# Evaluation of Non-Spherical Scattering Bodies for Ambisonic Microphone Arrays

Downloaded from: https://research.chalmers.se, 2026-10-15 11:11 UTC

Citation for the original published paper (version of record):

Ahrens, J., Hu, Z. (2022). Evaluation of Non-Spherical Scattering Bodies for Ambisonic Microphone Arrays. Proceedings of the AES International Conference, 2022-August: 58-66

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Audio Engineering Society

Conference Paper

Presented at the 2022 International Conference on
Audio for Virtual and Augmented Reality
2022 August 15–17, Redmond, WA, USA

*This conference paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This conference paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.*

# Evaluation of Non-Spherical Scattering Bodies for Ambisonic Microphone Arrays

Jens Ahrens and Ziyi Hu

*Division of Applied Acoustics*
*Chalmers University of Technology*
*412 96 Gothenburg, Sweden*

Correspondence should be addressed to Jens Ahrens (jens.ahrens@chalmers.se)

**ABSTRACT**

The XMA was recently presented, which is a higher-order ambisonic microphone array with a non-spherical scattering body. The approach is compatible with the also recently presented equatorial microphone array so that also XMAs can be designed with the microphones distributed solely on a circumferential contour around the scattering body. This greatly reduces the required number of microphones compared to classical spherical microphone arrays that require the microphones to be distributed over the entire surface of the scatterer. The equatorial XMA has so far only been evaluated as a head-mounted array, i.e. with a human head as the baffle. Other form factors of a range of sizes are also of practical relevance, particularly those form factors of 360 cameras as these are capable of capturing a complete panoramic audio-visual experience from a first-person view when combined with an equatorial XMA. We present a set of simulations based on which we identify what spherical harmonic orders can be obtained with what accuracy for a set of convex scattering body geometries that are of relevance in the given context. We demonstrate that the shape of the body is not very critical, and even corners are possible. The main limitation is that small bodies do not allow for extracting higher orders at low frequencies.

## 1 Introduction

The ambisonic format represents spatial audio content via spherical harmonic (SH) expansion coefficients [1]. These can be stored and transmitted in time domain like any multichannel audio signal. The convenience of the ambisonic format is that it allows for head-tracked binaural reproduction over headphones as well as for reproduction over loudspeaker arrays.

Ambisonic scenes are typically recorded with 1st or 2nd-order near-coincident microphones or higher-order microphone arrays that employ microphones distributed over a rigid spherical baffle [2]. A small number of such spherical microphone arrays (SMA) are available commercially whereby some models also comprise a set of video cameras so that the complete audio-visual data are recorded. These are then either used for analysis of the scene [3] or can be reproduced

**Fig. 1:** Photograph of a Live Planet (left) and GoPro MAX (right) 360 camera

for entertainment, for example, via a virtual reality headset. A variant of the SMA is the equatorial microphone array (EMA) [4], which also employs a spherical baffle whereby the microphones are positioned along the equator of the baffle. The limitation of the EMA is that it produces an ambisonic representation of a horizontal projection of the captured sound field rather than an ambisonic representation of the captured sound field itself. The upside of the EMA is that it saves significantly on microphones for a given maximum SH order $N$[1], which allows for creating arrays with SH orders that are beyond of what is conceivable for SMAs.

The audio quality with playback over headphones becomes very good for SH orders of $N = 5$ or higher [5, 6, 7]. The reproduction over loudspeaker arrays can work well even with somewhat lower orders [1]. It will therefore be desirable in many situations to be able to capture content with 3rd or higher order so that a microphone array with a baffle is required.

There are situations in which the use of an ambisonic microphone array with a spherical baffle is inconvenient. Wearable arrays are an example, and it may be useful to integrate ambisonic microphone arrays into 360 cameras, which do often not have a spherical form factor. Cf. Fig. 1, which depicts two commercial 360 cameras with a non-spherical body.

We presented recently the XMA, which is a variant of both SMA and EMA, yet with a non-spherical baffle [8, 9]. So far, the XMA has only been used as a head-mounted microphone array, which can be integrated into, for example, augmented reality glasses. A demonstration of the XMA in combination with binaural reproduction is available at[2].

