Converting hemispherical microphone array recordings

Hannes Pomberger, Franz Zotter
Institut für Elektronische Musik und Akustik, Universität für Musik und Darstellende Kunst, Graz, Austria,
Email: pomberger@iem.at

Introduction

For three-dimensional recording with hemispherical microphone arrays it is advantageous to assume an acoustic half-space bounded by a rigid horizontal plane from below. The reflection on this plane simplifies what the array records to limited-order even-symmetric spherical harmonics. However, due to the reflection, the representation is non-isotropic, which can be seen as distortion/interference of any recorded limited-order direction from the upper half space with its lower-half-space image source.
By contrast, conventional compact spherical arrays for Ambisonic recording capture a full set of limited-order spherical harmonics, i.e. include those with odd symmetry, and therefore record with isotropic directional resolution.

In this article we propose to reduce unwanted artifacts of hemispherical recordings by completion of the resulting even-symmetric spherical harmonic signals with odd-symmetric ones. The improvements are discussed based on the perceptually motivated performance measures E and $\boldsymbol{r}_\mathrm{E}$ that characterize direction dependencies of the loudness, mislocalization, and source width.

Hemispherical microphone arrays

The motivation to use compact hemispherical microphone arrays is to capture only sounds from the upper hemisphere. This is achieved by imposing a sound hard boundary condition at the equatorial plane. In spherical coordinates[1], the angular solutions of the Helmholtz equation fulfilling this boundary condition are those spherical harmonics with even symmetry with regard to z. These functions form an orthogonal and complete set of basis functions on the unit hemisphere S^2. On a hemisphere of radius r_M, the sound pressure due to incident sound is accordingly expressed as the expansion into spherical harmonics

$$p(r_M\boldsymbol{\theta}) = \sum_{n=0}^{N} \sum_{\substack{-n\le m\le n \\ 2|(m+n)}} Y_n^m(\boldsymbol{\theta})\, \psi_{nm}, \tag{1}$$

where ψ_{nm} are the expansion coefficients and $Y_n^m(\boldsymbol{\theta}) = N_n^{|m|}\, \mathsf{P}_n^{|m|}(\cos\theta) \begin{cases}\cos(m\varphi), & \text{for } m \ge 0\\ \sin(m\varphi), & \text{for } m < 0\end{cases}$ are the spherical harmonics of order n and degree m; $\mathsf{P}_n^{|m|}$ denotes the associated Legendre functions, and $N_n^{|m|}$ is a scalar normalization term. The second line of the sum over m includes the restriction to even-symmetric spherical harmonics: $n+m$ must be divisible by 2.

[1] Within this article, we define the position vector in terms of spherical coordinates as $\boldsymbol{r} = r\,\boldsymbol{\theta}$, whereby r is the radial distance and $\boldsymbol{\theta}$ is the direction vector $\boldsymbol{\theta} = [\cos(\varphi)\sin(\theta), \sin(\varphi)\sin(\theta), \cos(\theta)]^\mathrm{T}$ with φ and θ being the azimuth and zenith angle, respectively.

Modal sound field decomposition

Below an upper frequency limit, compact spherical microphone arrays for Ambisonic recording can be assumed to capture a limited-order sound pressure distribution. The infinite series in eq. (1) can be truncated to a maximum order N, and we may re-express it by the vector product

$$p(r_M\boldsymbol{\theta}) = \boldsymbol{y}_{e,\mathrm{N}}^\mathrm{T}(\boldsymbol{\theta})\; \boldsymbol{\psi}_{e,\mathrm{N}}, \tag{2}$$

where $\boldsymbol{y}_{e,\mathrm{N}}(\boldsymbol{\theta}) := [Y_n^m(\boldsymbol{\theta})]_{q=1\ldots(\mathrm{N}+1)(\mathrm{N}+2)/2}$ and $\boldsymbol{\psi}_{e,\mathrm{N}} := [\psi_n^m]_{q=1\ldots(\mathrm{N}+1)(\mathrm{N}+2)/2}$, with the integer index $q := \frac{n}{2}(n+2) + \frac{m}{2} + 1$.

