Spherical Harmonic Decomposition of a Sound Field Based on Microphones Around the Circumference of a Human Head

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

Ahrens, J., Helmholz, H., Alon, D. et al (2021). Spherical Harmonic Decomposition of a Sound Field Based on Microphones Around the Circumference
of a Human Head. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2021-October: 231-235. http://dx.doi.org/10.1109/WASPAA52581.2021.9632751

SPHERICAL HARMONIC DECOMPOSITION OF A SOUND FIELD BASED ON MICROPHONES AROUND THE CIRCUMFERENCE OF A HUMAN HEAD

Jens Ahrens, Hannes Helmholz**

Chalmers University of Technology
412 96 Gothenburg, Sweden
firstname.lastname@chalmers.se

David Lou Alon, Sebastià V. Amengual Garí

Facebook Reality Labs Research, Facebook
1 Hacker Way, Menlo Park, CA 94025
{davidalon,samengual}@fb.com

ABSTRACT

We present a method for decomposing a sound field into spherical harmonics (SH) based on observations of the sound field around the circumference of a human head. The method is based on the analytical solution for observations of the sound field along the equator of a rigid sphere that we presented recently. The present method incorporates a calibration stage in which the microphone signals for sound sources at a suitable set of calibration positions are projected onto the SH decomposition of the same sound field 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 binaural rendering of numerically simulated signals for an array of 18 microphones providing 8th SH order to demonstrate its effectiveness.

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

1. INTRODUCTION

Content for virtual reality and augmented reality 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 [1]. Augmented reality headsets may be equipped with outward facing cameras for enabling head and body tracking. These cameras may also be employed to record a panoramic video of the wearer's environment. We explore in this paper a solution for a head-worn microphone array that enables panoramic audio recording of the wearer's environment by means of a SH decomposition of the sound field captured by the array.

The literature on head-mounted microphone arrays is vast, and we can only provide a brief outline here. 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 [2, 3]. Prediction of the binaural signals from microphones distributed over the head of a person was investigated, for example, in [2, 4, 5] whereby the wearer's orientation was encoded in the binaural signals.

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 playback hardware [6] and also to perform head tracking if the playback is performed binaurally.

Our approach is an extension of the recently proposed equatorial microphone array (EMA) [7] that comprises microphones along the equator of a rigid spherical scatterer. We demonstrate how a non-spherical scatterer can be employed in an EMA. To achieve this, we revisit and refine the approach for SH decomposition of a sound field based on microphones densely distributed over the surface of a non-spherical scatterer proposed in [8, 9].

2. SPHERICAL MICROPHONE ARRAYS

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 special cases of the general SMA concept.

SMAs typically employ pressure sensors distributed over an acoustically rigid spherical scatterer. A sound pressure field $S^{\text{surf}}(\beta, \alpha, R, \omega)$ on the surface of such a scatterer of radius R that is centered at the coordinate origin is given by [10, Eq. (3.1.1)]

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

with

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

and [10, Eq. (4.2.13)]

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

$\mathring{S}^{\text{surf}}_{n,m}(R, \omega)$ are the SH coefficients of the sound pressure on the surface of the spherical scatterer with radius R. $\breve{S}_{n,m}(\omega)$ are the SH coefficients – and thereby a complete representation – of the incident sound field. $Y_{n,m}(\beta, \alpha)$ are the SH basis functions, which are dependent on colatitude β and azimuth α of a spherical coordinate system. $\omega = 2\pi f$ 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.

The computation of $\breve{S}_{n,m}(\omega)$ from the microphone signals $S^{\text{surf}}(\beta, \alpha, R, \omega)$ can be performed via [11]

$$\mathring{S}^{\text{surf}}_{n,m}(R, \omega) = \oint_O S^{\text{surf}}(\beta, \alpha, R, \omega)\; Y_{n,m}(\beta, \alpha)^*\, \mathrm{d}\Omega \qquad (4)$$

*We thank Facebook Reality Labs for funding the presented work.

and

$$\breve{S}_{n,m}(\omega) = \mathring{S}^{\text{surf}}_{n,m}(R,\omega)\ b_n^{-1}(R,\omega)\ , \tag{5}$$

or equivalently,

$$\breve{S}_{n,m}(\omega) = b_n^{-1}(R,\omega) \oint_O S^{\text{surf}}(\beta,\alpha,R,\omega)\ Y_{n,m}(\beta,\alpha)^*\,\mathrm{d}\Omega\ . \tag{6}$$

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 well documented in the SMA literature [12].

In practical implementations, the integrals in (4) and (6) 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 $\mathring{H}_{n,m}(\omega)$ of the user's head-related transfer functions (HRTFs) are known, then binaural rendering of the (order limited) captured sound field can performed using [13]

$$D(\omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} (-1)^m\ \breve{S}_{n,-m}(\omega)\ \mathring{H}_{n,m}(\omega)\ , \tag{7}$$

i.e., the signal $D(\omega)$ that arises at a given ear of the listener if she/he is exposed to the captured sound field can be computed.