A method that has similar capabilities like the XMA in that it allows for ambisonic encoding of the signals from non-spherically baffled microphone arrays is [10]. It performs the encoding parametrically by separating diffuse components from non-diffuse components in a frequency dependent manner, which is contrary to the XMA that performs linear encoding. It is demonstrated in [10] that parametric encoding produces higher perceptual quality for the particular head-mounted array that was employed ibidem. This array used a non-equatorial layout so that some of the potential of linear encoding may have stayed untapped. The XMA was demonstrated to be robust against, for example, mismatch and displacement of the microphones [11]. According data are not available for parametric encoding yet.

It was shown in many locations in the literature that head-size SMAs work very well for binaural reproduction, e.g. [5]. So far, the XMA has also been used only with baffles of similar size. In the present paper, we evaluate the performance of baffles that are smaller than a human head and that depart more from a spherical shape (a sphere is likely to be the ideal shape for such a baffle). Of primary interest are the form factors of commercial 360 cameras, most of which exhibit rather limited spatial audio capabilities at this point. For the different form factors, we investigate via numerical simulations which SH order can be reliably retrieved from the microphone signals at what frequency.

## 2 The XMA

In this section, we present the XMA method from an conceptual point of view to an extent that is sufficient for appreciating the subsequent analysis. We refer the reader to [8, 9] for the technical details. A similar approach was applied to SMAs in [2, 12], to a planar concentric array in [13], and to arrays with non-spherical volumetric baffles in [10, 14, 15].

The starting point of the derivation of the XMA is the observation that each channel of the ambisonic signal produced by an SMA is a linear combination of the microphone signals. The weights can be derived analytically.

[1] EMAs require $2N+1$ microphones; SMAs require $(N+1)^2$.

[2] `https://youtu.be/OPWCXFbOFxU`

The XMA follows the exact same concept whereby the weights of the linear combination are determined from a calibration measurement rather than an analytical formulation. The calibration measurement is conducted such that the XMA is exposed to sound fields of which the corresponding SH coefficients are known. A simple example is a loudspeaker that radiates sound in an anechoic environment. If the XMA is positioned at a distance of more than, say, 1 m from the loudspeaker, it can be assumed that an ideal microphone array would produce the SH coefficients of a plane wave propagating in the according direction. This measurement is performed for a sufficient amount of incidence directions. The frequency-dependent weights (i.e. the filters) $\chi_{n,m}^{(q)}(\omega)$ that are a function of the microphone index $q$ and the SH mode $(n,m)$ are determined via a regularized least-squares fit that expresses the SH coefficients as a linear combination of the measured microphone signals. The filters $\chi_{n,m}^{(q)}(\omega)$ then transform any sound field that the XMA captures into an SH representation.

A difference between our formulation and the ones from [2, 12, 14, 15] is the circumstance that the referenced works comprise the radial filter gain limitation in the optimization procedure whereas our optimization projects the microphone signals onto an SH decomposition of the captured sound field on the surface of a notional rigid sphere. We apply the radial filtering in a separate step. Additionally, we employ an circumferential microphone layout.

A noteworthy convenience of the XMA is that the baffle does not need to be acoustically rigid or have otherwise analytically describable acoustic impedance. The acoustic properties of the baffle are automatically taken into account by the calibration procedure.

If the layout of the microphones of a given XMA is a circumferential contour (rather than a distribution over the surface), then the XMA can be seen as a variant of the EMA. In a nutshell, the EMA solution is formulated such that the array output is correct for the case that the captured sound field is height-invariant. The array output for height-variant sound fields is not physically correct but still useful for auralization purposes [4]. This means that the XMA should be calibrated only for horizontal sound incidence for a height-invariant field (such as the field radiated by a loudspeaker at sufficient distance in the horizontal plane). This makes the calibration measurement much more straightforward compared to non-equatorial layouts where the calibration measurement needs to comprise incidence direction that are distributed over all $4\pi$ sr. Fig. 2 depicts an example of an equatorial XMA.

