CHALMERS
UNIVERSITY OF TECHNOLOGY

Spherical Harmonic Decomposition of a Sound Field Using Microphones on a Circumferential Contour Around a Non-Spherical Baffle

Citation for the original published paper (version of record):
Ahrens, J., Helmholz, H., Alon, D. et al (2022). Spherical Harmonic Decomposition of a Sound Field Using Microphones on a Circumferential
Contour Around a Non-Spherical Baffle. IEEE/ACM Transactions on Speech and Language Processing, 30: 3110-3119. http://dx.doi.org/10.1109/TASLP.2022.3209940

Spherical Harmonic Decomposition of a Sound Field Using Microphones on a Circumferential Contour Around a Non-Spherical Baffle

Jens Ahrens, Hannes Helmholz, David Lou Alon, Sebastià V. Amengual Garí

***Abstract*—Spherical harmonic (SH) representations of sound fields are usually obtained from microphone arrays with rigid spherical baffles whereby the microphones are distributed over the entire surface of the baffle. We present a method that overcomes the requirement for the baffle to be spherical. Furthermore, the microphones can be placed along a circumferential contour around the baffle. This greatly reduces the required number of microphones for a given spatial resolution compared to conventional spherical arrays. Our method is based on the analytical solution for SH decomposition based on observations along the equator of a rigid sphere that we presented recently. It incorporates a calibration stage in which the microphone signals due to a suitable set of calibration sound fields are projected onto the SH decomposition of those same sound fields on the surface of a notional rigid sphere by means of a linear filtering operation. The filter coefficients are computed from the calibration data via a least-squares fit. We present an evaluation of the method based on the application of binaural rendering of the SH decomposition of the signals from an 18-element array that uses a human head as the baffle and that provides 8th ambisonic order. We analyse the accuracy and robustness of our method based on simulated data as well as based on measured data from a prototype.**

***Index Terms*—Binaural rendering, spherical harmonics, microphone array, augmented reality**

I. Introduction

Content for virtual reality and augmented reality (AR) applications is typically captured with dedicated camera and microphone arrays. The consumer experiences the content from a first-person perspective whereby the audio signals are rendered binaurally. Spherical harmonics (SH) are a flexible basis for storage and transmission of the audio content as they allow for adapting the playback to rotations of the listener's head. The representation of spatial audio signals in terms of SH coefficients is also referred to as *ambisonics* format [1]. AR headsets on which such content is consumed may be equipped with outward facing cameras for enabling tracking of the user's head and body. These cameras may also be employed to record a 360 video of the wearer's environment so that the playback device can also function as a capture device. It has not been possible to integrate spatial audio capture with high spatial resolution into this form factor because the microphone arrays that have been available for this typically require being mounted onto a perfectly spherical scattering body [1], which is also referred to as *baffle*. This requirement is an obvious limitation.

In this article, we explore a method that follows a concept similar to conventional spherical microphone arrays (SMAs) in that a spatial transformation of the microphone signals is computed in order to obtain an orthogonal decomposition of the captured sound field and remove the effect of the baffle. Similarly to SMAs, we seek for an SH decomposition. Our method entails two major innovations compared to SMAs: 1) The shape of the baffle on which the microphones are mounted may depart from spherical and 2), we compute the spatial transform along a circumferential contour rather than across the entire spherical surface as with SMAs. The method enables new form factors for ambisonic microphone arrays including 360 video cameras [2] as well as using human heads as the baffle so that integration of such an array into AR glasses becomes possible. We will assume a head-shaped baffle in this article.

Head-mounted microphone arrays have been employed primarily for applications like beamforming, direction of arrival estimation, and noise suppression, particularly in the application area of hearing aids [3], [4]. Prediction of the binaural signals from microphones distributed over the head of a person was investigated, for example, in [3], [5], [6] whereby the wearer's orientation was encoded in the binaural signals. The methods from [7], [8] are able to produce head-tracked binaural signals from microphones on non-spherical baffles.

Contrary to previously proposed head-mounted microphone arrays, we aim at performing a decomposition of the sound field into SHs for being able to remove the scattering off the user's head and to compensate for head rotations during capture. This also allows for maintaining flexibility on the playback side in terms of the reproduction hardware [1].

SH decomposition of a sound field based on microphones densely distributed over the surface of a non-spherical scatterer is presented in [9], [10]. The SH representation of the captured sound field is obtained from the microphone signals by means of a linear filtering operation that is determined by means of a least-squares fit on calibration data. While distributing microphones over the entire surface of objects like, say, 360 camera arrays is conceivable, such a setup is less attractive for head-worn arrays as it would require a form factor that may be considered inconvenient.

A method that has similar capabilities like the presented one

Jens Ahrens and Hannes Helmholz are with Chalmers University of Technology, 412 96 Gothenburg, Sweden (e-mail: jens.ahrens@chalmers.se; hannes.helmholz@chalmers.se).

David Lou Alon and Sebastià V. Amengual Garí are with Reality Labs Research, 1 Hacker Way, Menlo Park, CA 94025 USA (e-mail: davidalon@fb.com; samengual@fb.com).

Manuscript received Mar. 29, 2022; revised Sep. 14, 2022.

in that it allows for ambisonic encoding of the signals from non-spherically baffled microphone arrays is [11]. It performs the encoding parametrically by separating diffuse components from non-diffuse components in a frequency dependent manner, which is contrary to the linear encoding performed in our method. It is demonstrated in [11] 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. Data on the robustness of parametric encoding are not available at this point. We demonstrate in Sec. V-B that our method is indeed robust against, for example, displacement of the microphones.

We presented an approach in [12], [13] that is an extension of the recently proposed equatorial microphone array (EMA) [14], which comprises microphones along the equator of a rigid spherical scatterer. To lift the requirement of the baffle to be spherical, we introduced a calibration stage into the EMA solution in which the microphone signals are projected onto the SH decomposition of the same sound field on the surface of a notional rigid sphere. The result is a method that obtains an SH decomposition of the captured sound field based on observations of the sound field on a circumferential contour around a scattering object the geometry of which may depart from spherical. Such a setup is convenient in many application scenarios such as the 360 camera arrays mentioned above, and it enables new applications such as integrating the microphones into an AR headset, which, together with an outward facing camera array, captures the audio-visual experience of the wearer from a first-person perspective.

The main limitation of circumferential arrays is the fact that they capture a horizontal projection of the impinging sound field rather than the impinging sound field itself. This does usually not constitute a limitation when binaural rendering of the captured sound field is targeted because the binaural elevation cues of interaural time difference (ITD) and interaural level difference (ILD) are preserved [12].

In this article, we revisit our approach from [12], [13] in Sec. II-IV. We present a thorough analysis of its properties including accuracy and robustness. The evaluation is performed both based on acoustic data that were simulated using the boundary element method (Sec. V) as well as based on data that we measured on a prototype head-mounted array (Sec. VI). We implemented the complete processing pipeline for binaural rendering including a method for equalizing the effects of the unavoidable SH order truncation (Appendix).

II. Spherical Microphone Arrays

SMAs are the standard setup for obtaining a SH decomposition of the captured sound field. We therefore review the underlying theory here. EMAs and the proposed head-mounted arrays may be considered derivatives of the general SMA concept.