Due to the orthogonality of the even symmetric spherical harmonics on the hemisphere S^2, expansion coefficients are obtained by $\psi_{nm} = 2\int_{S^2} Y_n^m(\boldsymbol{\theta})\, p(\boldsymbol{r}_M)\mathrm{d}\boldsymbol{\theta}$ for $2|(n+m)$. As the microphone array captures a spatially discrete sound pressure, the integral is typically re-formulated as

$$\boldsymbol{p} = \boldsymbol{Y}_{e,\mathrm{N}}\boldsymbol{\psi}_{e,\mathrm{N}} \quad \Rightarrow \quad \boldsymbol{\psi}_{e,\mathrm{N}} = \boldsymbol{Y}_{e,\mathrm{N}}^\dagger\, \boldsymbol{p}, \tag{3}$$

with the pseudo-inverse $()^\dagger$ of the spatially discretized matrix of harmonics $\boldsymbol{Y}_{e,\mathrm{N}} = [\boldsymbol{y}_{e,\mathrm{N}}^\mathrm{T}(\boldsymbol{\Theta}_l)]_l$ applied on the spatially discretized sound pressure $\boldsymbol{p} = [p(r_M\boldsymbol{\Theta}_l)]_l$.

The coefficients of a surrounding source distribution reproducing the recording are radial-filtered version thereof

$$\boldsymbol{\phi}_{e,\mathrm{N}} = \mathrm{diag}\left\{\boldsymbol{w}_\mathrm{N}(kr_M)\right\}\boldsymbol{\psi}_{e,\mathrm{N}}, \tag{4}$$

with $\boldsymbol{w}_\mathrm{N}(kr_M) = [w_0(kr_M), \ldots, \overbrace{w_\mathrm{N}(kr_M), \ldots, w_\mathrm{N}(kr_M)}^{\mathrm{N}+1}]$, and $w_\mathrm{n}(kr_M)$ describing the holographic radial filters.

Reproduction

In general, reproduction in an Ambisonics framework is conducted by mapping the modal representation of a (continuous) surrounding source distribution to signals for the particular (discrete) spherical loudspeaker setup. Stacking the loudspeaker signals in a vector $\boldsymbol{g}$ and the signals of surrounding source distribution in a vector $\boldsymbol{\phi}_\mathrm{N}$, this mapping is expressed by

$$\boldsymbol{g} = \boldsymbol{D}\boldsymbol{\phi}_\mathrm{N}. \tag{5}$$

The mapping matrix $\boldsymbol{D}$ is referred to as the *decoder* and it requires an elaborate design to provide a psychoacoustical accurate sound scene rendering. Existing techniques, e.g. [1], allow for accurate decoding of sound scenes captured by a full-spherical microphone array to hemispherical and other partial-spherical speaker setups.

Nevertheless this does not imply that recordings of a hemispherical microphone array are also reproduced without distortion by a standard decoder for, e.g., a surrounding hemispherical loudspeaker array. Standard decoders still require the full set of spherical harmonics signals, whereas hemispherical microphone arrays provide the even-symmetric ones, only. A direct rendering, is equivalent to zeroing the odd-symmetric components and thus to adding a symmetric lower-half space image sources to a full set of spherical harmonics. This causes the need for a conversion methods that optimally suppresses unwanted artifacts from image source interference from below by retrieving suitable odd-symmetric components.

Basic conversion

By proper radial filtering, the hemispherical array delivers the coefficients of the even symmetric spherical harmonics, cf. eq. (4), which yield an angular source distribution $\breve{f}(\boldsymbol{\theta})$ over the hemisphere S^2

$$\breve{f}(\boldsymbol{\theta}) = \begin{cases} \boldsymbol{y}_{e,\mathrm{N}}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{\phi}_{e,\mathrm{N}} & \text{for } \boldsymbol{\theta} \in S^2 \\ 0, & \text{for } \boldsymbol{\theta} \notin S^2. \end{cases} \tag{6}$$

The aim of an accurate conversion is to find an vector $\hat{\boldsymbol{\phi}}_{\mathrm{N}}$ that yields a distribution $\hat{f}(\boldsymbol{\theta})$ consisting of the full set of spherical harmonics up to order N,

$$\hat{f}(\boldsymbol{\theta}) = \boldsymbol{y}_{\mathrm{N}}^{\mathrm{T}}(\boldsymbol{\theta})\hat{\boldsymbol{\phi}}_{\mathrm{N}}, \tag{7}$$

which is optimal in some sense.

