End-to-End Magnitude Least Squares Binaural Rendering for Equatorial Microphone Arrays

Citation for the original published paper (version of record):
Helmholz, H., Deppisch, T., Ahrens, J. (2023). End-to-End Magnitude Least Squares Binaural Rendering for Equatorial Microphone Arrays. Fortschritte der Akustik - DAGA 2023: 1679-1682

(article starts on next page)

End-to-End Magnitude Least Squares Binaural Rendering for Equatorial Microphone Arrays

Hannes Helmholz, Thomas Deppisch, Jens Ahrens

Chalmers University of Technology, Gothenburg, Sweden

Email: {hannes.helmholz, thomas.deppisch, jens.ahrens}@chalmers.se

Abstract

We recently presented an end-to-end magnitude least squares (eMagLS) binaural rendering method for spherical microphone array (SMA) signals that integrates a comprehensive array model into a magnitude least squares objective to minimize reproduction errors. The introduced signal model addresses impairments due to practical limitations of spherical harmonic (SH) domain rendering, namely, spatial aliasing, truncation of the SH decomposition order, and regularized radial filtering. In this work, we improve the processing model when applied to the recently proposed equatorial microphone array (EMA) to facilitate three degrees-of-freedom head rotations during the rendering. EMAs provide similar accuracy to SMAs for sound fields from sources inside the horizontal plane while requiring a much lower number of microphones. We compare the proposed end-to-end renderers for both array types against a given binaural reference magnitude response. In addition to anechoic array simulations, the evaluation includes measured array room impulse responses to show the method's effectiveness in minimizing high-frequency magnitude errors for all head orientations from SMAs and EMAs under practical room conditions. The published reference implementation of the method has been refined and now includes the solution for EMAs.

Introduction

Accurately capturing and reproducing spatial sound fields is necessary for many virtual and augmented reality applications. Microphone arrays with a rigid spherical scattering body have thereby shown to be favorable for the signal-independent capture of arbitrary acoustic environments. In the following, we consider the popular use case of reproduction via headphones, resulting in a plausible and possibly authentic (i.e., indistinguishable) representation of the environment [5], including rotational movements under consideration of the listener's instantaneous head orientation. This is achieved through binaural rendering in the spherical harmonic (SH) domain using head-related transfer functions (HRTFs).

Spherical microphone arrays (SMAs) [11] provide a direction-independent spatial resolution by requiring a significant number of sensors, i.e., at least $(N+1)^2$ for a desired SH order N, to be equally distributed on a spherical surface. Equatorial microphone arrays (EMAs) reduce the required number of sensors to $2N+1$ on the circumference of a spherical scattering body [2], as shown in Fig. 1 (right). They provide a height-invariant representation of the captured sound field, which is compatible with conventional SH rendering methods for SMAs, i.e., to facilitate dynamic binaural reproduction and arbitrary head rotations with three degrees-of-freedom (3-DOF). The horizontal projection of the sound field may result in alterations of the magnitude spectrum in the binaurally rendered ear signals for non-horizontal sound incidence, which are in the same order of magnitude as the errors arising from conventional order-limited SMAs [2, Fig. 6].

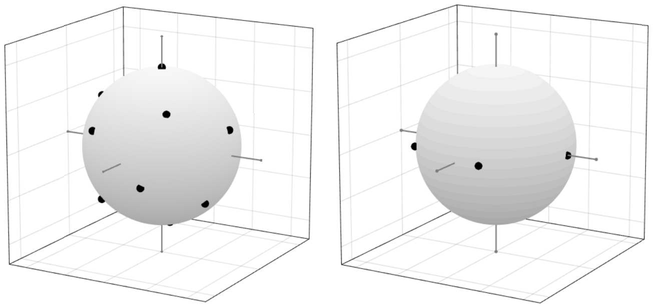

Figure 1: Spherical sampling grids at $N = 2$ for an SMA (left, t-design, 14 microphones) and an EMA (right, 5 microphones).