SMAs typically employ pressure sensors distributed over an acoustically rigid spherical baffle. A sound pressure field $S^{\text{surf}}(\beta, \alpha, R, \omega)$ on the surface of such a scattering object of radius R that is centered at the coordinate origin is given by [15, Eq. (3.1.1)]

$$S^{\text{surf}}(\beta, \alpha, R, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathring{S}^{\text{surf}}_{n,m}(R, \omega)\; Y_{n,m}(\beta, \alpha)\;, \quad (1)$$

with

$$\mathring{S}^{\text{surf}}_{n,m}(R, \omega) = \breve{S}_{n,m}(\omega)\; b_n(R, \omega)\;, \quad (2)$$

and [15, Eq. (4.2.13)]

$$b_n(R, \omega) = -\frac{\mathrm{i}}{\left(\omega \frac{R}{c}\right)^2} \frac{1}{h_n'^{(2)}\left(\omega \frac{R}{c}\right)}\;. \quad (3)$$

$\mathring{S}^{\text{surf}}_{n,m}(R, \omega)$ are the SH coefficients of the sound pressure on the surface of the spherical scatterer. $\breve{S}_{n,m}(\omega)$ are the SH coefficients – and thereby a complete representation – of the incident sound field. $\omega = 2\pi f$ is the radian frequency in $\mathrm{rad/s}$, f is the frequency in Hz, c is the speed of sound in $\mathrm{m/s}$, and i is the imaginary unit. $h_n'^{(2)}(\cdot)$ denotes the derivative of the nth order spherical Hankel function of second kind with respect to the argument. $Y_{n,m}(\beta, \alpha)$ are the SH basis functions, which are dependent on colatitude β and azimuth α of a spherical coordinate system and are defined as [15, Eq. (2.1.59)]

$$Y_{n,m}(\beta, \alpha) = (-1)^m \sqrt{\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}}\, P_n^{|m|}(\cos\beta)\; \mathrm{e}^{\mathrm{i}m\alpha}\;. \quad (4)$$

$P_n^{|m|}(\cdot)$ are the associated Legendre functions.

The computation of $\breve{S}_{n,m}(\omega)$ from the signals $S^{\text{surf}}(\beta, \alpha, R, \omega)$ of pressure microphones on the baffle can be performed via [16]

$$\mathring{S}^{\text{surf}}_{n,m}(R, \omega) = \oint_O S^{\text{surf}}(\beta, \alpha, R, \omega)\; Y_{n,m}(\beta, \alpha)^*\, \mathrm{d}\Omega \quad (5)$$

and

$$\breve{S}_{n,m}(\omega) = \mathring{S}^{\text{surf}}_{n,m}(R, \omega)\; b_n^{-1}(R, \omega)\;, \quad (6)$$

or equivalently,

$$\breve{S}_{n,m}(\omega) = b_n^{-1}(R, \omega) \oint_O S^{\text{surf}}(\beta, \alpha, R, \omega)\; Y_{n,m}(\beta, \alpha)^*\, \mathrm{d}\Omega\;. \quad (7)$$

The asterisk * denotes complex conjugation, and $b_n^{-1}(R, \omega)$ is termed *radial filters* in the SMA literature. These filters exhibit impractically high gains at low frequencies at high orders (because $b_n(R, \omega)$ tends to 0 there) so that they require regularization. The effect of this is that those high SH orders n cannot be extracted from the microphone signals at low frequencies. It is well documented in the SMA literature [17] that this circumstance does not constitute a noteworthy limitation in practice as sound fields do typically not exhibit a considerable amount of energy in the aforementioned range.

In practical implementations, the integrals in (5) and (7) are approximated by summations over the microphone signals, which bounds the maximum order n that can be extracted to $n \leq N$. One speaks of an Nth order decomposition.

If the SH coefficients $\mathring{H}^{\text{L,R}}_{n,m}(\omega)$ of the user's left and right head-related transfer functions (HRTFs) are known, then

binaural rendering of the (order-limited) captured sound field can be performed using [18]

$$B^{\mathrm{L,R}}(\omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{1}{4\pi \mathrm{i}^{-n}} \breve{S}_{n,-m}(\omega, \Omega)\ \mathring{H}^{\mathrm{L,R}}_{n,m}(\omega)\ , \quad (8)$$

i.e., the signal $B^{\mathrm{L,R}}(\omega)$ that arises at a given ear of the listener if she/he is exposed to the captured sound field can be computed. We refer the reader to [19] for a summary of the peculiarities of (8).

III. Equatorial Microphone Arrays

The EMA was proposed in [14], which is a generalization of the solution from [20]. An EMA is essentially an SMA as described in Sec. II but with microphones placed solely along the equator of the scatterer. The EMA solution performs a circular harmonic decomposition of the captured sound field from which an SH representation is computed from which the effect of the scattering object is removed. The minimum number of microphones that are required for an Nth order decomposition is $2N{+}1$ for EMAs, which is opposed to $(N{+}1)^2$ for SMAs. At an order of $N = 8$, EMAs therefore require 17 microphones whereas SMAs require 81 microphones or more depending on their placement.

We omit details of the EMA solution as they are not of primary relevance for the present work. What is relevant here are the high-level conclusions that can be drawn from the EMA solution [14]:

An EMA cannot deduce all information on the captured sound field. The solution requires assumptions to be made to circumvent ambiguities. It turned out to be sufficient to design the array processing such that it computes the correct SH coefficients for height-invariant impinging sound fields. The consequence is that the array will always output a horizontally propagating sound field. If the impinging sound field is not height invariant, then the array outputs a horizontal projection of it.

The sound fields from compact sound sources at close distances to the array, which are not height-invariant, or the sound fields from sources that are located outside of the horizontal plane therefore produce undesired deviations of the array output. When binaural rendering is performed, the deviations in the binaural signals are in the order of a few dB or smaller at some frequencies in the range below the spatial aliasing frequency.

IV. Arrays with Arbitrarily-Shaped Scatterers

The SMA and EMA solutions are only applicable to arrays that comprise a spherical scattering object. We propose a solution in this section that applies the SMA and EMA concepts to arrays that comprise arbitrarily-shaped compact scatterers. We use the term *spherical XMA* (sXMA) for arrays that employ microphones that are distributed over the entire surface of the scatterer and the term *equatorial XMA* (eXMA) for arrays whose microphones are located along an equator-like contour.

The problem of recovering the incident sound field from observations of the sound field on the surface of an arbitrary scatterer can generally not be solved analytically. We therefore seek for a numerical solution similarly to [9], [10], where a numerical fit of filter coefficients onto a set of known microphone signals and the corresponding known SH coefficients of the incident field is performed. A similar approach was applied to SMAs in [21], [22] and to planar concentric arrays in [23].

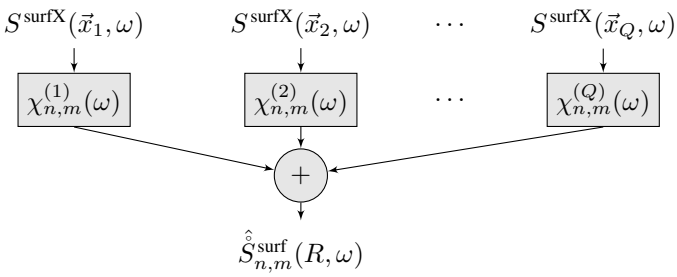

Fig. 1. Block diagram representation of (9)

Unlike previous approaches, we do not aim at extracting the SH coefficients $\breve{S}_{n,m}(\omega)$ of the impinging sound field directly because these generally diverge at low frequencies for sound fields from sources at finite distances so that the numerical solution may be ill conditioned. We rather propose to numerically extract $\mathring{S}^{\mathrm{surf}}_{n,m}(R, \omega)$ given by (2) as it is numerically well conditioned under all circumstances. In other words, we propose to project the pressure signal $S^{\mathrm{surfX}}(\vec{x}, \omega)$ at position $\vec{x}$ on the surface of the arbitrarily-shaped scatterer onto the SH coefficients of the pressure distribution that the same incident sound field would evoke on the surface of a virtual rigid spherical scatterer of radius R. $\breve{S}_{n,m}(\omega)$ can then be computed via (7) using the well-known gain-limited radial filters.

