



A Head-Mounted Microphone Array for Binaural Rendering

Downloaded from: <https://research.chalmers.se>, 2022-07-02 09:42 UTC

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

Ahrens, J., Helmholz, H., Alon, D. et al (2021). A Head-Mounted Microphone Array for Binaural Rendering. Immersive and 3D Audio: from Architecture to Automotive (I3DA).
<http://dx.doi.org/10.1109/I3DA48870.2021.9610892>

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

A Head-Mounted Microphone Array for Binaural Rendering

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 recently presented a method for obtaining a spherical harmonic representation of a sound field based on microphones along the equator of a rigid spherical object that ideally has a size similar to that of a human head. We refer to this setup as an equatorial microphone array. Even more recently, we presented an extension of this method that allows for employing a scattering object that is approximately spherical such as a human head. The present paper provides an overview as well as a juxtaposition of the two solutions. We present an instrumental evaluation based on the application of binaural rendering of the captured sound fields by analysing simulated binaural transfer functions of both methods for a variety of scenarios.

Index Terms—Binaural audio, spherical microphone array, equatorial microphone array, virtual reality

I. INTRODUCTION

Content for virtual reality (VR) and augmented reality (AR) applications are typically captured with dedicated camera and microphone arrays. The videos from the camera array are stitched together to produce a panoramic video in which the user can look “around” when consuming the content on VR or AR headsets. Several of the commercially available headsets are equipped with outward-facing cameras that are primarily used for tracking the user’s head movements at this stage, but it is conceivable that these can also capture a panoramic video of the user’s environment, particularly with AR headsets.

The audio equivalent to panoramic video is head-tracked binaural audio, the content for which can be captured with rigid spherical microphone arrays (SMA) like the one depicted in Fig. 1 (left). What has not been available is a solution for a microphone array that can be mounted onto AR headsets additionally to a camera array so that the headset can serve as a self-sufficient multimedia capture device.

We recently proposed the equatorial microphone array (EMA) [1], which is essentially an SMA but with microphones only around the equator of the spherical scattering object that the array comprises as depicted in Fig. 1 (right). This comes at the price of not being able to capture elevation information, but it allows for reducing the number of microphones for a given spherical harmonic (SH) order N from $(N+1)^2$ for SMAs to $2N+1$. As an example, SMAs require at least 81 microphones for $N = 8$, whereby EMAs require not more than 17.

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

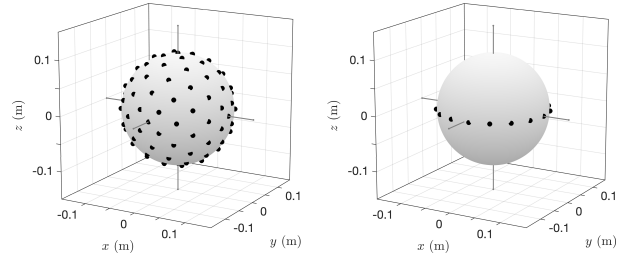


Fig. 1. 8th-order SMA (left) with 110 microphones on a Lebedev grid. 8th-order EMA (right) with 17 equiangularly spaced microphones.

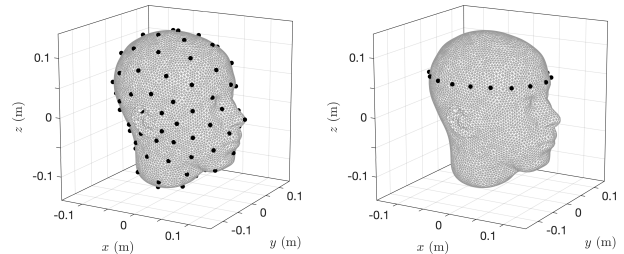


Fig. 2. 8th-order sXMA (left) with 110 microphones on a grid that is inspired by the Lebedev grid. 8th-order eXMA (right) with 18 equiangularly spaced microphones. The scatterer is a torso-less human head.