The basic conversion is optimal by approximating $\breve{f}(\boldsymbol{\theta})$ in terms of $\hat{f}(\boldsymbol{\theta})$ in the least-square-error sense. This yields the following optimization problem on the full sphere $\mathbb{S}^2$,

$$\min \int_{\mathbb{S}^2} |\hat{f}(\boldsymbol{\theta}) - \breve{f}(\boldsymbol{\theta})|^2 \mathrm{d}\boldsymbol{\theta}. \tag{8}$$

Its solution can be shown to be

$$\hat{\boldsymbol{\phi}}_{\mathrm{N}} = \boldsymbol{M}\boldsymbol{\phi}_{e,\mathrm{N}}, \tag{9}$$

with the conversion matrix defined by the integral over the hemisphere S^2

$$\boldsymbol{M} = \int_{S^2} \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta})\boldsymbol{y}_{e,\mathrm{N}}(\boldsymbol{\theta})^{\mathrm{T}} \mathrm{d}\boldsymbol{\theta}. \tag{10}$$

Performance measures

An error-free retrieval of the unknown odd-symmetric components is infeasible. Quantifying the conversion performance requires error measures of perceivable features. The following measures are perceptually motivated and their spatially discrete equivalents have proven to be psychoacoustically relevant estimates for the performance of discrete amplitude panning functions, cf.[2].

The *energy* measure,

$$E = \int_{\mathbb{S}^2} |f(\boldsymbol{\theta})|^2 \mathrm{d}\boldsymbol{\theta}, \tag{11}$$

is proportional to the perceived loudness of the surrounding source distribution, cf.[3]. The $\boldsymbol{r}_{\mathrm{E}}$ measure,

$$\boldsymbol{r}_{\mathrm{E}} = \frac{\int_{\mathbb{S}^2} \boldsymbol{\theta}|f(\boldsymbol{\theta})|^2 \mathrm{d}\boldsymbol{\theta}}{E}, \tag{12}$$

is a vector pointing in the perceived direction of the surrounding source distribution, and its length is proportional to its angular spread, cf.[3]. Similarly as in [3], we define the angular mapping error as the directional deviation of $\boldsymbol{r}_{\mathrm{E}}$ from the actual source direction $\boldsymbol{\theta}_s$

$$\epsilon_{\mathrm{E}} = \arccos \frac{\boldsymbol{\theta}_s^{\mathrm{T}} \boldsymbol{r}_{\mathrm{E}}}{\|\boldsymbol{r}_{\mathrm{E}}\|}. \tag{13}$$

Similar to [3], the length $\|\boldsymbol{r}_{\mathrm{E}}\|$ is mapped to an angular spread by

$$\sigma_{\mathrm{E}} = 2\arccos(\|\boldsymbol{r}_{\mathrm{E}}\|). \tag{14}$$

For a single plane wave impinging from the direction $\boldsymbol{\theta}_s$, the coefficients of the order-limited full-spherical source distribution, $f(\boldsymbol{\theta}) = \boldsymbol{y}_{\mathrm{N}}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{\phi}_{\mathrm{N}}$, are $\boldsymbol{\phi}_{\mathrm{N}} = \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}_s)$. In this case, the above measures yield ideal results: $E(\boldsymbol{\theta}_s) = \text{const.}$, $\epsilon_{\mathrm{E}}(\boldsymbol{\theta}_s) = 0$, and $\sigma_{\mathrm{E}}(\boldsymbol{\theta}_s) = \text{const}$. This ideal behavior is a consequence of the isotropy of the full set of spherical harmonics.