It seems intuitive that it will be favorable if the arbitrarily-shaped scatterer does not depart too much from spherical and if the diameter of the notional rigid sphere is chosen similar to the size of the XMA. Yet, we demonstrated in [2] that the proposed method works well even if the baffle has corners.

The extraction of $\mathring{S}^{\mathrm{surf}}_{n,m}(R, \omega)$ from the microphone signals $S^{\mathrm{surfX}}(\vec{x}, \omega)$ is a linear operation, which means that $\hat{\mathring{S}}^{\mathrm{surf}}_{n,m}(R, \omega)$ can be represented by a linear combination of $S^{\mathrm{surfX}}(\vec{x}, \omega)$ observed at different positions $\vec{x}_q$ as

$$\mathring{S}^{\mathrm{surf}}_{n,m}(R, \omega) \approx \underbrace{\sum_{q=1}^{Q} \chi^{(q)}_{n,m}(\omega)\ S^{\mathrm{surfX}}(\vec{x}_q, \omega)}_{=\hat{\mathring{S}}^{\mathrm{surf}}_{n,m}(R,\omega)}\ , \quad (9)$$

whereby $S^{\mathrm{surfX}}(\vec{x}_q, \omega)$ is the sound pressure on the surface of the arbitrarily-shaped scatterer at the position / microphone with index q. $\chi^{(q)}_{n,m}(\omega)$ are the complex weights of the Q microphone signals. A block diagram representation of (9) is depicted in Fig. 1.

With conventional SMAs, the coefficients $\hat{\mathring{S}}^{\mathrm{surf}}_{n,m}(R, \omega)$ can be computed in practice via quadrature of the integral in (5) as

$$\hat{\mathring{S}}^{\mathrm{surf}}_{n,m}(R, \omega) = \sum_{q=1}^{Q} w_q\ S^{\mathrm{surf}}(\beta_q, \alpha_q, \omega)\ Y_{n,m}(\beta_q, \alpha_q)^*\ , \quad (10)$$

whereby w_q are the quadrature weights of the Q microphone locations. Comparing (10) and (9) makes it obvious that, in the case of the conventional SMA,

$$\chi_{n,m}^{(q)}(\omega) = \chi_{n,m}^{(q)} = w_q \; Y_{n,m}(\beta_q, \alpha_q)^* \; . \tag{11}$$

For the XMA, we have to assume that the complex weights $\chi_{n,m}^{(q)}(\omega)$ are frequency dependent. We can obtain them for sets of (n, m, q) from a least-squares fit according to (9). This requires a set of microphone signals $S^{\text{surfX}}(\vec{x}, \omega)$ and the corresponding known coefficients $\mathring{S}_{n,m}^{\text{surf}}(R, \omega)$ for at least $Q+1$ different sound fields to establish an over-determined system of linear equations. These data can be obtained from calibration measurements of defined sound fields that impinge on the XMA. Once $\chi_{n,m}^{(q)}(\omega)$ is known, we can straightforwardly apply (9) to the microphone signals due to arbitrary incident sound fields to obtain an approximation $\hat{\mathring{S}}_{n,m}^{\text{surf}}(R, \omega)$ of their according SH coefficients. We found that Tikhonov regularization [24, Eq. (6.10)] in the least-squares fit can be very effective in increasing the robustness but is not imperative.

We propose to use plane waves that carry a unit-amplitude time-domain impulse as sound fields for this calibration due to the convenient implementation. $S^{\text{surfX}}(\vec{x}_q, \omega)$ is also referred to as *steering vector* of the array in this case. When assuming an XMA of the size of a human head, the impinging sound field due to a loudspeaker in a free field can be approximated by a plane wave if the distance between the loudspeaker and the XMA is at least $1\,\text{m}$ [25], [26]. Note that this procedure applies to sXMAs and eXMAs alike with the only difference that eXMAs should be calibrated only with horizontally propagating plane waves.

For a plane wave propagating into the direction (ϕ, θ) defined by colatitude ϕ and azimuth θ [15, Eq. (2.3.6)]

$$\mathring{S}_{n,m}^{\text{surf,pw}}(R, \omega) = 4\pi \text{i}^{-n} \; Y_{n,m}(\phi, \theta)^* \; b_n(R, \omega) \tag{12}$$

holds. When using (12) for calibration in (9), it is important to be aware of the implicit time reference that (12) comprises. Eq. (12) represents a spatio-temporal transfer function, which is the frequency-domain representation of the spatio-temporal impulse response $\mathring{s}_{n,m}^{\text{surf,pw}}(t)$. Eq. (12) implies that $t = 0$ is the moment when the notional planar wave front carrying a time-domain impulse, which $\mathring{s}_{n,m}^{\text{surf,pw}}(t)$ is the response to, passes the coordinate origin if no scattering object were present. The right hand side of (9) should be using the same time reference.

V. Simulation Results

We evaluate the proposed method in its eXMA variant as this one is more interesting from an application point of view and the properties of sXMAs can be directly deduced from it. We assume an eXMA that is composed of microphones that are mounted around the circumference of a torso-less acoustically rigid human head as depicted in Fig. 2. We used the *mesh2hrtf* implementation of the boundary element method (BEM) from [27], [28] to simulate the microphone signals due to sound originating from point sources at different locations. We obtained the head mesh from the same resource where its suitability for the BEM simulation was demonstrated.

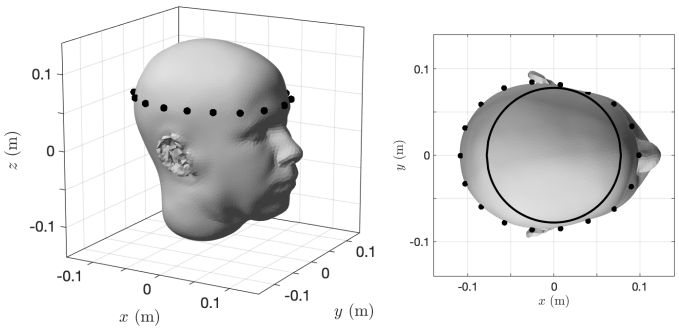

Fig. 2. Geometry of the eXMA on the acoustically rigid head "MH". The ears are located on the y-axis. The black dots denote the locations of the 18 microphones, which are located at $z = 60\,\text{mm}$. Left: Perspective view. Right: Top-down view. The black solid line denotes the virtual rigid sphere of radius $R = 78\,\text{mm}$ onto which the microphone signals are projected.

We made the following parameter choices:

- We employed 18 approximately evenly spaced microphones and target a sound field decomposition of an SH order of $N = 8$. An EMA of comparable geometry would require at least $2N+1 = 17$ microphones for this. As it is not straightforward to determine if the microphone grid maintains orthogonality of the implicit inherent circular harmonic decomposition, we chose to add one extra microphone to the minimum required number of 17. The spatial aliasing frequency f_A can be estimated via the relation $N = \omega_\text{A}/c\,R$ [29], which yields $f_\text{A} \approx 5.5\,\text{kHz}$ for the present case.
- We calibrated the eXMA via (9) using spherical waves that originated from 360 equal-angularly spaced locations in the horizontal plane at a distance of $3\,\text{m}$. This is a distance that is sufficient to assume that the impinging wave fronts are planar at the XMA so that (12) can be employed in the calibration. Recall that the minimum required number of calibration sound fields is $Q+1 = 19$ in the present case. We chose a higher number as well as to apply Tikhonov regularization with $\lambda = 1$ to increase the robustness.
- We added Gaussian noise with an RMS amplitude of $-80\,\text{dB}$ below the maximum amplitude of the impulse responses to both calibration and test data to emulate measurement errors and sensor self-noise.