Practical arrays comprise a limited number of microphones and therefore undersample the sound field in the spatial domain, leading to spatial aliasing above a certain frequency [12]. The sound field is decomposed with limited spatial resolution, resulting in an equivalent truncation of the SH representation of the HRTF data during the binaural rendering [3]. The perceptual relevance of SH order truncation and spatial aliasing has been in combination with different methods developed to mitigate the effects of spatial undersampling, e.g. [15, 8]. To reduce the influence of the scattering body from the captured microphone signals, so-called radial filtering is applied with a need for magnitude regularization at low frequencies [9].

The end-to-end magnitude least squares (eMagLS) binaural rendering method [6] considers above-mentioned impairments caused by spatial undersampling and radial filtering to minimize reproduction errors in a least-squares sense at low frequencies, and in a magnitude-least-squares sense at high frequencies. In this contribution, we extend the method to support rendering signals from EMAs while facilitating 3-DOF sound field rotations.

Rendering Method

Fig. 2 provides a high-level overview of the different steps involved in the binaural rendering of SMA and EMA signals in combination with different methods for mitigating spatial undersampling artifacts. For further details, we refer the reader to [1] and the original publications of the methods mentioned in Fig. 2.

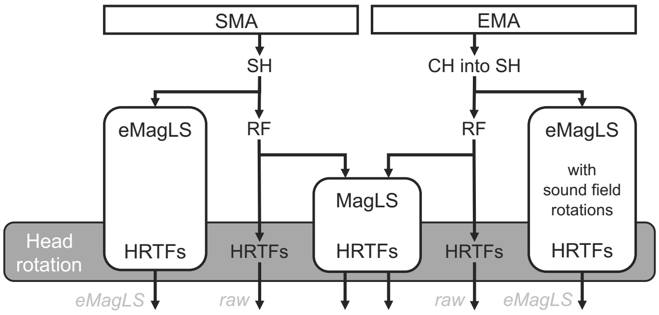

Figure 2: Procedures to represent SMA or EMA signals in the spherical harmonic (**SH**) and circular harmonic (**CH**) domains, application of radial filters (**RF**), subsequent rotation and binaural rendering with head-related transfer functions (**HRTFs**) to obtain ear signals (*raw*). Different preprocessing methods can be implemented for the mitigation of spatial undersampling, e.g. magnitude least squares (**MagLS**) and end-to-end MagLS (**eMagLS**).

eMagLS Rendering for EMA

MagLS and eMagLS are methods for the rendering of SMA signals. They essentially provide a set of SH-domain filters to replace the HRTFs used in conventional rendering. The original MagLS approach [13, 15] compensates for the inherent SH order truncation of the employed HRTFs. The eMagLS method additionally compensates for practical limitations of the employed microphone array, namely spatial aliasing and regularization of the radial filters [6].

The eMagLS rendering filters are computed by mapping the array output for given impinging sound fields onto the corresponding ideal binaural signals. Typically, plane waves in a free field are considered for which the ideal binaural signals are the HRTFs for the incidence direction of the plane wave under consideration. A suitable set of plane waves for this has to cover various incidence directions so that the computation of the rendering filters is well conditioned. Below a transition frequency f_T, the rendering filters minimize the least-squares reproduction error considering magnitude and phase. Above f_T, accurate reproduction of the phase is neglected in favor of accurate magnitude reproduction and the rendering filters are computed such that only the deviation of the spectral magnitude of the binaural output signals from the ideal output is minimized in a least-squares sense [13, 6]. An implementation of eMagLS is available[1].

Distributing plane waves over the required incidence angles is straightforward for SMAs. It is also convenient for SMAs that rotations of the listener's head and changes in the incidence angle of a plane wave are equivalent, so the rendering filters automatically cover arbitrary head rotations. The challenge with EMAs is the inherent projection of the impinging sound field onto the horizontal plane. Independent of the incidence direction of a plane wave, an EMA outputs a SH representation of a horizontally propagating sound field. This does not constitute a limitation if one permits the user to perform head rotations only along the azimuth, as demonstrated in [6].