3. EQUATORIAL MICROPHONE ARRAYS

The EMA was proposed in [7], which is a generalization of the solution from [14]. An EMA is essentially an SMA as described in Sec. 2 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.

We omit details of the EMA solution as they are not of primary interest for the present work. What is relevant here are the high-level conclusions that can be drawn from the EMA solution [7]:

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 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 produce undesired deviations of the binaural output signals. Below the spatial aliasing frequency, these deviations are in the order of a few dB at some frequencies or smaller.

4. ARRAYS WITH ARBITRARILY-SHAPED SCATTERERS

The SMA and EMA solutions are only applicable to arrays that comprise a rigid 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 *sphere-like 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.

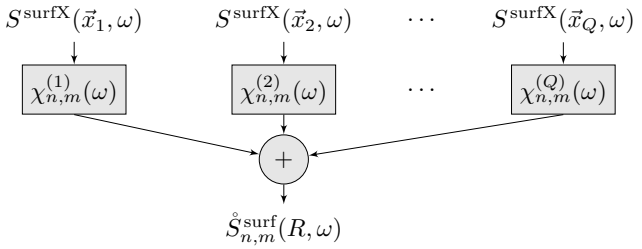

Figure 1: Block diagram representation of (8)

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], which performs a numerical fit of filter coefficients onto a set of known microphone signals and the corresponding known SH coefficients of the incident field.

Unlike [8, 9], 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 $\mathring{S}^{\text{surf}}_{n,m}(R,\omega)$ given by (2) as it is numerically well conditioned under all circumstances. In other words, we propose to project the pressure signal $S^{\text{surfX}}(\vec{x},\omega)$ 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 (6) 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.

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

$$\mathring{S}^{\text{surf}}_{n,m}(R,\omega) = \sum_{q=1}^{Q} \chi^{(q)}_{n,m}(\omega)\ S^{\text{surfX}}(\vec{x}_q,\omega)\ , \tag{8}$$

whereby $S^{\text{surfX}}(\vec{x}_q,\omega)$ 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 (8) is depicted in Fig. 1.

With conventional SMAs, the coefficients $\mathring{S}^{\text{surf}}_{n,m}(R,\omega)$ are computed in practice via quadrature of the integral in (4) as

$$\mathring{S}^{\text{surf}}_{n,m}(R,\omega) = \sum_{q=1}^{Q} w_q\ S^{\text{surf}}(\beta_q,\alpha_q,\omega)\ Y_{n,m}(\beta_q,\alpha_q)^*\ , \tag{9}$$

whereby w_q are the quadrature weights of the Q microphone locations. Comparing (9) and (8) makes it obvious that, in the case of the conventional SMA,

$$\chi^{(q)}_{n,m}(\omega) = \chi^{(q)}_{n,m} = w_q\ Y_{n,m}(\beta_q,\alpha_q)^*\ . \tag{10}$$

For the XMA, we have to assume that the complex weights $\chi^{(q)}_{n,m}(\omega)$ are frequency dependent. We can obtain them for sets

of (n, m, q) from a least-squares fit according to (8). This requires a set of microphone signals $S^{\text{surfX}}(\vec{x}, \omega)$ and the corresponding known coefficients $\mathring{S}^{\text{surf}}_{n,m}(R, \omega)$ 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 $\chi^{(q)}_{n,m}(\omega)$ is known, we can straightforwardly apply (8) to the microphone signals due to arbitrary incident sound fields to obtain their according SH coefficients $\mathring{S}^{\text{surf}}_{n,m}(R, \omega)$. We found that Tikhonov regularization [15, 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 as sound fields for this calibration due to the convenient implementation. 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 m [16, 17]. 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 [10, Eq. (2.3.6)]

$$\mathring{S}^{\text{surf,pw}}_{n,m}(\omega) = 4\pi \mathrm{i}^{-n}\, Y_{n,m}(\phi, \theta)^*\, b_n(R, \omega) \tag{11}$$

holds. When using (11) for calibration in (8), it is important to be aware of the implicit time reference that (11) comprises. Eq. (11) represents a spatio-temporal transfer function, which is the frequency-domain representation of the spatio-temporal impulse response $\mathring{s}^{\text{surf,pw}}_{n,m}(t)$. Eq. (11) implies that $t = 0$ is the moment when the notional planar wave front carrying a time-domain impulse, which $\mathring{s}^{\text{surf,pw}}_{n,m}(t)$ is the response to, passes the coordinate origin if no scattering object were present. The right hand side of (8) should be using the same time reference, too.