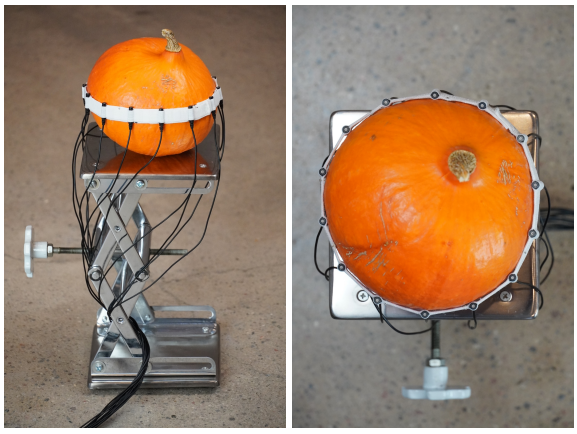

**Fig. 2:** Side view (left) and top view (right) of our very first prototype of an equatorial XMA based on a pumpkin. The array comprises 14 omnidirectional microphones and provides 6th ambisonic order.

We demonstrated in [8] that the horizontal projection of the captured sound field that is inherent to equatorial XMAs may be less of a limitation that what one may think at first sight. Even equatorial XMA preserve correct interaural elevation cues when the signals are reproduced binaurally. This can be sufficient to actually perceive elevation correctly when head-tracking is applied [16]. A demonstration of binaural rendering of an elevated sound source by the array from Fig. 2 is available at[3]. We are not aware that it has even been proven that conventional SMAs are able to preserve monoaural elevation cues given that the corresponding frequency range is typically corrupted by spatial aliasing.

## 3 Method

We will limit our considerations to equatorial XMAs because positioning microphones along a circumferential contour is much more convenient than having to distribute them over a surface as it would be required for SMA-like XMAs. All microphones need to be located inside a plane. This is a requirement that is a direct consequence of treating a horizontal XMA as a

[3] https://youtu.be/fY8rfushmwM

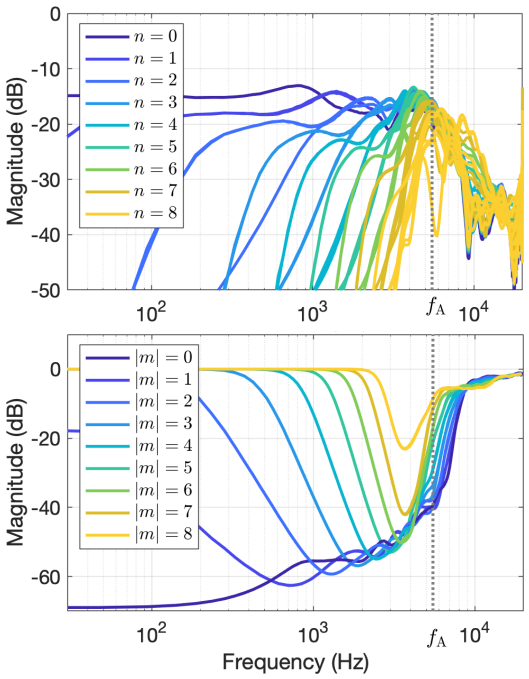


**Fig. 3:** Top: $20\log_{10}|\chi_{n,m}^{(q)}(\omega)|$ for a selected microphone of the equatorial XMA depicted in Fig. 4. Bottom: Normalized calibration error $E(\omega)$, Eq. (1), of that same XMA. The data are from [11].

variant of the EMA. The effect of deviations from this requirement are unclear at this point.

We used the *mesh2hrtf* implementation of the boundary element method from [17, 18] to simulate XMA calibration measurements for a variety of candidate baffles. We thereby simulated the microphone signals due to sound originating from point sources at different locations in the horizontal plane at a distance of 3 m. The XMAs that we simulated comprised rather high numbers of microphones. As we demonstrated in [11] and in the video in Footnote 2, binaural reproduction can work very well with only a handful of microphones. We still put more microphones than required for a good performance as this makes it easier to interpret the data visualizations.

We will employ graphs like the ones that are depicted in Fig. 3 in our analysis. The top plot shows the magnitude transfer function of the filters $\chi_{n,m}^{(q)}(\omega)$ that convert the microphone signals of the XMA depicted in Fig. 4

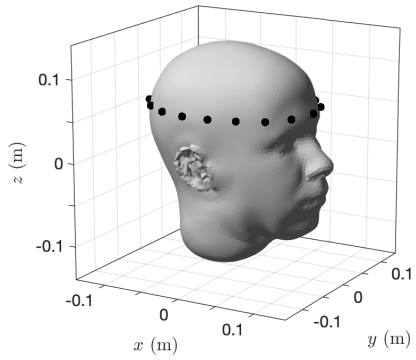


**Fig. 4:** The equatorial XMA whose data are depicted in Fig. 3. It comprises 18 omnidirectional microphones, which are marked by the black dots. The mesh of the head that wears the XMA is part of the mesh2hrtf package.

