviation across all frequencies was calculated as

$$\Delta G_{sp}(\Omega_t) = \frac{1}{N_\omega} \sum_{\omega} \mid 20lg \frac{\mid p_{DIR,REF}(\omega, \Omega_t) \mid}{\mid p_{DIR,TEST}(\omega, \Omega_t) \mid} \mid . \quad (8)$$

Fig. 4 shows the spectral differences $\Delta G_{sp}(\Omega_t)$ per sampling point for order-limited interpolation and for the SUpDEq method at $N = 4$ and $f \leq 8\,\text{kHz}$. In this case, the test sampling grid Ω_t is full spherical and calculated for ϕ and θ in steps of 1°. The plots show that, independent of the method for spatial upsampling, the spectral differences are maximal for directions to the rear. Generally, the order-limited interpolation results in distinct spectral differences spread over the entire angular range, while the SUpDEq method leads to differences mainly for sound radiation to the rear.

Conclusion

We presented an SH based approach for spatial upsampling (interpolation) of sparse human speaker directivities, and applied the SUpDEq method which removes directional components of the dataset prior to the upsampling by directional equalization. The analysis of the results showed that the highest inaccuracies and deviations to the reference occur for rearward sound incidence. Due to constructive interferences of the sound radiated, a bright spot can be observed here. However, the interference pattern changes rapidly for adjacent directions, especially towards higher frequencies corresponding to a high spatial order N_{max} in the SH domain. The evaluation revealed that for a human speaker already at $N_{low} = 4$ with 38 measured directions on a Lebedev grid, a decent full-spherical dense directivity set can be generated, which might be sufficient for various applications in the field of virtual acoustics.

In subsequent research we will investigate to what extent the results can be transfered to surrounding arrays, e.g. to a pentakis dodecahedron with 32 microphones. Such arrays have been realized e.g. in [4, 5] and allow to measure directivities at $N_{low} = 4$ in real time. Thus, they are not restricted to impulse-response based measurements, but allow for measuring time-variant influences of the directivity pattern. Furthermore, it has to be investigated to what extent the model of a sound source located at a considerably small mouth opening holds for a natural human speaker. Here influences of sound radiation by glottis, nose and even towards lower frequencies of radiation by the human skull need to be considered.

The research presented in this paper has been carried out in the Research Project NarDasS funded by the Federal Ministry of Education and Research in Germany; Support Code: BMBF 03FH014IX5-NarDasS. The authors thank Raphael Gillioz for supporting the measurements.

References

[1] Brandner, M., Frank, M., and Rudrich, D., "DirPat - Database and Viewer of 2D/3D Directivity Patterns of Sound Sources and Receivers," in *144th Audio Eng. Soc. Conv.*, 2018.

[2] Kocon, P. and Monson, B. B., "Horizontal directivity patterns differ between vowels extracted from running speech," *J. Acoust. Soc. Am.*, 144(1), pp. EL7–EL12, 2018.

[3] Katz, B. and D'Alessandro, C., "Directivity measurements of the singing voice," in *Proceedings of the 19th International Congress on Acoustics*, 2007.

[4] Pollow, M., *Directivity Patterns for Room Acoustical Measurements and Simulations*, Logos Verlag Berlin GmbH, 2015.

[5] Arend, J. M., Stade, P., and Pörschmann, C., "Binaural reproduction of self-generated sound in virtual acoustic environments," in *Proc. of the 173rd Meeting of the Acoust Soc. of Am.*, 2017.

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

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

[8] Marshall, A. H. and Meyer, J., "The directivity and auditory impressions of singers," *Acustica*, 58, pp. 130–140, 1985.

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

[10] Bernschütz, B., Pörschmann, C., Spors, S., and Weinzierl, S., "Entwurf und Aufbau eines variablen sphärischen Mikrofonarrays für Forschungsanwendungen in Raumakustik und Virtual Audio," in *Proc. of the 36th DAGA*, pp. 717–718, 2010.

[11] Bernschütz, B., "A Spherical Far Field HRIR / HRTF Compilation of the Neumann KU 100," in *Proc. of the 39th DAGA*, pp. 592–595, 2013.