A. Accuracy

1) Calibration: We analyze the accuracy of the calibration by analyzing the error between the left-hand side $\mathring{S}_{n,m}^{\text{surf}}(R, \omega)$ and the right-hand side $\hat{\mathring{S}}_{n,m}^{\text{surf}}(R, \omega)$ of (9). We define the normalized calibration error $E(\omega)$ as

$$E(\omega) = 20 \log_{10} \frac{1}{L} \left| \sum_{l=1}^{L} \frac{\hat{\mathring{S}}_{n,m}^{\text{surf},l}(R, \omega) - \mathring{S}_{n,m}^{\text{surf},l}(R, \omega)}{\mathring{S}_{n,m}^{\text{surf},l}(R, \omega)} \right| \; . \tag{13}$$

l is the index of a total of L sounds fields for which calibration data are available. We only evaluate $E(\omega)$ for horizontally propagating plane waves.

$E(\omega)$ is depicted in Fig. 3 (top) for $R = 78\,\text{mm}$, which is the largest radius of a sphere that is centered at the coordinate

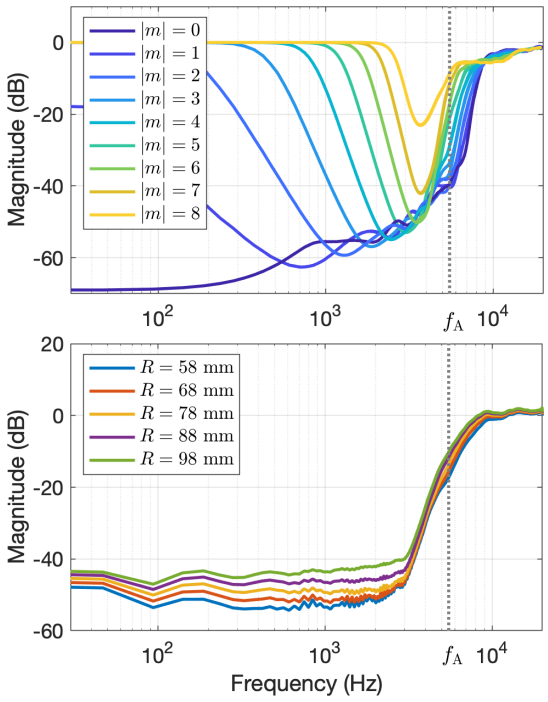

Fig. 3. Top: Normalized calibration error (13) as a function of frequency and for different $|m|$ for $R = 78\,\mathrm{mm}$. The error is independent of n for a given $|m|$. Bottom: Normalized error of the reconstruction of the sound pressure on the surface of the notional sphere due to the captured sound field for different radii R of the sphere. The error was averaged over 1891 positions distributed over the target sphere and over 100 incidence azimuths of horizontally propagating plane waves. The reference sound pressure distribution is computed from (1) with 35th order.

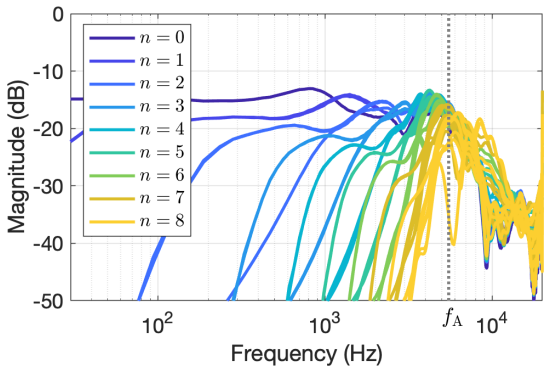

Fig. 4. $20\log_{10}\left|\chi^{(q)}_{n,m}(\omega)\right|$ for microphone q located at $(x, y) = (0.1, 0)$ m

origin that assures that the sphere fits entirely into the head (cf. Fig. 2 (right)). As expected, the error is large at frequencies above f_A. Fig. 3 (top) depicts also a large error for high $|m|$ at low frequencies, which will be discussed below.

Fig. 4 depicts $20\log_{10}\left|\chi^{(q)}_{n,m}(\omega)\right|$ for the present scenario. It indicates that this large error at low frequencies is due to the circumstance that the higher orders are not extracted from the microphone signals at low frequencies. Note that $20\log_{10}\left|\chi^{(q)}_{n,m}(\omega)\right|$ is very low there in Fig. 4. Recall from (11) that $\chi^{(q)}_{n,m}(\omega) = \chi^{(q)}_{n,m}$ is frequency independent for a conventional SMA.

The unavailability of high-order modes (n, m) at low frequencies is usually not a limitation in practice as these modes tend to exhibit very little energy. This aspect is confirmed by Fig. 3 (bottom) where the normalized error of the reconstruction of the captured sound pressure on the surface of the notional rigid sphere is depicted. We performed the reconstruction for a set of different projection radii R to also cover the effect thereof. The reconstruction error is very low below f_A.

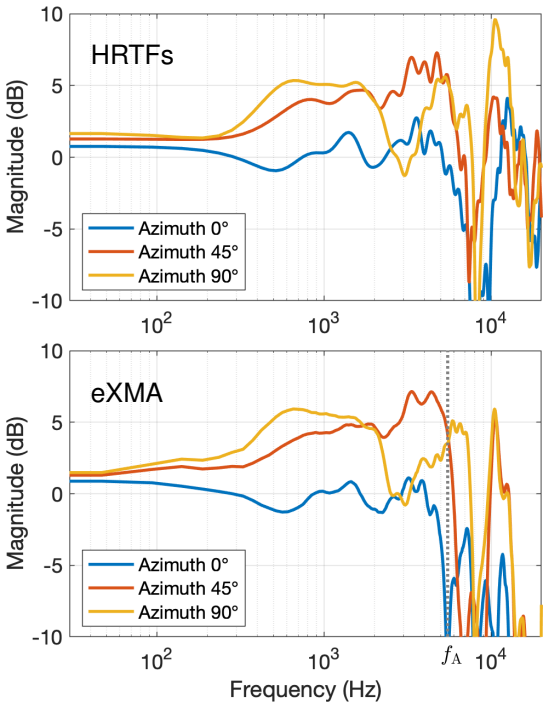

Fig. 5. Top: Magnitude of the left-ear HRTFs of the dummy head for different azimuth angles of horizontal sound incidence. Bottom: Magnitude of the left-ear BTFs of the eXMA for spherical waves originating from point sources located in the horizontal plane at the corresponding azimuth angles at a distance of 1 m.

The coefficients $\hat{\hat{S}}^{\mathrm{surf},l}_{n,m}(R,\omega)$ tend to exhibit more energy at high n (and thereby at high $|m|$) for sound fields that originate from sources that are close to the array. We will confirm in Sec. VI that the array output is accurate even in such situations.

Lastly, it is remarkable how clearly it is identifiable in Fig. 4 what frequency range a given SH order n contributes to primarily.

Fig. 3 (bottom) demonstrates that projection onto spheres of smaller radius R is more accurate than on larger R. This effect is small in magnitude and may be attributed to the fact that the (inaccurately deduced) higher orders n contribute less to the sound pressure if it is reconstructed on a smaller sphere. This apparent advantage of smaller projection spheres is rendered void by the inevitable gain limitation of the radial filters $b_n^{-1}(R,\omega)$ from (2) as this gain limitation causes higher inaccuracies with smaller spheres so that R may actually be chosen freely. It is rather an abstract parameter of the performed transformation and does not exhibit a physical meaning. We use $R = 78\,\mathrm{mm}$ in the remainder of this section.

2) Binaural output: In the following, we analyse the performance of the proposed eXMA based on its binaural transfer

functions (BTFs). The BTFs are given by the binaural output of the array processing pipeline in the case that the array captures a given sound pressure field like a plane wave that carries a time-domain Dirac impulse. In the remainder, we use the HRTFs of a *Neumann KU100* dummy head from [17] for the binaural rendering of the signals from the eXMA via (8).