into an SH representation. Fig. 3 (bottom) depicts the normalized error $E(\omega)$ of the SH coefficients that the array computes, which we define as

$$E(\omega) = 20\log_{10}\frac{1}{L}\left|\sum_{l=1}^{L}\frac{\hat{\mathring{S}}_{n,m}^{(l)}(\omega) - \mathring{S}_{n,m}^{(l)}(\omega)}{\mathring{S}_{n,m}^{(l)}(\omega)}\right|. \quad (1)$$

$\hat{\mathring{S}}_{n,m}^{(l)}(\omega)$ are the SH coefficients that the XMA computes and $\mathring{S}_{n,m}^{(l)}(\omega)$ are the correct SH coefficients of the incident sound field. $l$ is the index of a total of $L$ horizontally propagating plane waves for which calibration data are available. We use $L = 100$ throughout this paper.

We use Fig. 3 to illustrate a set of important principles that are apparent in all data that we present in the evaluation in Sec. 4:

- In Fig. 3 (top), a line that is depicted in a given color can mask all lines of the same color that depict the data for all azimuthal modes $m$ that correspond to the indicated order $n$.
- Similarly, each line in Fig. 3 (bottom) masks the lines for all other orders $n$ that correspond to the indicated azimuthal mode $m$.
- In Fig. 3 (bottom), a normalized error of 0 dB means that the error is of the same magnitude like the data themselves. The corresponding SH coefficient is therefore corrupted. We propose to use as error threshold of $-40$ dB as a ballpark to

determine if a given SH coefficient was extracted reliably or not. This threshold is on the conservative side and corresponds to an error of 1 %.

- As evident from Fig. 3 (bottom), each angular mode $m$ has a sweet spot in terms of the frequency range in which it can be extracted with highest accuracy. Towards higher frequencies, spatial aliasing corrupts the computations (the spatial aliasing frequency $f_\mathrm{A}$ is indicated in Fig. 3, which we estimated via the well-known relation $N = \omega_\mathrm{A}/cR$ [19]. $R$ is the nominal radius of the baffle.) Towards lower frequencies, the limiting mechanism is the finite aperture of the array. If the wave length of the sound waves in air is much longer that the aperture, then the ability of the array to extract spatial information is limited. At very low frequencies, only the 0th mode, which represents purely omnidirectional information, can be extracted.

- The low-frequency end of an SH mode's sweet spot is higher for higher mode indices $|m|$. Although the error is high in Fig. 3 (bottom) for most modes at low frequencies, Fig. 3 (top) shows that the filters $\chi_{n,m}^{(q)}(\omega)$ actually attenuate the microphone signals at those frequencies. In other words, the error is high because the modes are not extracted from the microphone signals. The information is simply missing in the ambisonic representation. We deem this preferable compared to obtaining corrupted modes at these low frequencies.

- This is contrary to the frequency range above $f_\mathrm{A}$. There, the extracted SH coefficients do exhibit energy, but most of the spatial information is not correct. This has been shown in various locations in the literature to cause only a minor perceptual impairment if at all [7]. Usually, it is more beneficial to have spatially aliased higher orders available than not to have them at all because they help reducing unwanted angle dependencies in the reproduced signals.

## 4 Results

We chose not to present data on the array responses to sound incidence from non-horizontal directions. We found that the results are qualitatively and quantitatively similar to what we observed for EMAs in [20] and for head-mounted XMAs in [8]. I.e., the XMA outputs a horizontal projection of the captured sound field, and when binaural rendering is employed, the interaural cues are correct, and the magnitude of the binaural signals deviates from the correct one by not more than a few dB for frequencies below the spatial aliasing frequency of the array. Above the spatial aliasing frequency, the deviations are larger, but their perceptual impact seems small. Recall that a demonstration of binaural rendering of an elevated sound source by the array from Fig. 2 is available at Footnote 3.

Fig. 5 presents an overview of the baffle shapes that we considered. We will not present all data for all the shapes because of space constraints. We discuss some observations on a conceptual level such that the reader can apply the concepts on any shape of their interest.