The second main advancement that the EMA constitutes is the fact that it provides an SH decomposition of the captured sound field from microphones that are on a circumferential contour on the scattering object. Adapting the conventional SMA solution to non-spherical scatterers such as it was done in [2], [3] is not sufficient for creating a viable head-mounted array as the microphones would have to be distributed over the surface of the entire head, as illustrated in Fig. 2 (left), which is obviously impractical. Extending the EMA solution to non-spherical scatterers, however, would allow for a head-mounted array like the one depicted in Fig. 2, which can be integrated into an AR headset.

We presented a solution for this type of head-mounted array in [4], which is an extension of the solutions from [2], [3] merged with the concept of EMAs. The present paper provides an overview of the evolution of the head-mounted array from the SMA via the EMA. We refer the reader to [1] for an in-depth treatment of the EMA and to [4] for an in-depth treatment of the head-mounted array.

II. SPHERICAL MICROPHONE ARRAYS

SMA typically employ pressure sensors distributed over an acoustically rigid spherical scatterer as depicted in Fig. 1 (left). 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 [5, Eq. (3.1.1)]

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

with

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

and [5, Eq. (4.2.13)]

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

$\check{S}_{n,m}^{\text{surf}}(R, \omega)$ are the SH coefficients of the sound pressure on the surface of the spherical scatterer. $\check{S}_{n,m}(\omega)$ 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 rad/s, f the frequency in Hz, and c is the speed of sound in m/s. $h_n^{(2)}(\cdot)$ denotes the derivative of the n th order spherical Hankel function of second kind.

$\check{S}_{n,m}(\omega)$ can be computed by integrating the sound pressure $S^{\text{surf}}(\beta, \alpha, R, \omega)$ on the surface of the scatterer, which is provided by pressure sensors in practice, over the entire surface O as [6]

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

and using

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

or equivalently,

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

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 [7].

In practical implementations, the integrals in (4) and (6) are approximated by summations over the microphone signals, which bounds the maximum SH order n that can be extracted to $n \leq N$. One speaks of an N th-order decomposition.

If the SH coefficients $\check{H}_{n,m}^{\text{L,R}}(\omega)$ of the user's left and right head-related transfer functions (HRTFs) are known, then the (order-limited) captured sound field can be rendered binaurally via [8]

$$B_{\text{L,R}}(\omega) = \sum_{n=0}^N \sum_{m=-n}^n (-1)^m \check{S}_{n,-m}(\omega) \check{H}_{n,m}^{\text{L,R}}(\omega), \quad (7)$$

i.e., the signal $B_{\text{L,R}}(\omega)$ that occurs at a given ear of the listener if the listener is exposed to the captured sound field can be computed. We refer the reader to the literature such as [1, Sec. II.B] for details on this.

III. EQUATORIAL MICROPHONE ARRAYS

EMAs use the same scattering body like SMAs with the difference that the microphones are placed exclusively around the equator of the scatterer. Cf. Fig. 1 (right). Such a setup cannot reconstruct arbitrary impinging sound fields because of ambiguities that occur. As a simple example, the array cannot differentiate two sound fields that are copies of one another mirrored on the horizontal plane. We proposed the following solution in [1] to mitigate this:

We assume that the impinging sound field $S(\vec{x}, \omega)$ is height-invariant, which allows for representing it by a continuum of horizontally propagating plane waves as [9, Sec. (4.32.3.4.4)]

$$S(\vec{x}, \omega) = \frac{1}{2\pi} \int_0^{2\pi} \bar{S}(\theta, \omega) e^{-i\frac{\omega}{c} r \cos(\theta-\alpha)} d\theta, \quad (8)$$

whereby $\bar{S}(\theta, \omega)$ are the plane wave coefficients, and θ is the azimuthal propagation direction of a given plane wave $e^{-i\frac{\omega}{c} r \cos(\theta-\alpha)}$. Despite the fact that (8) represents a sound field in terms of propagating plane waves, it is a general representation of 2D sound fields.