Fig. 5 juxtaposes the left-ear HRTFs of the dummy head for sound incidence directions straight ahead, $45°$ to the left, and $90°$ to the left with the according BTFs of the eXMA for point sources located at the corresponding azimuth angles at a distance of $1\,\mathrm{m}$. A perfect eXMA would produce BTFs that would be essentially identical to the dummy head HRTFs.

It is worth noting that the HRTFs that we employed in the rendering were measured from a distance of $3\,\mathrm{m}$ whereas the sources that we used for testing of the eXMA were located at a distance of $1\,\mathrm{m}$. This means that even an ideal eXMA would theoretically produce BTFs that deviate from the reference HRTFs. It is evident from the data from [25], [26] that these deviations are negligible for the source distances that we employ. We made this choice in order to employ data in the evaluation that are not identical to the data used for the calibration.

It can be deduced from Fig. 5 that the BTFs of the eXMA are similar to the corresponding HRTFs up to a frequency of approx. $6\,\mathrm{kHz}$ whereby deviations in the order of $1\,\mathrm{dB}$ to $2\,\mathrm{dB}$ arise.

Significant attenuation of the eXMA BTFs compared to the HRTFs occurs above $6\,\mathrm{kHz}$. This is due to the truncation of the SH series in (8), and is well-known from the literature on SMAs. A range of methods have been proposed to equalize this for SMAs most which have been shown to be effective [30]. Such methods are not necessarily applicable with XMAs as the properties of the scattering body as well as the consequences of the microphone placement can vary greatly. We present results in Sec. VI that employ an equalization method that we proposed in [31] that is flexible enough to be applicable with all microphone array types that have been discussed in this article – SMAs, EMAs, sXMAs, and eXMAs.

The BTFs of the eXMA change with source elevation, and they also deviate from the according HRTFs as evident from Fig. 6. This is expected and is qualitatively and quantitatively similar to the deviations that an EMA produces in the same situation [14]. It is worth noting that the deviations of the BTFs change continuously with the elevation of sound incidence, which tends to be less disturbing from a perceptual point of view than erratic changes.

While the eXMA is not capable of producing monaural elevation cues, ITD and mostly also ILD are preserved even for non-horizontal sound incidence as evident from Fig. 7. This is a direct consequence of the fact that the eXMA outputs a horizontal projection of the captured sound field. In the context of head-tracked pseudobinaural rendering where the signals from microphones on the surface of a rigid sphere of head size are directly played to the listener without further processing, it was demonstrated that correct interaural auditory localization cues can lead to correct perception of elevation even in the complete absence of monaural cues [32].

Close sound sources often constitute a critical situation for microphone array processing methods as such sound fields have more energy in higher modes (n, m) particularly at low frequencies compared to far sources. Many times, intolerable amplification of the low-frequency output occurs if the short source distance violates assumptions of the processing method. We observed this in [20] where we proposed an earlier EMA

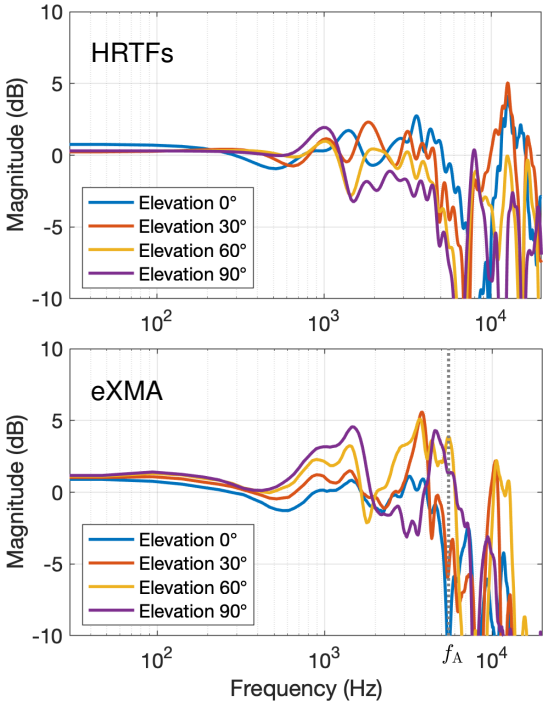

Fig. 6. Same as Fig. 5 for sound incidence from straight ahead from different source elevations. Top: Dummy head HRTFs. Bottom: eXMA BTFs.

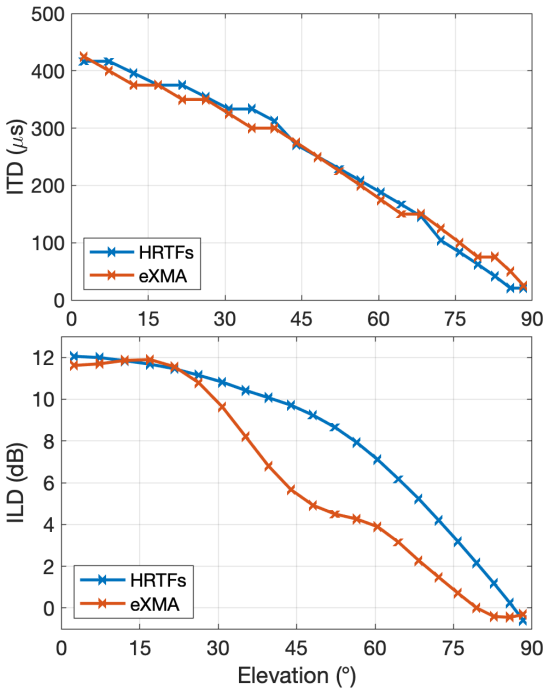

Fig. 7. ITD and ILD as a function of elevation for a source at azimuth $45°$. Top: ITD defined as the lag at which the maximum interaural correlation occurs. Bottom: ILD computed in the frequency band $1\,\mathrm{kHz}$ to $8\,\mathrm{kHz}$.

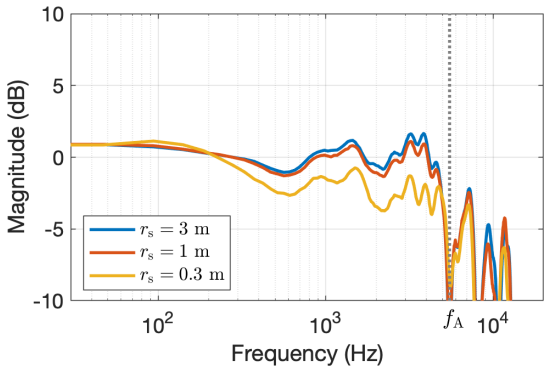

Fig. 8. Left-ear BTFs of the eXMA for a point source in the horizontal plane at different distances. Data are normalized with the source distance from the array center. Reference HRTF data are not available.

solution in which the scattering due to the array baffle was approximated rather than modeled accurately. The inaccuracies of this approximation produced a massive low-frequency boost for sources closer than, say, 1 m.

The sound fields due to close sound sources do indeed violate the assumption of height invariance that underlies EMAs and eXMAs. We demonstrated in [14] that the EMA is robust towards this violation due to close sources, and Fig. 8 demonstrates that also our eXMA solution is robust towards this. A minor change of the spectral balance is apparent particularly for the closest source distance of $r_s = 0.3$ m compared to the farther distances. This is expected and is a general phenomenon with the scattering off rigid bodies [25]. SMAs and EMAs produce BTFs with very similar dependencies [14].

B. Robustness

1) Microphone placement: To evaluate the robustness of the proposed solution regarding the placement of the microphones, we calibrated the array as described above and then evaluated the BTFs of the array after applying a random displacement of the microphones tangentially to the head surface taken from normal distributions with zero mean and standard deviations σ of 1 mm, 3 mm, 10 mm, and 30 mm. Cf. Fig. 9 (top) for an illustration of the geometry. All microphone transfer functions where again computed with *mesh2hrtf*.