Max-$\boldsymbol{r}_{\mathrm{E}}$ weighting: Weighting the components of each order by a suitable factor a_n, $\|\boldsymbol{r}_{\mathrm{E}}\|$ can be maximized, cf. [1]. The spherical source distribution with order weighting is $f(\boldsymbol{\theta}) = \boldsymbol{y}_{\mathrm{N}}^{\mathrm{T}}(\boldsymbol{\theta})\,\mathrm{diag}\,\{\boldsymbol{a}_{\mathrm{N}}\}\,\boldsymbol{\phi}_{\mathrm{N}}$, where the vector $\boldsymbol{a}_{\mathrm{N}} = [a_0, \ldots, \overbrace{a_{\mathrm{N}}, \ldots, a_{\mathrm{N}}}^{2\mathrm{N}+1}]$ contains the weights.

Performance of the basic conversion: Figure 1 shows the performance measures for the basic conversion of a plane wave with max-$\boldsymbol{r}_{\mathrm{E}}$ weighting for different maximum orders N. The results are only shown in dependence of the zenith angle of the recorded plane wave. The representation is invariant under azimuthal rotation. The dashed lines in fig. 1c represent the direction-invariant spread one would obtain from recording the full set of limited-order spherical harmonics with max-$\boldsymbol{r}_{\mathrm{E}}$ weights.

Obviously, the energy measure of a converted hemispherical array recording becomes direction dependent, in particular for the first order. For the zenith direction, the angular mapping error is zero in all conditions, oscillates around zero elsewhere, and causes an elevated image for sources from the horizon. Generally the mapping errors decrease with the order. The ideal spread is nearly achieved, except for first order.

Energy-completing conversion

The direction-dependent energy mapping of the basic conversion motivates a further inspection of how the energy distributes over the spherical harmonic components, in the ideal case of an isotropic mapping. The spherical harmonic components of a single plane wave with amplitude s, impinging from $\boldsymbol{\theta}_s$, are $\phi_{nm} = s\,Y_n^m(\boldsymbol{\theta}_s)$. From the closure relation,

$$\sum_{m=-n}^{m} |Y_n^m(\boldsymbol{\theta})|^2 = \frac{2n+1}{4\pi}, \tag{15}$$

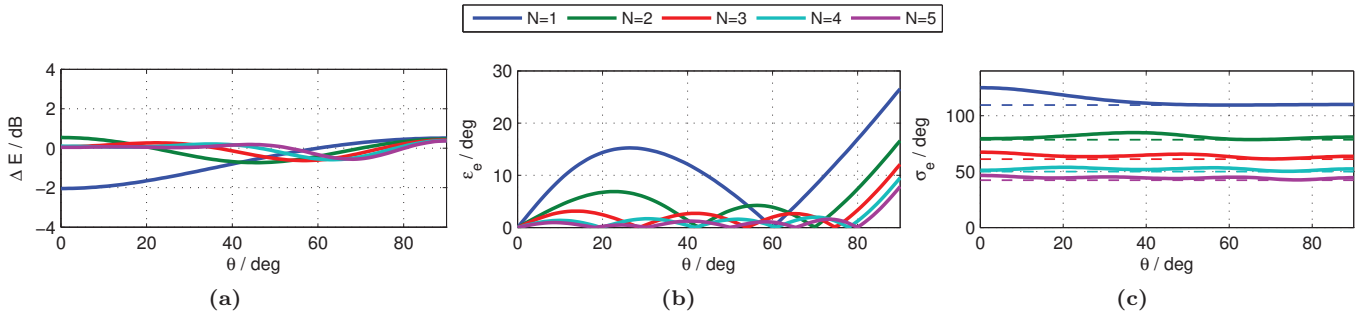

Figure 1: Performance measures for the basic conversion of a plane wave with max-$\boldsymbol{r}_{\mathrm{E}}$ weighting for different maximum orders N in dependence of its incident zenith angle.