5. RESULTS

In order to test the capabilities of our proposed method to project the observed sound pressure onto a notional surface at a different radial distance from the origin, we perform the following simulations: We observe the sound pressure on the surface of a rigid spherical scatter of radius $R_{\text{XMA}} = 80\,\text{mm}$ (cf. (1)) and then use (8) to predict the sound pressure that the exact same sound field would evoke on the surfaces of concentric spherical scatterers of different radii R. We

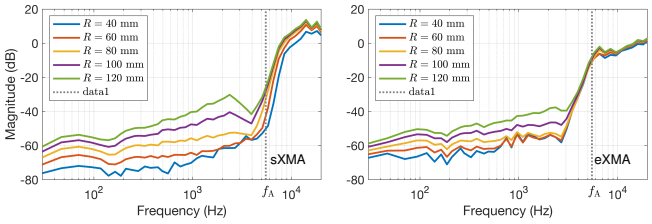

Figure 2: Average normalized error of 8th-order XMAs of spherical shape and radius $R_{\text{XMA}} = 80\,\text{mm}$ while (8) is used to project a horizontally propagating incident plane wave onto rigid spheres of different radii R. The error was evaluated at 1891 positions distributed over the target sphere. The reference sound pressure distribution is computed from (1) with 35th order. Left: 110-node sXMA. Right: 17-node eXMA. The spatial aliasing frequency f_{A} is also indicated.

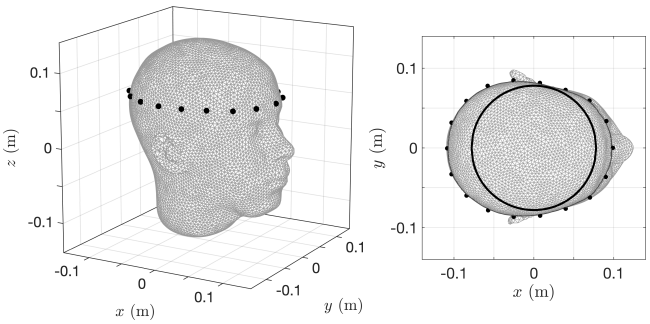

Figure 3: Mesh of the acoustically rigid head based on which the calibration and test data were computed. The ears are located on the y-axis. The black dots denote the locations of the 18 microphones, which are located at $z = 60\,\text{mm}$. Left: Perspective view. Right: Top-down view. The black solid line denotes the virtual rigid sphere of radius $R = 78\,\text{mm}$ onto which the microphone signals are projected.

compute $\chi^{(q)}_{n,m}(\omega)$ from calibration measurements of plane waves. The incident sound field for testing is a horizontally propagating plane wave, which we chose because both SMAs and EMAs are theoretically able to reconstruct the sound field perfectly in this case. We normalized the spheres' spatio-temporal impulse responses to 1 and added Gaussian noise with a root-mean-square (RMS) amplitude of $-80\,\text{dB}$ to the calibration signals and did not apply regularization.

We target an 8th order reconstruction and employed an sXMA with 110 microphones on a Lebedev grid, which was calibrated with 146 directions spherically distributed also on a Lebedev grid, and an eXMA with 17 microphones, which was calibrated with 35 horizontally propagating plane waves with constant azimuthal spacing.

Fig. 2 depicts the average normalized error as defined in [7, Eq. (21)] for above described scenarios. It is evident that the projection is accurate up to approx. 5 kHz for both sXMA and eXMA. The projection is inaccurate at higher frequencies because of spatial aliasing and SH order truncation, which is a well-studied phenomenon. This cutoff frequency can be estimated via the relation $N = \omega/c\,R$ [18], which yields $f_{\text{A}} \approx 5.5\,\text{kHz}$ for the present case.

Interestingly, the projection onto spheres that are smaller than the one on which the sound pressure is observed is more accurate than the projection onto a sphere of the same or larger radius. Gauss's theorem may partly explain this observation.

Finally, we evaluate 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. 3. We used the *mesh2hrtf* implementation of the boundary element method (BEM) from [19, 20] 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. We use the HRTFs of a *Neumann KH100* dummy head from [12] for the binaural rendering of the signals from the eXMA via (7).

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 orthog-

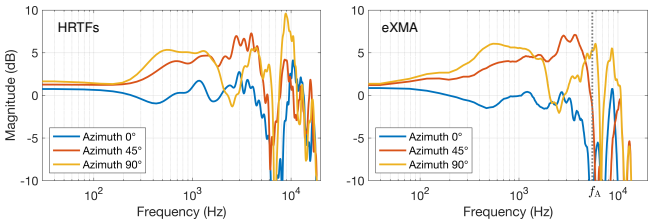

Figure 4: Left: Magnitude of the left-ear HRTFs of the dummy head for different azimuth angles of horizontal sound incidence. Right: Magnitude of the left-ear TFs of the eXMA for spherical waves originating from point sources located in the horizontal plane at the corresponding azimuth angles at a distance of 1 m.

onality of the implicit inherent circular harmonic decomposition, we chose to add one extra microphone to the minimum required number of 17.