All simulations were carried out at a sampling frequency of 32 kHz. The filters $\chi_{n,m}^{(q)}(\omega)$ were computed with a length of 1024 taps apart from the GoPro MAX inspired shape in Fig. 5 (bottom right), for which we used a length of 512 taps. All $\chi_{n,m}^{(q)}(\omega)$ were computed with Tikhonov regularization with the regularization parameter $= 1$.

### 4.1 No Baffle

Not using any baffle, i.e. using a *open* microphone array, is not an option because in this case, similar ambiguities arise like with SMAs without a baffle [21]. This prevents certain SH modes from being extracted at certain frequencies. Arrays with two or more layers have been proposed to mitigate this by switching between the layers depending on frequency and mode [21]. This type of microphone array never reached widespread use.

### 4.2 Size of the Baffle

Scaling the size of a baffle by a given factor inversely scales the frequency axis of $E(\omega)$ by the same factor. Applied to Fig. 3, this means that if we scale the head in Fig. 4 by a factor of 0.5, then all graphs in Fig. 3 (top) and (bottom) move upwards on the frequency axis by a factor of 2. We omit presenting a plot here that illustrates this.

Refer to Sec. 4.6 for data from a very small baffle.

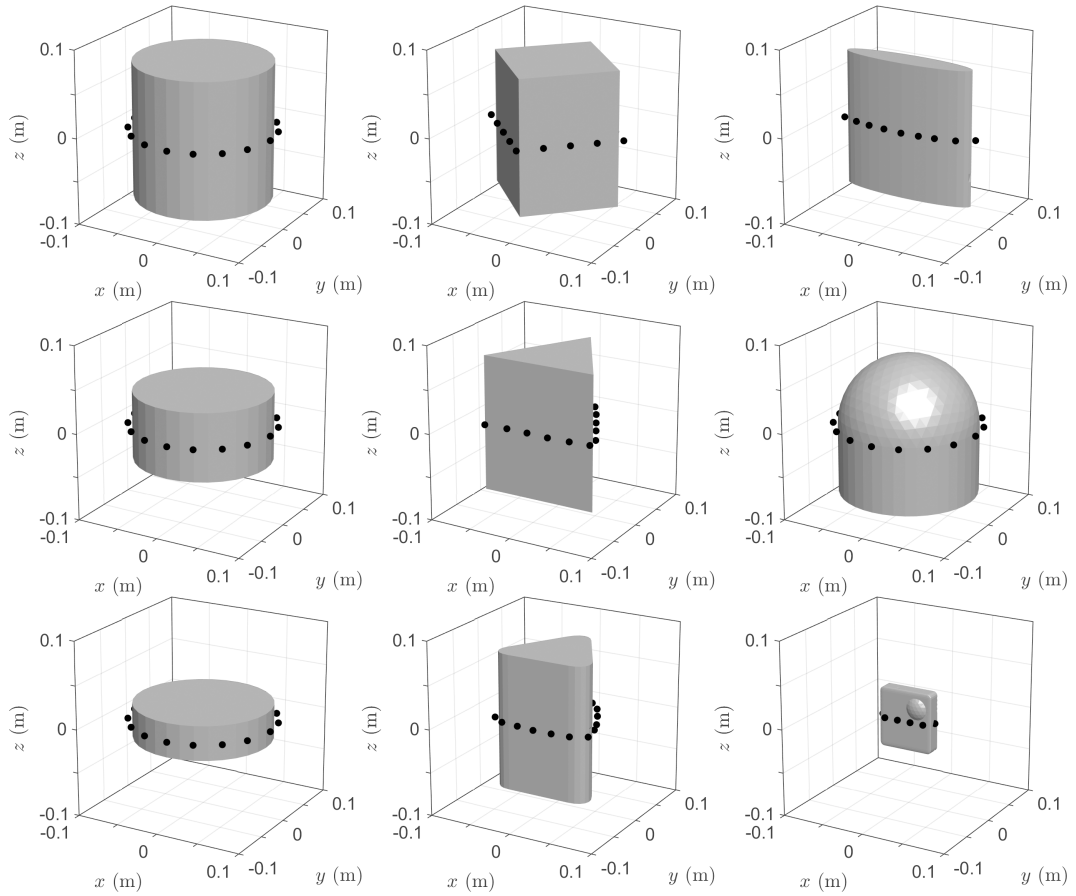


**Fig. 5:** The baffle shapes that we considered in our study. All arrays employ 17 omnidirectional microphones and produce 8th ambisonic order apart from the bottom right one, which employs 10 microphones and produces 4th ambisonic order. Left column: Cylinders with radius $r = 78\,\text{mm}$ and different height. The shape of the Live Planet camera from Fig. 1 (left) is comparable to the top-left shape but 30 % smaller. Middle column: Baffles with corners. Right column: Specialty shapes. The bottom right shape is inspired by the GoPro MAX 360 camera that is depicted in Fig. 1 (right).

### 4.3 Height of the Baffle

We use the example of a vertically positioned cylinder, Fig. 5 (left column), to illustrate the effect of the height of the baffle on the accuracy with which given SH modes can be extracted. We found that the accuracy that is provided by a spherical baffle in the XMA is not significantly higher compared to a cylindrical baffle. Above a certain height, it does not provide any advantage (and no disadvantage) to make the cylinder even higher. The lower the cylinder is, the more do the XMA's properties approach those of an open array.