The computation of the eMagLS filters for EMAs is ill-conditioned if arbitrary head rotations are permitted. Exposing an EMA to plane waves produces a SH output that does not trigger non-horizontal information in the HRTFs. The only way to trigger sufficient information in the HRTFs so that general eMagLS rendering filters can be determined is by introducing head rotations during the computation. There are a variety of ways how this can be set up. We chose the following approach:

- For each sound incidence direction for which an HRTF in the considered set is available, we expose the EMA to a simulated plane wave that impinges on the array from a direction corresponding to the horizontal projection of the HRTF direction. We compute the SH representation that the array outputs without radial filtering for this case.
- We rotate the SH representation such that the relative incidence direction of the plane wave corresponds to the actual direction of the available HRTF. We use the SH rotations [7] in the implementation from[2] [10]. Explicitly, we rotate the array output about the z-axis such that the plane wave appears to impinge from straight ahead (the direction of the positive x-axis). We then rotate it about the y-axis to obtain the desired elevation angle. Finally, we rotate the array output about the z-axis back to its original azimuth angle.
- We do this for all available HRTFs. The result is a set of array output signals onto which the optimization objective for computing the eMagLS rendering filters [6, Eq. (12)] towards the corresponding ideal binaural signals can be applied. Analogous to the SMA solution, the least-squares objective is used at low frequencies below the eMagLS transition frequency.

Instrumental Evaluation

We evaluate the accuracy of the binaurally rendered ear signals from SMAs and EMAs, with and without eMagLS processing. We show results with horizontal and vertical head rotations for an anechoic environment based on a simulated plane wave impinging from the frontal direction on an SMA at $N = 44$ with[3] [10], as well as a room environment based on sequentially measured SMA impulse responses at $N = 29$ of the *Large Broadcast Studio* (LBS) from the WDR Cologne dataset [14]. Lower-order array data is accurately synthesized by subsampling the high-resolution SMA at the maximum available order in the SH domain, i.e., in our case SMAs and EMAs with radius $r = 8.75\,\mathrm{cm}$ at SH order $N = 2$ (cf. Fig. 1 for the respective sampling grids). This results in a spatial aliasing frequency of around 1.2 kHz, as approximated by $f_\mathrm{A} = c\,N \,/\, (2\pi\, r)$ [12], with the speed of sound in air c.

We render binaural ear signals for the *Neumann KU100* [4] dummy head HRTFs, using the SMA and EMA functions from[4]. Thereby, the radial filters for the raw renderings are Tikhonov-regularized [9] at 18 dB – corresponding to 40 dB for the convention of EMA radial filters as for-