For convenience, we depict only the results for $\sigma = 10$ mm in Fig. 9 (bottom). This figure depicts the difference between the BTFs when the array is evaluated with the displaced microphones and the BTFs when the array is evaluated with the microphones at the same positions like during calibration. We found that the deviations of the BTFs are smaller for smaller displacements and larger for larger displacements than the depicted one so that we omit presenting more detailed data here. It is evident from Fig. 9 (bottom) that despite the considerable displacement of the microphones, the BTFs are only moderately affected below f_A. Large deviations arise above f_A whereby this is not very critical because we found that the microphone displacement causes primarily a change in the fine structure of the spatial aliasing. This shows as large errors but seems perceptually hardly relevant as we found

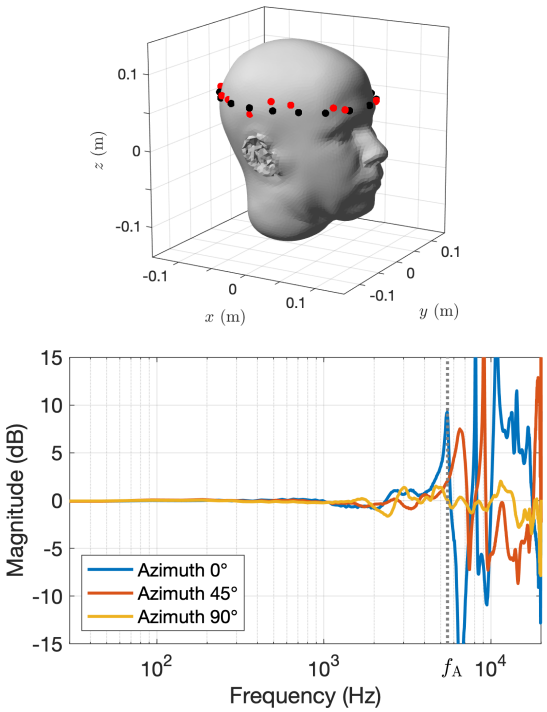

Fig. 9. Top: Head MH from Fig. 2 with the microphone positions during calibration (black dots) and the microphone positions during evaluation (red dots) for $\sigma = 10$ mm. Bottom: Difference between the left-ear BTFs of the eXMA when evaluated using the red-marked microphone positions and the left-ear BTFs of the eXMA when evaluated using the black-marked microphone positions.

during informal listening. The observations are very similar also for non-horizontal sound incidence.

2) Calibration: We simulated four different heads and tested in how far a deviation of the geometry of the baffle at evaluation affects the BTFs.

Head CT (from [33]) depicted in Fig. 10 is approx. 15 % smaller than head MH used above. The microphone positions were chosen such that the microphones occur at the same azimuth angles for both heads. We also produced scaled versions of each of the heads. In the remainder, CT(s) – the *small* version of CT – refers to head CT in its original size, CT(l) – the *large* version of CT – refers to head CT scaled to the original size of MH, MH(s) refers to head MH scaled to the original size of head CT, and MH(l) refers to head MH in its original size. Cf. Fig. 11 (top). The intention was to investigate the influence of the head shape and the head size separately.

We selected each of the four heads for calibration and performed the evaluation on the remaining heads. We found no systematic tendencies that would allow for differentiating the influence of the head size vs. the head shape. Both head size differences and head shape differences produce deviations of the BTFs of similar magnitude. Fig. 11 (bottom) depicts representative results. We omit presenting further results here as they are very similar of all combinations of heads that we tested.

The data in Fig. 11 (bottom) are similar to the data from Fig. 9 (bottom): Below f_A the deviations are moderate, and

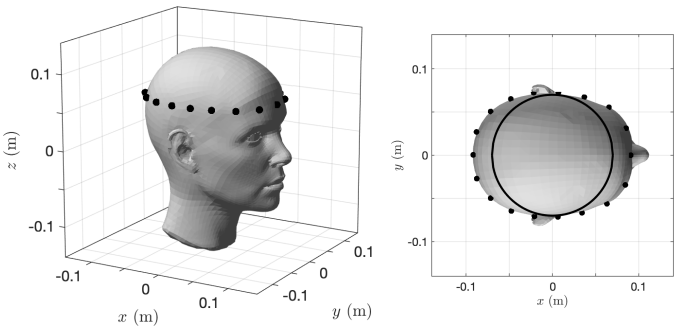

Fig. 10. Geometry of head CT, which is approx. 15 % smaller than head MH from Fig. 2. We use $R = 70\,\mathrm{mm}$ for CT.

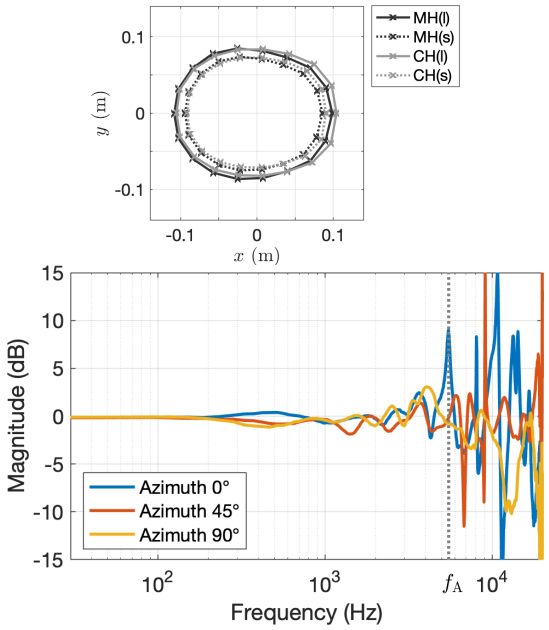

Fig. 11. Top: Top-down view on the microphone contours of all four heads that are employed in the robustness analysis. Bottom: Difference between the left-ear BTFs when head MH(l) is used for calibration and CT(s) for evaluation and when head MH(l) is used for both calibration and evaluation.

the fine structure of the spatial aliasing changes. This suggests that it may be tolerable to calibrate the array on a different head that on which is being used.

VI. Measurement Results

We built a prototype based on a wooden artificial head with geometrical dimensions that were comparable to head MH from Fig. 2. Photographs of the prototype are depicted in Fig. 12. We deployed 18 small pressure microphones along a horizontal contour with approximately constant spacing between them. The prototype was calibrated in the anechoic chamber at the Division of Applied Acoustics at Chalmers University of Technology using a loudspeaker in the horizontal plane at a distance of $3.0\,\mathrm{m}$ at 360 azimuths. The loudspeaker was still, and the prototype was rotated by means of a turntable.

Measured BTFs are presented in Fig. 13. We deployed the complete end-to-end processing pipeline as it is described

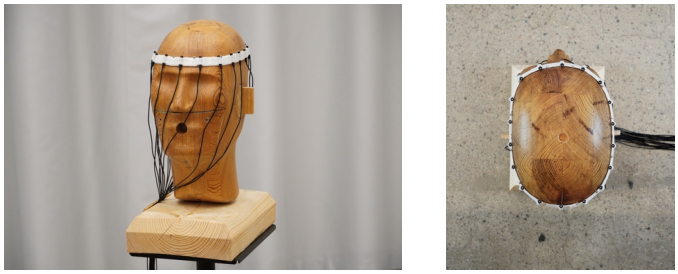

Fig. 12. Perspective view (left) and top-down view (right) of the prototype. The head dimensions are $14.5\,\mathrm{cm}$ (width) and $21.5\,\mathrm{cm}$ (length).