Fig. 6 shows that $E(\omega)$ for a cylinder that has a height $h$ that is equal to half its radius $r$ (Fig. 5 (left column, bottom)) has more pronounced open-array-like peaks in the error compared to a cylinder with $h = r$ (Fig. 5 (left column, middle)).

Note that the shape of the commercial camera depicted in Fig. 1 (left) is similar to the cylinder depicted in Fig. 5 (top left) but 30 % smaller. The performance is similar to the one depicted in Fig. 6 (left) but without the hump in the 0th order around 1 kHz.

### 4.4 Shape of the Cross-Section

Changing the shape of the cross-section of the cylinder in Fig. 5 (left column, top) from circular to square or to

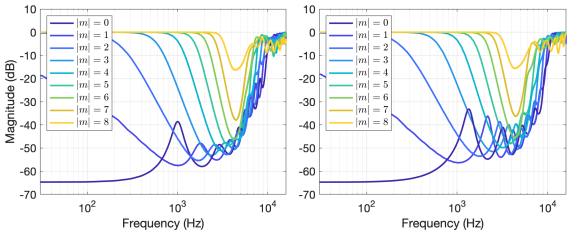


**Fig. 6:** $E(\omega)$ of a cylinder with radius $r = 78\,\text{mm}$ and different height $h$, cf. Fig. 5 (left column, middle and bottom). Left: $h = r$. Right: $h = 0.5r$.

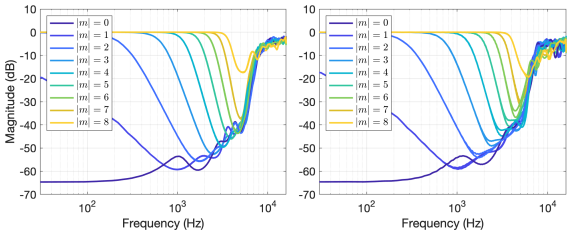


**Fig. 7:** $E(\omega)$ of the objects from Fig. 5 (middle column, top and middle). Left: Square cross-section. Right: Triangular cross-section with sharp corners.

triangular as in Fig. 5 (middle column, top and middle) does not make a significant difference in terms of $E(\omega)$ as depicted in Fig. 7. It is therefore not surprising that rounding the corners of the triangular cross-section as in Fig. 5 (middle column, bottom) does not have a noteworthy benefit either. We omit showing the data.

### 4.5 Vertical Position of the Microphones

In all examples presented so far, we mounted the microphones in the middle of the baffle in terms of the baffle's height, as it is depicted in Fig. 5. Fig. 8 demonstrates that the placement of the microphones is relatively uncritical unless the microphones are very close the upper or lower edge of the baffle where $E(\omega)$ exhibits peaks in the lower orders that are similar to what one observes with open arrays.

Positioning the microphones 20 mm from the upper edge hardly increases the error (Fig. 8 (left)). Positioning microphones exactly on the edge makes the problem somewhat less well conditioned and increases the $E(\omega)$ by a moderate amount (Fig. 8 (right)).