[1] https://github.com/thomasdeppisch/eMagLS

[2] https://github.com/polarch/Spherical-Harmonic-Transform

[3] https://github.com/polarch/Array-Response-Simulator

[4] https://github.com/AppliedAcousticsChalmers/ambisonic-encoding

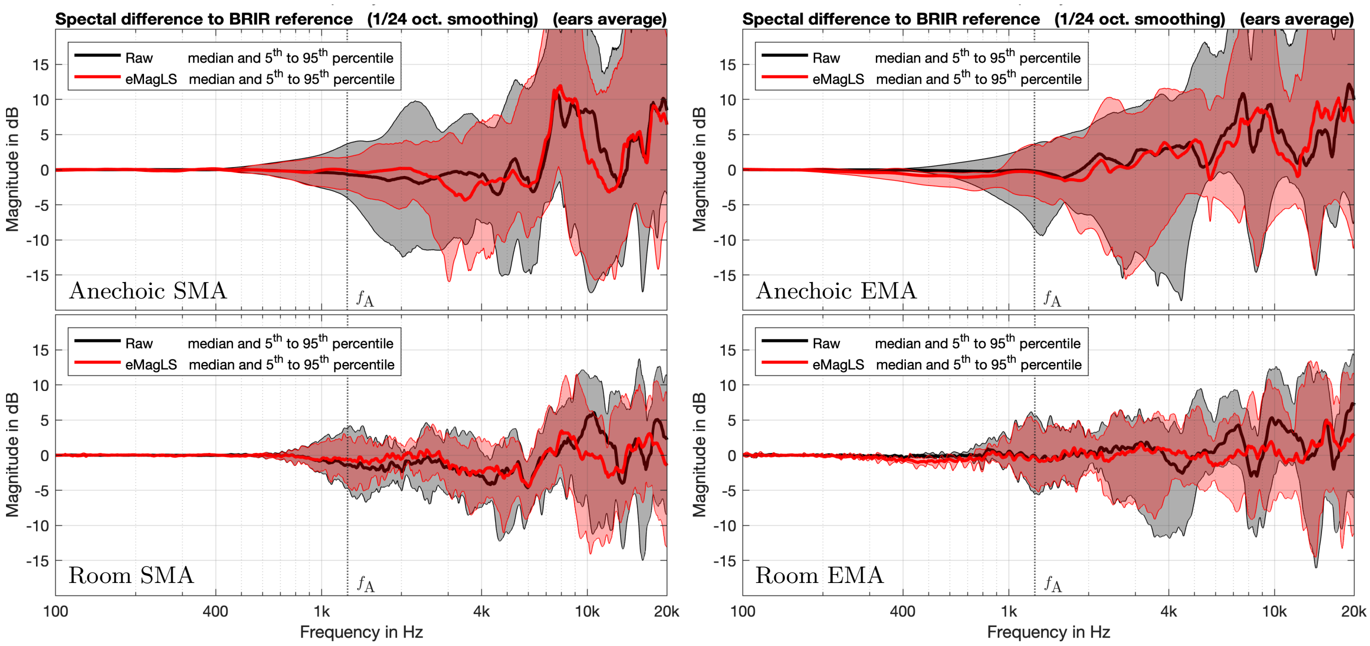

Figure 3: Spectral error during **horizontal head rotations** with raw (black) and eMagLS (red) rendering at SH order 2 for SMA (left) and EMA (right) in anechoic environment (top) and reverberant room environment (bottom).

mulated in [2, Eq. (18), Eq. (20)]. The eMagLS filters are generated with a target length of 512 samples and other parameters identical to the evaluation in [6]. All head rotations are performed with 3-DOF in the SH domain[2] [7].

We compare the rendered ear signals to a reference to summarize the strong direction and frequency-dependent errors in the reproduction from lower-order arrays. For all cases, the reference ear signals are generated from a high-resolution SMA rendering at $N = 29$ with identical head rotations and rendering parameters. We determine the average spectral difference for each configuration with the following procedure:

- Apply 1/24 fractional octave smoothing with[5] to the magnitude spectra of all binaural ear signals.
- Compute the spectral differences by dividing the smoothed magnitude spectra of the rendered signals by the reference ear signals.
- Compute the frequency-dependent median and 5th to 95th percentiles of the spectral differences for all head orientations (horizontal or vertical).
- Compute the depicted average result as the RMS of the percentiles from both ears.

Results for Horizontal Head Rotations

Fig. 3 shows the resulting spectral differences computed over a yaw head rotation from 0° to 360° in 1° steps. It confirms that the EMA and SMA outputs produce comparable errors in the raw rendering case (black). With eMagLS (red), both array types improve their overall spectral deviation (flatter median error) and direction-dependent deviations (narrower error percentiles). The improvement is more apparent in the room environment (bottom) than in the anechoic scenario (top) for the present array configurations. The response for SMAs (raw and eMagLS) strongly depends on the employed sampling grid, particularly at lower SH orders.

Note that the anechoic results from Fig. 3 (top) show median errors for a frontal plane-wave incidence direction and yaw-only rotations. The eMagLS renderer minimizes the average reproduction error for arbitrary (equally-distributed) incidence directions and head rotations. An evaluation of the SMA with arbitrary incidence directions would thus show reproduction errors for the entire frequency range close to 0 dB analogous to [6, Fig. 4].

The results from eMagLS for EMA with horizontal rotations in SHs exhibit slight deviations below f_A in both acoustic environments, as seen in Fig. 3 (right) from 200 Hz to 1 kHz, which are not present in the raw rendering or the previous solution in CHs [6].

Results for Vertical Head Rotations

Fig. 4 shows the resulting spectral differences computed over a pitch head rotation from 0° to 360° in 1° steps. Again, SMA and EMA show spectral differences compared to the reference above f_A of similar magnitude. The EMA's inherent horizontal projection does not impose a limitation on 3-DOF head rotations neither in the anechoic (top) nor the present room (bottom) scenario.