It can then be shown that the SH coefficients $\check{S}_n^m(\omega)$ (cf. (6)) of an impinging height-invariant sound field can be obtained from the sound pressure $S^{\text{surf}}(\pi/2, \alpha, R, \omega)$ along the equator of the spherical scatterer via a Fourier series expansion (a.k.a. circular harmonic expansion) as [1]

$$\check{S}_n^m(\omega) = \frac{4\pi i^{-n} Y_n^m(\pi/2, 0)}{\sum_{n'=|m|}^{\infty} 4\pi i^{-n'} b_{n'}(R, \omega) [Y_{n'}^m(\pi/2, 0)]^2} \times \frac{1}{2\pi} \int_0^{2\pi} S^{\text{surf}}(\pi/2, \alpha, R, \omega) e^{-im\alpha} d\alpha, \quad (9)$$

whereby the scattering off the scatterer is described in 3D and is removed.

Comparing (9) with the SMA equivalent (6) allows for identifying the EMA radial filters, which are essentially represented by the fraction in the upper line of (9). We demonstrated in [1] that they exhibit the same qualitative properties like SMA radial filters. The EMA filters require similar regularization of the gain at low frequencies at high SH orders whereby the methods from the SMA literature can be applied.

The evaluation presented in Sec. V will illustrate what consequences a violation of the requirement for a height-invariant impinging sound field has on the binaural output signals.

Previous solutions for the EMA such as [10], [11] used a pure 2D formulation of the problem. This causes a boost of low frequencies by dozens of dB for sound sources that are close to the array [1].

IV. HEAD-MOUNTED MICROPHONE ARRAYS

The head-mounted array from Fig. 2 (right) may be interpreted as an EMA with a non-spherical scattering body. One way of describing the acoustic effect of a non-spherical scattering body mathematically is the so-called T-matrix [5, Eq. (4.2.11)], which is a mathematical tool that allows for representing the scattering off an arbitrary object of finite spatial extent in terms of SHs. This approach, however, can only describe the resulting sound field in the exterior domain i.e., outside of a sphere that encloses the scattering object completely. The setup under consideration has microphones located in the interior domain (for example, those on the side of the head) so that the T-matrix cannot be employed, and we seek for a different solution.

We note that the coefficients $\check{S}_{n,m}(\omega)$ of the impinging sound field generally diverge at low frequencies for sound fields from sources at finite distances so that the solution may be ill-conditioned. We, therefore, propose to instead extract the coefficients $\mathring{S}_{n,m}^{\text{surf}}(R, \omega)$ given by (4) for the SMA. For the EMA, (9) can be reformulated to also yield $\mathring{S}_{n,m}^{\text{surf}}(R, \omega)$. The coefficients $\check{S}_{n,m}(\omega)$ that are required for the binaural rendering can then be computed via (6) using the well-known gain-limited SMA radial filters. This is in contrast to previous comparable methods that extract $\check{S}_{n,m}(\omega)$ directly [2], [3].

In either case, the SMA in (4) and the EMA in the according variant of (9), the extraction of the desired coefficients $\mathring{S}_{n,m}^{\text{surf}}(R, \omega)$ from the microphone signals can be represented by a linear combination of the microphone signals $S^{\text{surf}}(\vec{x}, \omega)$ observed at different positions \vec{x}_q on the scatterer's surface. In practice, the integrals are replaced by summations because of the employment of a discrete set of Q microphones so that the computation of $\mathring{S}_{n,m}^{\text{surf}}(R, \omega)$ can be formulated as

$$\mathring{S}_{n,m}^{\text{surf}}(R, \omega) = \sum_{q=1}^Q \chi_{n,m}^{(q)}(\omega) S^{\text{surf}}(\vec{x}_q, \omega), \quad (10)$$

whereby $\chi_{n,m}^{(q)}(\omega)$ are the complex linear weights of the microphone signals. A block diagram representation of (10) is depicted in Fig. 3.