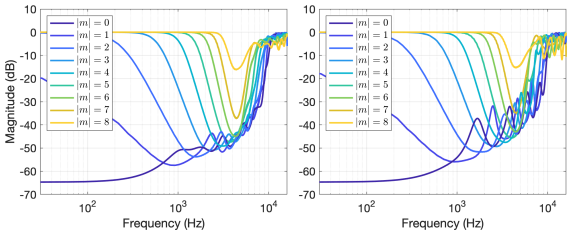


**Fig. 8:** $E(\omega)$ of the cylinder from Fig. 5 (left column, top) for different positions of the microphones. Left: Microphones positioned 20 mm below the upper edge. Right: Microphones positioned on the upper edge.

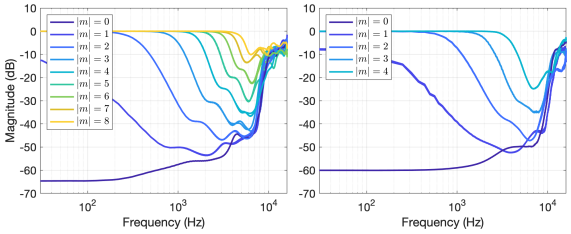


**Fig. 9:** $E(\omega)$ of two of the specialty shapes from Fig. 5 (right column). Left: Sideways squashed cylinder, cf. Fig. 5 (right column, top). Right: GoPro MAX, cf. Fig. 5 (right column, bottom)

### 4.6 Specialty Shapes

We employed the following specialty shapes in order to further identify limitations that may arise: 1) A cylinder that is squashed sideways whereby the thickness at the thickest part is 20 % of the height. 2) A cylinder with a dome. 3) A cuboid-like baffle that is inspired by the GoPro MAX commercial 360 camera that is 69 mm high and 16 mm deep. The geometries are depicted in Fig. 5 (right column).

The cylinder with the dome produces an error $E(\omega)$ that is equal to that of a regular cylinder. We therefore omit presenting the data here. A short spatial dimension as it occurs with the sideways squashed cylinder as well as with the GoPro-MAX-inspired baffle increases $E(\omega)$ by a moderate amount, cf. Fig. 9. It seems that an array aperture that is long in one of the Cartesian dimension is sufficient for maintaining acceptable performance. It was shown in [13] that even a thin plate can be sufficient of a baffle, though not with equatorial microphone layouts. The main limitations of the GoPro-MAX-inspired baffle arise due to its small size.

## 5 Discussion and Conclusions

We analyzed the normalized calibration error of higher-order microphone arrays with non-spherical baffles as a function of the baffle shape. The calibration error represents the error with which the spherical harmonic (SH) coefficients of the captured sound field are extracted from the microphone signals. We found that the upper frequency limit of the bandwidth with which SH coefficients can be extracted reliably is determined by spatial aliasing in a way that is very similar to how it is the case for spherical microphone arrays.

The lower limit of the frequency range with high accuracy is determined by the aperture of the array. It is sufficient for the aperture to be long only in one Cartesian dimension. Still, at very low frequencies, only the 0th order is available.

We applied a simple Tikhonov-regularized least-squares fit in this paper. It may be possible to increase the accuracy at low frequencies by some amount using the least-squares solution based on the singular value decomposition (SVD) proposed in [12], which allows for a more elegant regularization. This is difficult to anticipate because we perform the radial filter gain limitation in a separate step after the least-squares fit, which is contrary to the SVD solution from [12], which performs all steps at once.

It is unclear at this point how small such a baffle may be while still producing satisfactory perceptual results when the captured signals are auralized. The main mechanism to consider is the circumstance that higher-order spatial information cannot be extracted from an array with a small aperture. It was shown in [20] that, for arrays of bowling ball size that deliver 5th or higher order in the mid frequency range, the undesired loss of the spatial information at low frequencies is not audible when binaural playback is employed. The important aspect to investigate in the future will be at what point the loss of the spatial information will actually become audible in the different conceivable playback formats. This will allow for evaluating baffles for higher-order microphone arrays more reliably in terms of their suitability for spatial audio capture.

## References

[1] Zotter, F. and Frank, M., *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, Springer, Berlin, Heidelberg, 2019.

[2] Moreau, S., Daniel, J., and Bertet, S., "3D sound field recording with higher order ambisonics - objective measurements and validation of a 4th order spherical microphone," in *120th Convention of the AES*, Paris, France, 2006.