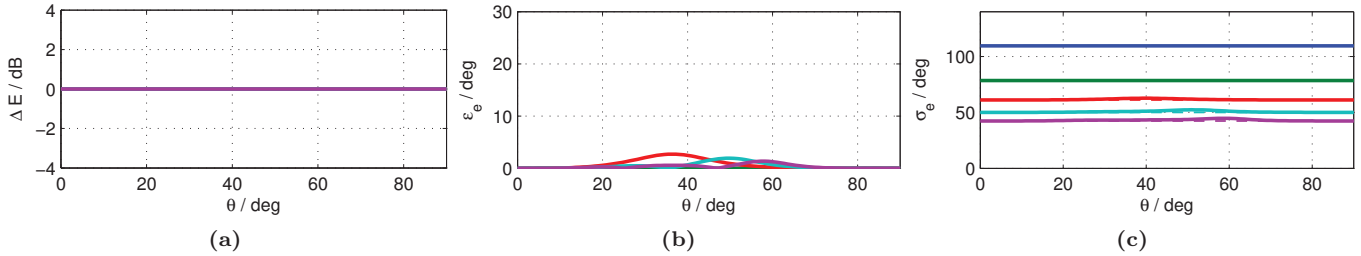

Figure 2: Performance measures for the energy-completing conversion of a single plane wave with max-$\boldsymbol{r}_{\mathrm{E}}$ weighting for different maximum orders N in dependence of its incident zenith angle.