With conventional SMAs, the coefficients $\mathring{S}_{n,m}^{\text{surf}}(R, \omega)$ are

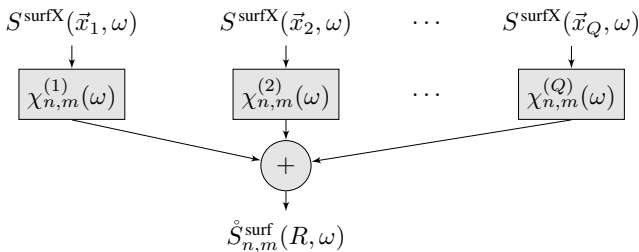


Fig. 3. Block diagram representation of (10) (from [4])

computed in practice via quadrature of the integral in (4) as

$$\mathring{S}_{n,m}^{\text{surf}}(R, \omega) = \sum_{q=1}^Q w_q S^{\text{surf}}(\beta_q, \alpha_q, \omega) Y_{n,m}(\beta_q, \alpha_q)^*, \quad (11)$$

whereby w_q are the quadrature weights of the microphone locations that are chosen such that the orthonormality of the SH bases is maintained. Comparing (11) and (10) makes it obvious that, in the case of the conventional SMA,

$$\chi_{n,m}^{(q)}(\omega) = \chi_{n,m}^{(q)} = w_q Y_{n,m}(\beta_q, \alpha_q)^*. \quad (12)$$

The equivalent of (11) for an EMA is a little more involved as it is given by a discrete variant of (9), which yields frequency-dependent linear weights $\chi_{n,m}^{(q)}(\omega)$ for the microphone signals.

The core of our solution for the head-mounted array is the following assumption: The extraction of the desired SH coefficients for SMA and EMA can be accomplished via (10). Given that the geometry of a human head departs only moderately from that of a sphere, it should be possible to find weights $\chi_{n,m}^{(q)}(\omega)$ that compute the coefficients $\mathring{S}_{n,m}^{\text{surf}}(R, \omega)$ from the sound pressure picked up by the microphones on the surface of the head. In other words, we propose to project the sound pressure $S^{\text{surf}}(\vec{x}, \omega)$ at position \vec{x} on the surface of the head 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 .

We use the term *sphere-like XMA* (sXMA) for arrays that employ microphones that are distributed over the entire surface of the scatterer like in Fig. 2 (left) and the term *equatorial XMA* (eXMA) for arrays whose microphones are located along an equator-like contour like in Fig. 2 (right).

As the linear weights $\chi_{n,m}^{(q)}(\omega)$ cannot be computed analytically for arbitrarily-shaped scatterers, we proposed in [4] to obtain them for sets of (n, m, q) from a least-squares fit according to (10). This requires a set of microphone signals $S^{\text{surf}}(\vec{x}, \omega)$ 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 for which the coefficients $\mathring{S}_{n,m}^{\text{surf}}(R, \omega)$ are known. Once $\chi_{n,m}^{(q)}(\omega)$ is determined, we can straightforwardly apply (10) to the microphone signals due to arbitrary incident sound fields to obtain their according SH coefficients $\mathring{S}_{n,m}^{\text{surf}}(R, \omega)$. Fig. 4 summarizes the processing pipeline for all discussed array variants.

The only differentiation between sXMA and eXMA that we have to make is that eXMAs should be calibrated based on sound fields that fulfill the requirement of height-invariance as much as possible. We found that point sources in the horizontal plane at a moderate distance (or farther) are suitable [4]. Given that an sXMA solution is not practical for a head-mounted array (cf. Fig. 2 (left)), we will therefore focus only on the eXMA solution in the remainder of this paper.