[3] O'Donovan, A., Duraiswami, R., and Zotkin, D., "Imaging concert hall acoustics using visual and audio cameras," in *IEEE ICASSP*, pp. 5284–5287, 2008.

[4] Ahrens, J., Helmholz, H., Alon, D., and Amengual Garí, S. V., "Spherical Harmonic Decomposition of a Sound Field Based on Observations Along the Equator of a Rigid Spherical Scatterer," *J. Acoust. Soc. Am.*, (150), 2021.

[5] Bernschütz, B., "Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording," PhD thesis, Technische Universität Berlin, 2016.

[6] Zaunschirm, M., Schörkhuber, C., and Höldrich, R., "Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *J. Acoust. Soc. Am.*, 143(6), pp. 3616–3627, 2018.

[7] Lübeck, T., Helmholz, H., Arend, J. M., Pörschmann, C., and Ahrens, J., "Perceptual Evaluation of Mitigation Approaches of Impairments due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data," *JAES*, 68(6), pp. 428–440, 2020.

[8] Ahrens, J., Helmholz, H., Alon, D. L., and Amengual Garí, S. V., "A Head-Mounted Microphone Array for Binaural Rendering," in *Int. 3D Audio Conference (I3DA)*, Bologna, Italy, 2021.

[9] Ahrens, J., Helmholz, H., Alon, D. L., and Amengual Garí, S. V., "Spherical Harmonic Decomposition of a Sound Field Based on Microphones Around the Circumference of a Human Head," in *IEEE WASPAA*, New Paltz, NY, USA, 2021.

[10] McCormack, L., Politis, A., Gonzalez, R., Lokki, T., and Pulkki, V., "Parametric Ambisonic Encoding of Arbitrary Microphone Arrays," *IEEE/ACM TASLP*, 30, pp. 2062–2075, 2022.

[11] Ahrens, J., Helmholz, H., Alon, D., and Amengual Garí, S. V., "Spherical Harmonic Decomposition of a Sound Field Using Microphones

on a Circumferential Contour Around a Non-Spherical Baffle," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2022, (submitted).

[12] Politis, A. and Gamper, H., "Comparing modeled and measurement-based spherical harmonic encoding filters for spherical microphone arrays," in *WASPAA*, pp. 224–228, 2017.

[13] Berge, S., "Acoustically Hard 2D Arrays for 3D HOA," in *AES Int. Conf. on Immersive and Interactive Audio*, Redmond, WA, USA, 2019.

[14] Tourbabin, V. and Rafaely, B., "Direction of Arrival Estimation Using Microphone Array Processing for Moving Humanoid Robots," *IEEE/ACM Trans. on Audio, Speech, and Lang. Proc.*, 23(11), pp. 2046–2058, 2015.

[15] Zotkin, D. N., Gumerov, N. A., and Duraiswami, R., "Incident field recovery for an arbitrary-shaped scatterer," in *IEEE ICASSP*, pp. 451–455, 2017.

[16] Ackermann, D., Fiedler, F., Brinkmann, F., Schneider, M., and Weinzierl, S., "On the Acoustic Qualities of Dynamic Pseudo-Binaural Recordings," *JAES*, 68(6), pp. 418–427, 2020.

[17] Ziegelwanger, H., Kreuzer, W., and Majdak, P., "Mesh2HRTF: Open-source software package for the numerical calculation of head-related transfer functions," in *22nd ICSV*, Florence, Italy, 2015.

[18] Ziegelwanger, H., Majdak, P., and Kreuzer, W., "Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization," *J. Acoust. Soc. Am.*, 138, pp. 208–222, 2015.

[19] Ward, D. B. and Abhayapala, T. D., "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Trans. on Speech and Audio Proc.*, 9(6), pp. 697–707, 2001.

[20] Ahrens, J. and Andersson, C., "Perceptual evaluation of headphone auralization of rooms captured with spherical microphone arrays with respect to spaciousness and timbre," *J. Acoust. Soc. Am.*, 145, pp. 2783–2794, 2019.

[21] Balmages, I. and Rafaely, B., "Open-Sphere Designs for Spherical Microphone Arrays," *IEEE TASLP*, 15(2), pp. 727–732, 2007.