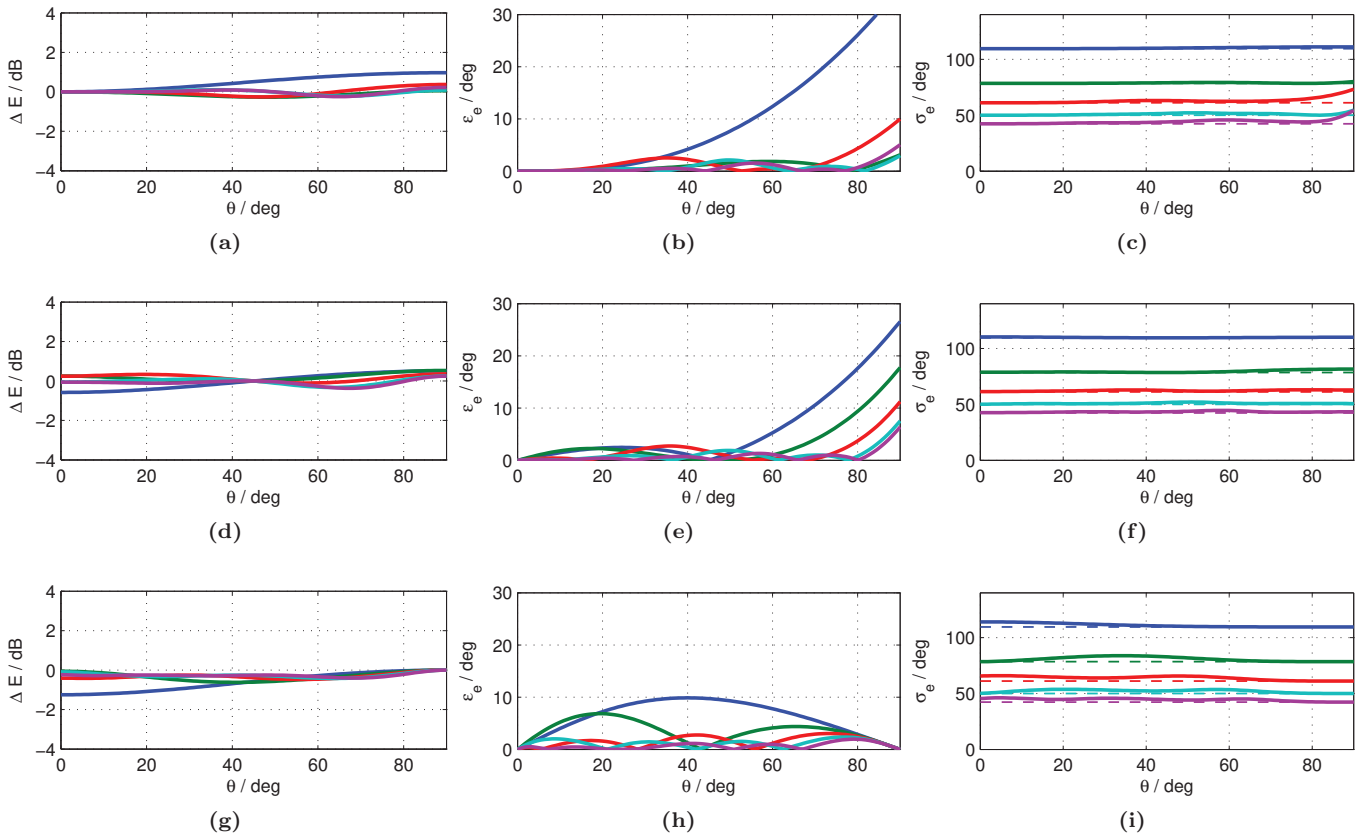

Figure 3: Performance measures for the energy-completing conversion of a plane wave with max-$\boldsymbol{r}_{\mathrm{E}}$ weighting for different maximum orders N in dependence of its incident zenith angle in presence of a second plane wave with an incident zenith angle of (a) to (c) $\theta_2 = 0°$, (d) to (f) $\theta_2 = 45°$, and (d) to (f) $\theta_2 = 90°$.

it follows that for an order n the sum of absolute squares over all m for a plane wave yields $\sum_{m=-n}^{n} |\phi_{nm}|^2 = \frac{2n+1}{4\pi} s^2$.

This sum for the n^{th} order relates to squared absolute value of the zeroth order component by

$$\sum_{m=-n}^{n} |\phi_{nm}|^2 = (2n+1)|\phi_{00}|^2, \tag{16}$$

what can be used as an additional criterion for an energetically completing conversion.

For a sum of uncorrelated plane waves, the above equation still describes the energy distribution of an ideally isotropic capture after replacing the squared absolute values by their expected values $\mathrm{E}\left[|\phi_{nm}|^2\right]$. Thus for capture with a hemispherical array, it seems more than well-justified to enforce an energy constraint by a reformulated eq. (16) in addition to the original minimization problem:

$$\begin{aligned} &\min \int_{\mathbb{S}^2} |\hat{f}(\boldsymbol{\theta}) - \breve{f}(\boldsymbol{\theta})|^2 \mathrm{d}\boldsymbol{\theta}, \\ \text{s.t.} \quad &\sum_{m=-n}^{n} \mathrm{E}\left[|\hat{\phi}_{nm}|^2\right] = (2n+1)\,\mathrm{E}\left[|\phi_{00}|^2\right], \\ &\text{for } n = 1, \ldots, N. \end{aligned} \tag{17}$$

For notational convenience, we split the solution of the above optimization problem into even- and odd-symmetric components $\hat{\boldsymbol{\phi}}_{e,\mathrm{N}}$, $\hat{\boldsymbol{\phi}}_{o,\mathrm{N}}$. The solution of the even-symmetric part yields

$$\hat{\boldsymbol{\phi}}_{e,\mathrm{N}} = \frac{1}{2}\boldsymbol{\phi}_{e,\mathrm{N}}. \tag{18}$$

For a particular order n, the odd-symmetric part yields

$$\hat{\boldsymbol{\phi}}_{o,n} = \frac{1}{2}\alpha_n \boldsymbol{M}_{o,n}\, \boldsymbol{\phi}_{e,\mathrm{N}}, \tag{19}$$

with the matrix $\boldsymbol{M}_{o,n} = \int_{S^2} \boldsymbol{y}_{o,n}(\boldsymbol{\theta})^{\mathrm{T}} \boldsymbol{y}_{e,\mathrm{N}}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}$ and scalar

$$\alpha_n = \sqrt{\frac{(2n+1)\,\mathrm{E}[|\phi_{00}|^2] - E\left[\|\boldsymbol{\phi}_{e,n}\|^2\right]}{\mathrm{E}\left[\|\boldsymbol{M}_{o,n}\,\boldsymbol{\phi}_{e,\mathrm{N}}\|^2\right]}}. \tag{20}$$

Interestingly, this result is similar to the basic conversion of eq. (9), except for the scalar factor α_n that enforces the constraint in every order n.