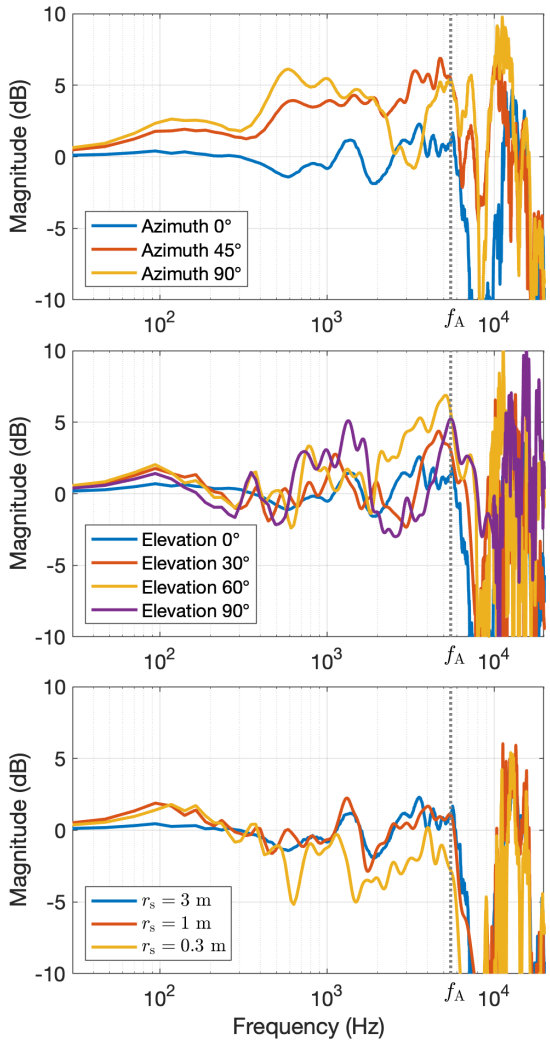

Fig. 13. Measured BTFs of the eXMA from Fig. 12. The BTFs are equalized according to (15) and (16). Top: Horizontal sound incidence from different azimuths (equivalent to Fig. 5). Middle: Frontal sound incidence from different elevations (equivalent to Fig. 6). Bottom: Frontal sound incidence from different distances (data are normalized with the source distance, equivalent to Fig. 8).

in Sec. II-IV including equalization of the BTFs according to [31]. The equalization method is outlined in the Appendix.

The measurement data in Fig. 13 exhibit very similar properties like the corresponding simulated data from Fig. 5, 6, and 8. Moreover, the measurement data for horizontal sound incidence in Fig. 13 (top) deviate from the reference HRTFs

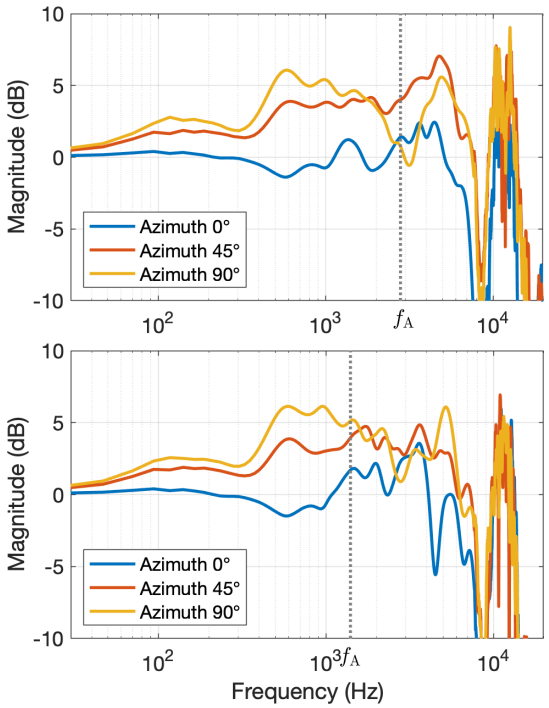

Fig. 14. Same as Fig. 13 (top) but for a total of 9 microphones (top) and 6 microphones (bottom). The corresponding SH orders are 4 ($f_A \approx 2.7\,\mathrm{kHz}$) and 2 ($f_A \approx 1.4\,\mathrm{kHz}$) respectively.

depicted in Fig. 5 (top) – the HRTFs that are used for the rendering – by less than 2 dB below f_A and somewhat more above f_A whereby the global spectral balance is maintained well. The measurement data for elevation and distance in Fig. 13 (middle) and (bottom) also exhibit the same tendencies like the simulated data but with slightly larger deviations. We attribute these deviations to the measurement setup, which was not free of reflecting surfaces like the turntable on which the prototype was positioned, a multichannel pre-amplifier that was positioned close to it, as well as rigging frames for mounting the loudspeaker. Some of the wigglyness of the yellow curve in Fig. 13 (bottom) may be attributed to reflections bouncing between the loudspeaker and the head. The loudspeaker baffle was positioned only approx. 15 cm from the head's face.

It is difficult to state at this point how many microphones are sufficient with the present setup to achieve a perceptually viable binaural output. The physical accuracy is closely linked to the spatial aliasing frequency f_A through the number of microphones. It was demonstrated to various locations in the literature including [34], [30] that physical accuracy is not imperative above approx. 2 kHz. Careful equalization of the binaural signals above 2 kHz using, for example, magnitude least-squares in its original formulation [35] or in our formulation presented in the Appendix can compensate perceptually for a lack of physical accuracy. Fig. 14 indicates this for our prototype. It depicts data equivalent to Fig. 13 (top) whereby we used only every other or every third of the microphones to obtain a total of 9 and 6 microphones, respectively. The deviations of the magnitude BTFs are small to moderate.

A demonstration of binaural rendering of signals captured with the prototype from Fig. 12 with different microphone counts is available at[1].

VII. Conclusions

We presented a proof-of-concept for a head-mounted circumferential microphone array that performs a spherical harmonic decomposition of the captured sound field. The evaluation of binaurally rendered signals from the array showed that the accuracy is only slightly lower than that of a comparable equatorial microphone array with a spherical scatterer. We confirmed via analysis of accuracy and robustness based on simulations as well as via measurements on a 18-element prototype array mounted around a human head that the method is deployable in practice.

Besides the head-mounted form factor, the presented method is applicable in all scenarios where a spherical array baffle is inconvenient and where the microphones cannot be distributed over the entire surface of the baffle but need to be confined to a circumferential contour.

Appendix

To equalize the spectral balance of the array's BTFs, we introduce the modal equalization filter $\breve{E}_{n,m}(\omega)$ into (8) as [31]

$$B^{\mathrm{L,R}}_{(\mathrm{Eq})}(\omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{1}{4\pi \mathrm{i}^{-n}}\, \breve{E}_{n,m}(\omega)\; \hat{\breve{S}}_{n,-m}(\omega,\Omega)\; \mathring{H}^{\mathrm{L,R}}_{n,m}(\omega)\,. \tag{14}$$

We design $\breve{E}_{n,m}(\omega)$ such that it minimizes the error between the chosen HRTFs $H^{\mathrm{L,R}}(\omega,\Omega)$ and the BTFs $B^{\mathrm{L,R}}_{(\mathrm{Eq})}(\omega)$ of the array pipeline given by (14) for a plane wave from incidence direction Ω in a least-squares sense, whereby in the general case, we minimize across all incidence directions $\Omega \in S^2$ with S^2 denoting the unit sphere and $\|\cdot\|$ denoting the l_2 norm:

$$\breve{E}^{\mathrm{LS}}_{n,m}(\omega) = \arg \min_{\breve{E}_{n,m}(\omega)} \oint_{\Omega\in S^2} \Big\| H^{\mathrm{L,R}}(\omega,\Omega) - \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{\breve{E}_{n,m}(\omega)}{4\pi \mathrm{i}^{-n}}\; \hat{\breve{S}}^{\mathrm{pw}}_{n,-m}(\omega,\Omega)\; \mathring{H}^{\mathrm{L,R}}_{n,m}(\omega,\Omega) \Big\|^2 \mathrm{d}\Omega\,. \tag{15}$$

In the case of a circumferential microphone contour like in the present case, the integral in (15) reduces to an integral along the azimuth in the horizontal plane. $\hat{\breve{S}}^{\mathrm{pw}}_{n,m}(\omega,\Omega)$ are the SH coefficients of a plane wave with incidence direction Ω that are computed from the corresponding array signals. The integral over Ω has to be discretized in practice, which does not pose any restrictions so long as the discretization is sufficiently dense.