- We calibrated the eXMA via (8) using spherical waves that originated from 90 equal-angularly spaced locations in the horizontal plane at a distance of 3 m. This is a distance that is sufficient to assume that the impinging wave fronts are planar at the XMA so that (11) 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 normalized the impulse responses to 1 and added Gaussian noise with an RMS amplitude of -80 dB to both calibration and test data to emulate measurement errors and sensor self-noise.
- We set the radius of the notional rigid sphere onto which (8) projects the signals to $R=78$ mm (cf. Fig. 3 (right)). This is the largest radius of a sphere that is centered at the coordinate origin that assures that the sphere fits entirely into the head.
- We evaluated the eXMA solution based on binaurally rendered signals for source positions other than those that were employed in the calibration. The binaural signals are computed using (8), (5), and (7).

Fig. 4 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 binaural transfer functions (TFs) of the eXMA for point sources located at the corresponding azimuth angles at a distance of 1 m. A perfect eXMA would produce binaural

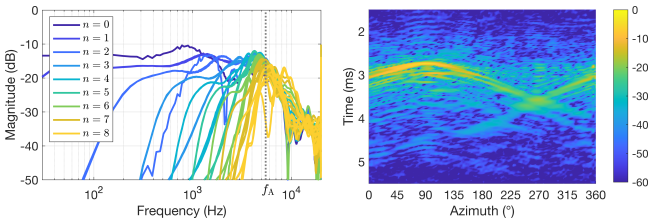

Figure 5: Left: $20\log_{10}\left|\chi_{n,m}^{(q)}(\omega)\right|$ for microphone q located at $(x,y)=(0.1,0)$ m. Right: eXMA left-ear impulse responses on a logarithmic scale for point sources in the horizontal plane at a distance of 1 m at different azimuth angles.

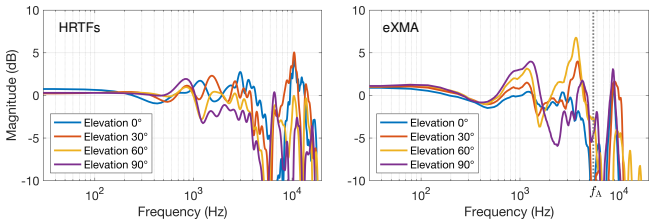

Figure 6: Same as Fig. 4 for sound incidence from straight ahead from different source elevations. Left: Dummy head HRTFs. Right: eXMA TFs.

TFs 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 m whereas the sources that we used for testing of the eXMA were located at a distance of 1 m. This means that even an ideal eXMA would theoretically produce binaural TFs that deviate from the reference HRTFs. It is evident from the data from [16, 17] that these deviations are negligible for the source distances that we employ.

It can be deduced from Fig. 4 that the binaural TFs of the eXMA are similar to the corresponding HRTFs up to a frequency of approx. 6 kHz whereby deviations in the order of 1 dB to 2 dB arise. The binaural TFs tend be slightly smoother than the HRFTs.

Significant attenuation of the eXMA TFs compared to the HRTFs occurs above 6 kHz. This is due to the truncation of the SH series in (7), and is well-known from the literature on spherical microphone arrays. Cf. also Fig. 2. A range of methods have been proposed to equalize this most which have been shown to be effective [21].

Fig. 5 (left) depicts $20\log_{10}\left|\chi_{n,m}^{(q)}(\omega)\right|$ for the present scenario. Remarkably, it is clearly identifiable what frequency range a given SH order n contributes to primarily.

For completeness, Fig. 5 (right) plots the binaural impulse responses of the eXMA for various horizontal angles of incidence to demonstrate that the binaural time cues are preserved.

The binaural TFs 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 [7]. It is worth noting that the deviations of the magnitude spectrum change continuously with the elevation of sound incidence, which tends to be less disturbing from a perceptual point of view than erratic changes.

We also tested our eXMA solution for sound sources at very close and very far distances and found no fundamental limitations [22]. The interaural time differences and partly also the interaural level differences are preserved even for non-horizontal sound incidence [22].

6. 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. Binaural audio examples of the scenarios covered in this paper are available at [23].

7. REFERENCES

[1] T. Magariyachi and Y. Mitsufuji, "Analytic error control methods for efficient rotation in dynamic binaural rendering of ambisonics," *The Journal of the Acoust. Soc. of America*, vol. 147, no. 1, pp. 218–230, 2020.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

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

[8] 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.

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

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

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

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

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

[23] "Audio examples, WASPAA 2021," http://www.ta.chalmers.se/research/audio-technology-group/audio-examples/waspaa-2021a/, accessed: 2021-04-06.