As for horizontal head rotations, eMagLS (red) improves the median and direction-dependent deviations compared to the raw rendering (black) for both array types during vertical head rotations. Thereby, EMAs show again slightly increased direction-dependent errors below f_A, as seen in Fig. 4 (right). Analysis of the spectral differences for individual head orientations reveals that broadband errors from 200 Hz to 2 kHz arise with increasing magnitude towards the ±90° head pitch directions. This suggests that the least squares implementation below f_T (1 kHz for the shown arrays at $N = 2$) is not yet optimal for EMAs.

[5] https://www.tu.berlin/ak/forschung/publikationen/open-research-tools/aktools

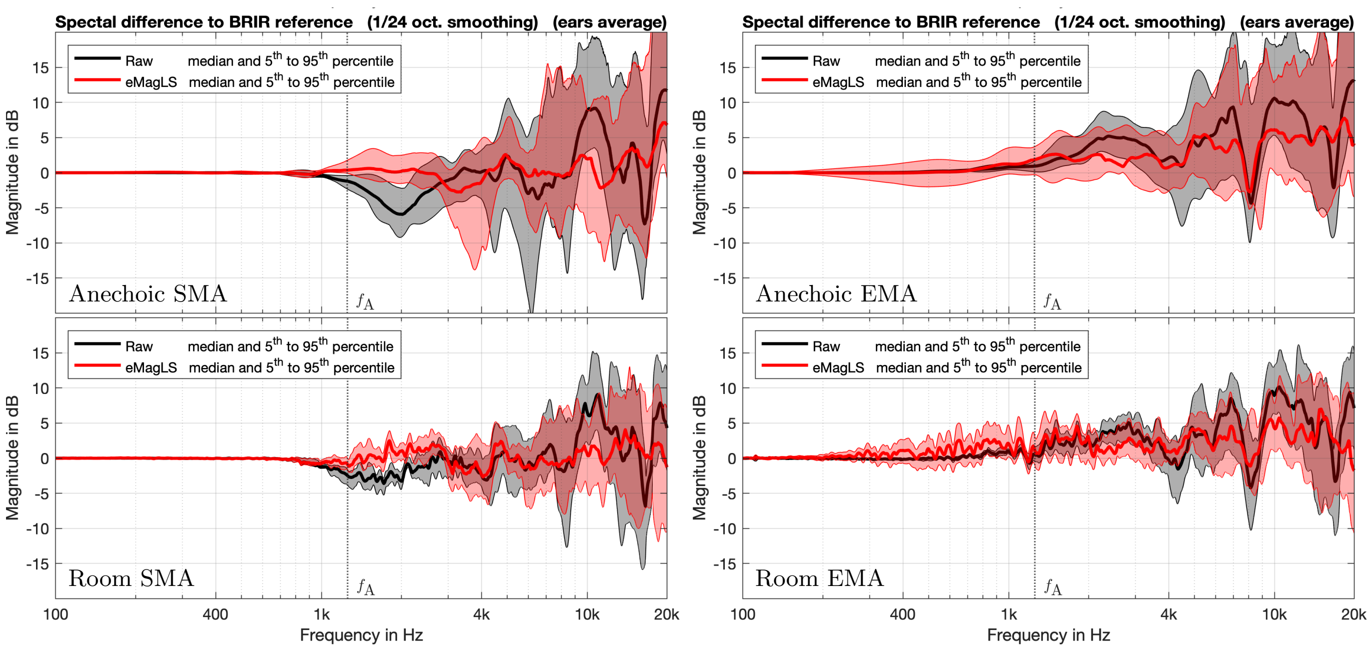

Figure 4: Spectral error during **vertical head rotations** with raw (black) and eMagLS (red) rendering at SH order 2 for SMA (left) and EMA (right) in anechoic environment (top) and reverberant room environment (bottom).

Conclusions

Practical acoustic scenarios often consist of primary sound sources close to the horizontal plane that excite early reflections and diffuse reverberation of the surrounding room from non-horizontal directions. In such applications, an EMA provides similar spectral accuracy as an equivalent SMA of the same SH order, whereby the direction-dependent spectral errors of both array types can be improved through eMagLS rendering (cf. Fig. 3). The resulting horizontal projection of the sound field from an EMA can be rotated in SHs with 3-DOF, providing binaural ear signals similar to an SMA for arbitrary listener head movements (cf. Fig. 4).