The problem in (15) can only be solved with high accuracy at low to moderate frequencies. Given that maintaining the spectral magnitudes of the BTFs is perceptually much more important than maintaining the phase at high frequencies, it was proposed in [34], [35] to minimize the *magnitude* error

[1] https://youtu.be/OPWCXFbOFxU

in an LS sense above a cutoff frequency $f_\mathrm{c} \leq f_\mathrm{A}$. Adopted to our proposed formulation, this reads

$$\breve{E}_{n,m}^{\mathrm{MagLS}}(\omega) = \arg \min_{\breve{E}_{n,m}(\omega)} \oint_{\Omega \in S^2} \Big\| \big| H^{\mathrm{L,R}}(\omega, \Omega) \big| - \left| \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{\breve{E}_{n,m}(\omega)}{4\pi \mathrm{i}^{-n}} \; \hat{S}_{n,-m}^{\mathrm{pw}}(\omega, \Omega) \; \mathring{H}_{n,m}^{\mathrm{L,R}}(\omega, \Omega) \right| \Big\|^2 \mathrm{d}\Omega \;, \tag{16}$$

which tends to be solved with much lower error even if spatial aliasing and order truncation are apparent. $|\cdot|$ denotes the absolute value. A cutoff frequency of $f_\mathrm{c} = 2\,\mathrm{kHz}$ has been shown to be a useful choice [34], [35].

A detailed presentation of the equalization method is available in [31].

ACKNOWLEDGMENTS

The authors thank Reality Labs for funding the presented work and Zamir Ben-Hur for valuable comments on the manuscript.

REFERENCES

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

[2] J. Ahrens and Z. Hu, "Evaluation of non-spherical scattering bodies for ambisonic microphone arrays," in *Proc. of the AES Int. Conf. AVAR*, Redmond, WA, USA, Aug. 2022.

[3] J. E. Greenberg and P. M. Zurek, "Microphone-array hearing aids," in *Microphone Arrays*, M. Brandstein and D. Ward, Eds. Berlin, Heidelberg: Springer, 2001, ch. 11, pp. 229–253.

[4] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*, K. R. Liu and S. Haykin, Eds. Hoboken, NJ: John Wiley & Sons, Inc., 2010, ch. 9, pp. 269–302.

[5] A. W. Bronkhorst and J. A. Verhave, "A microphone-array-based system for restoring sound localization with occluded ears," in *Meeting Proceedings RTO-MP-HFM-123*, Neuilly-sur-Seine, France, 2017, p. paper 20.

[6] P. Calamia, S. Davis, C. Smalt, and C. Weston, "A conformal, helmet-mounted microphone array for auditory situational awareness and hearing protection," in *IEEE WASPAA*, New Paltz, NY, USA, 2017, pp. 96–100.

[7] L. Madmoni, J. Donley, V. Tourbabin, and B. Rafaely, "Beamforming-based Binaural Reproduction by Matching of Binaural Signals," in *Proc. of the AES Int. Conf. AVAR*, Redmond, WA, USA, Aug. 2020.

[8] J. Fernandez, L. McCormack, P. Hyvärinen, A. Politis, and V. Pulkki, "Enhancing binaural rendering of head-worn microphone arrays through the use of adaptive spatial covariance matching," *The Journal of the Acoustical Society of America*, vol. 151, no. 4, pp. 2624–2635, 2022.

[9] V. Tourbabin and B. Rafaely, "Direction of arrival estimation using microphone array processing for moving humanoid robots," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 2046–2058, 2015.

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

[11] L. McCormack, A. Politis, R. Gonzalez, T. Lokki, and V. Pulkki, "Parametric ambisonic encoding of arbitrary microphone arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, pp. 1–14, 2022.

[12] J. Ahrens, H. Helmholz, D. L. Alon, and S. V. Amengual Garí, "A head-mounted microphone array for binaural rendering," in *Int. 3D Audio Conference (I3DA)*, Bologna, Italy, 2021.

[13] ——, "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.

[14] J. Ahrens, H. Helmholz, D. Alon, and S. V. Amengual Garí, "Spherical Harmonic Decomposition of a Sound Field Based on Observations Along the Equator of a Rigid Spherical Scatterer," *J. Acoust. Soc. Am.*, no. 150, 2021.

[15] N. Gumerov and R. Duraiswami, *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*. Amsterdam: Elsevier, 2005.

[16] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. on Speech and Audio Proc.*, vol. 13, no. 1, pp. 135–143, 2005.

[17] B. Bernschütz, "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100," in *Proceedings of AIA/DAGA*. Meran, Italy: DEGA, Mar. 2013, pp. 592–595.

[18] B. Rafaely and A. Avni, "Interaural cross correlation in a sound field represented by spherical harmonics," *J. Acoust. Soc. Am.*, vol. 127, no. 2, pp. 823–828, 2 2010.

[19] J. Ahrens, "Binaural Audio Rendering in the Spherical Harmonic Domain: A Summary of the Mathematics and Its Pitfalls," Technical note v. 2, Chalmers University of Technology, 2022.

[20] J. Ahrens, H. Helmholz, D. Alon, and S. V. Amengual Garí, "The far-field equatorial array for binaural rendering," in *IEEE ICASSP*, Toronto, Canada, 2021.

[21] S. Moreau, J. Daniel, and S. Bertet, "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, May 2006.

[22] A. Politis and H. Gamper, "Comparing modeled and measurement-based spherical harmonic encoding filters for spherical microphone arrays," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2017, pp. 224–228.

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

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[25] H. Wierstorf, M. Geier, and S. Spors, "A free database of head related impulse response measurements in the horizontal plane with multiple distances," in *130th Convention of the AES, e-Brief 6*. London, UK: AES, May 2011.

[26] S. Spagnol, "On distance dependence of pinna spectral patterns in head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 137, no. 1, pp. EL58–EL64, 2015.

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

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

[29] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Trans. on Speech and Audio Processing*, vol. 9, no. 6, pp. 697–707, 2001.

[30] T. Lübeck, H. Helmholz, J. M. Arend, C. Pörschmann, and J. Ahrens, "Perceptual Evaluation of Mitigation Approaches of Impairments due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data," *JAES*, vol. 68, no. 6, pp. 428–440, 2020.

[31] J. Ahrens, T. Deppisch, and etal, "Generalized angle-dependent equalization of spherical harmonic representations of sound fields for spatial audio rendering applications," *unpublished manuscript*, 2022.

[32] D. Ackermann, F. Fiedler, F. Brinkmann, M. Schneider, and S. Weinzierl, "On the acoustic qualities of dynamic pseudo-binaural recordings," *JAES*, vol. 68, no. 6, pp. 418–427, July 2020.

[33] Mono, "Charlize Theron_Head," Online, https://skfb.ly/6RBTQ, 2022, last viewed: 2022-02-07.

[34] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3616–3627, June 2018.

[35] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares," in *Fortschritte der Akustik – DAGA*, 2018, pp. 339–342.