Performance for a single plane wave: For a single plane wave with max-$\boldsymbol{r}_{\mathrm{E}}$ weighting, fig. 2 shows the performance measures for the energy-completing conversion eqs. (18) to (20) using different maximum orders N. Obviously, the energy measure in fig. 2a yields a perfect result due to the new constraint. In addition, compared to the basic conversion in fig. 1, also the angular mapping error as well as the angular spread become almost ideal.

Non-additivity: In contrast to the basic conversion, which is accomplished by a constant matrix cf. eq. (9), the energy-completing conversion is non-additive due to its adaptive scaling of the odd-components by the non-linear factor α_n. Hence, for a sum of two or more plane waves, one could expect distortions in the mapping of the individual plane wave components.
To illustrate this, we consider two plane waves with uncorrelated unit-variance signals, impinging from two different directions $\boldsymbol{\phi}_{e,\mathrm{N}} = s_1 \boldsymbol{y}_{e,\mathrm{N}}(\boldsymbol{\theta}_{s1}) + s_2 \boldsymbol{y}_{e,\mathrm{N}}(\boldsymbol{\theta}_{s2})$. Figure 3 shows the resulting performance measures for one plane wave of varying zenith angle, assuming that the second plane wave is coming from the same azimuth and its zenith angle is fixed at $\theta_2 = 0°$ in (a) to (c), $\theta_2 = 45°$ in (d) to (f), and $\theta_2 = 90°$ in (d) to (f). If the directions of both plane waves are the same, the performance is the same as for a single plane wave. There is a slight tendency for a plane wave close to the horizon to be louder. As for the basic conversion, there is no angular error for the zenith direction. The improvement the energy-completing conversion achieved for one plane wave generally reduces by the presence of the second one. Still, except for the increased angular mapping error of the first order and $\theta_2 = 0°$, there is generally a slight improvement.

Considering the dependence on azimuth and zenith angle separately, i.e. $\boldsymbol{y}_{e,\mathrm{N}}(\varphi, \theta)$, the norm of the n^{th} order components is independent of the azimuth direction of either of the plane waves, $\|\boldsymbol{y}_{e,n}(\varphi, \theta)\| = \|\boldsymbol{y}_{e,n}(\varphi', \theta)\|$. Similarly the norm of $\boldsymbol{M}_{o,n}\boldsymbol{y}_{e,n}(\varphi, \theta)$ does not depend on the azimuth direction, i.e. $\|\boldsymbol{M}_{o,n}\boldsymbol{y}_{e,n}(\varphi, \theta)\| = \|\boldsymbol{M}_{o,n}\boldsymbol{y}_{e,n}(\varphi', \theta)\|$. Consequently, the performance measures in fig. 3 remain unchanged when the plane wave directions do not coincide in azimuth.

Conclusion

In this contribution we proposed a new energy-completing conversion for hemispherical array recordings. For one single plane wave, the conversion improves the performance in terms direction dependent loudness, source width, and angular mapping accuracy. The improvement compared to the basic conversion deteriorates when a second plane wave is present. However, there is a slight performance improvement in almost all cases.

The $\boldsymbol{r}_{\mathrm{E}}$ performance measures concerning the angular mapping error and source width become meaningless in a diffuse sound field, which contains an infinitely large number of uncorrelated plane waves. Because the energy constraint of the energy-complementing conversion already enforces the energy distribution of an ideally isotropic recording, we expect it to perform optimally for the conversion of diffuse sound fields, again.

References

[1] F. Zotter and M. Frank, "All-round ambisonic panning and decoding," *Journal of the AES*, 2012.

[2] M. Frank, "Localization using different amplitude-panning methods in the frontal horizontal plane," *Proc. of the EAA Joint Symposium on Auralization and Ambisonics, Berlin*, 2014.

[3] F. Zotter, H. Pomberger, and M. Noisternig, "Energy-preserving ambisonic decoding," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 37–47, 2012.