The GitHub repository containing the initially proposed MATLAB reference implementation of the eMagLS rendering method has been improved and extended by the variant for EMAs[1]. Static binaural listening examples of various room and array configurations at different SH orders are available[6].

Acknowledgement

We thank Reality Labs by Meta for funding this project.

References

[1] Jens Ahrens. Binaural Audio Rendering in the Spherical Harmonic Domain: A Summary of the Mathematics and its Pitfalls. Technical Report 1, arXiv, 2022.

[2] Jens Ahrens, Hannes Helmholz, David Lou Alon, and Sebastià V. Amengual Garí. Spherical Harmonic Decomposition of a Sound Field Based on Observations Along the Equator of a Rigid Spherical Scatterer. *J. of the Acou. Soc. of America*, 150(2):805–815, 2021.

[3] Zamir Ben-Hur, Fabian Brinkmann, Jonathan Sheaffer, Stefan Weinzierl, and Boaz Rafaely. Spectral Equalization in Binaural Signals Represented by Order-Truncated Spherical Harmonics. *J. of the Acou. Soc. of America*, 141(6):4087–4096, 2017.

[4] Benjamin Bernschütz. A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100. In *Fortschritte der Akustik – AIA/DAGA 2013*, pages 592–595, Meran, Italy, 2013. Deutsche Gesellschaft für Akustik.

[5] Fabian Brinkmann, Alexander Lindau, and Stefan Weinzierl. On the Authenticity of Individual Dynamic Binaural Synthesis. *J. of the Acou. Soc. of America*, 142(4):1784–1795, 2017.

[6] Thomas Deppisch, Hannes Helmholz, and Jens Ahrens. End-to-End Magnitude Least Squares Binaural Rendering of Spherical Microphone Array Signals. In *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pages 1–7, Bologna, Italy, 2021. IEEE.

[7] Joseph Ivanic and Klaus Ruedenberg. Rotation Matrices for Real Spherical Harmonics. Direct Determination by Recursion. *Journal of Physical Chemistry*, 100(15):6342–6347, 1996.

[8] Tim Lübeck, Hannes Helmholz, Johannes M. Arend, Christoph Pörschmann, and Jens Ahrens. Perceptual Evaluation of Mitigation Approaches of Impairments due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data. *Journal of the Audio Engineering Society*, 68(6):428–440, 2020.

[9] Sébastien Moreau, Jérôme Daniel, and Stéphanie Bertet. 3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of Spherical Microphone. In *120th Convention of the Audio Engineering Society*, pages 1–24, Paris, France, 2006. Audio Engineering Society.

[10] Archontis Politis. *Microphone array processing for parametric spatial audio techniques*. Phd thesis, Aalto University, 2016.

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

[12] Boaz Rafaely, Barak Weiss, and Eitan Bachmat. Spatial Aliasing in Spherical Microphone Arrays. *IEEE Transactions on Signal Processing*, 55(3):1003–1010, 2007.

[13] Christian Schörkhuber, Markus Zaunschirm, and Robert Höldrich. Binaural Rendering of Ambisonic Signals via Magnitude Least Squares. In *Fortschritte der Akustik – DAGA 2018*, pages 339–342, Munich, Germany, 2018. Deutsche Gesellschaft für Akustik.

[14] Philipp Stade, Benjamin Bernschütz, and Maximilian Rühl. A Spatial Audio Impulse Response Compilation Captured at the WDR Broadcast Studios. In *27th Tonmeistertagung – VDT International Convention*, pages 551–567, Cologne, Germany, 2012. Verband Deutscher Tonmeister e.V.

[15] Markus Zaunschirm, Christian Schörkhuber, and Robert Höldrich. Binaural Rendering of Ambisonic Signals by Head-Related Impulse Response Time Alignment and a Diffuseness Constraint. *J. of the Acou. Soc. of America*, 143(6):3616–3627, 2018.

[6] `http://www.ta.chalmers.se/research/audio-technology-group/audio-examples/daga-2023a/`