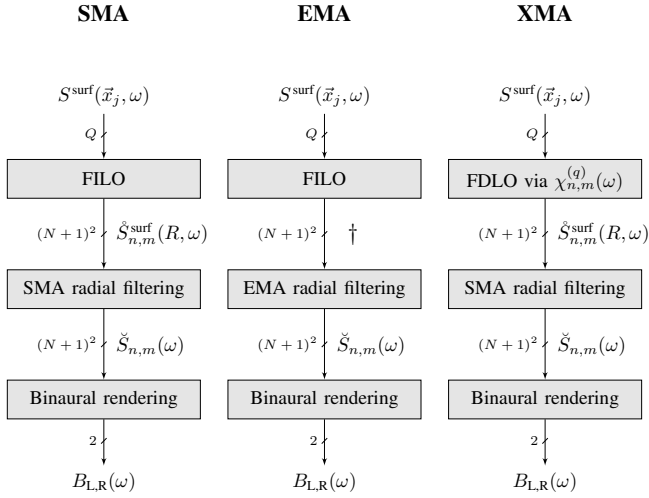


Fig. 4. Juxtaposition of the processing steps for SMA (left), EMA (middle), and XMA (right). The XMA pipeline applies to both sXMA and eXMA. FILO: Frequency-independent linear operation. FDLO: Frequency-dependent linear operation. $B_{L,R}$ are the binaural output signals of the left and right ear computed via (7). The signals denoted by the dagger \dagger do not appear explicitly in the mathematical formulation. We, therefore, do not assign a mathematical symbol to them. SMA radial filtering is given by (5). EMA radial filtering is implicitly contained in (9).

V. EVALUATION

This section provides an overview of the performance of both the EMA depicted in Fig. 1 (right) as well as the head-mounted eXMA depicted in Fig. 2 (right). All results are based on computer simulations. More scenarios are covered by the audio examples that accompany the present paper [12].

The EMA has a radius of $R = 0.0875$ m. The radius of the eXMA is similar to that. We used the *mesh2hrtf* implementation of the boundary element method (BEM) from [13], [14] 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 KU100* dummy head from [15] for the binaural rendering of the signal from both EMA and eXMA using [1, Eq. (9)].

The EMA solution is entirely analytical and does not require calibration. We calibrated the eXMA via (10) 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 eXMA. We refer the reader to [4] for more information on the calibration and evaluation procedures.

A. Binaural Transfer Functions

Fig. 5 depicts a comparison between the HRTFs that are employed for the rendering (top) and the corresponding binaural transfer functions (BTFs) of the EMA (middle) and the eXMA (bottom). Ideal EMAs and eXMAS would yield BTFs that are identical to the HRTFs. For the presented angles of sound incidence, HRTFs and the EMA and eXMAS

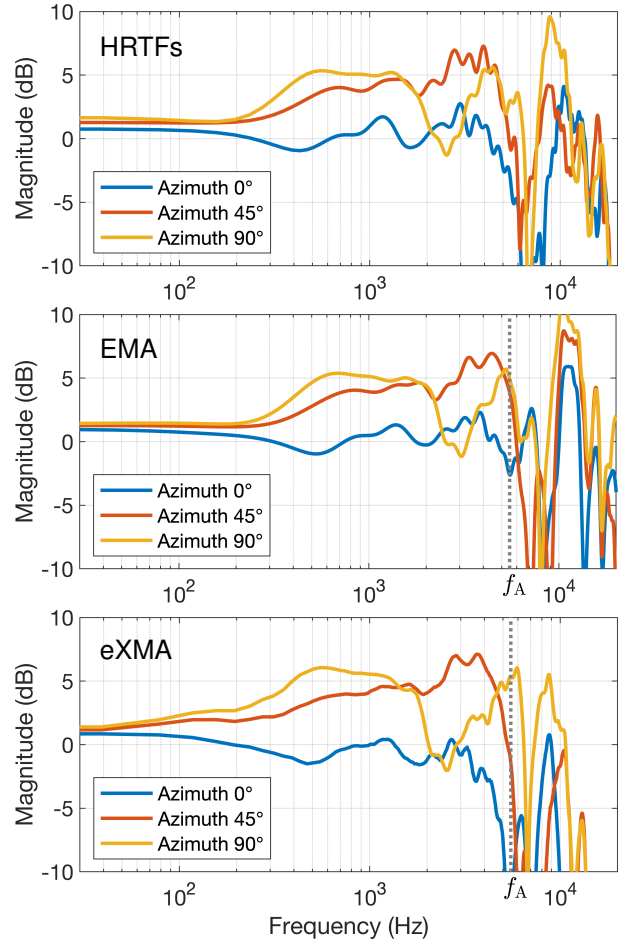


Fig. 5. Magnitude of the left-ear BTFs for different azimuth angles of horizontal sound incidence. Top: HRTFs of the dummy head. Middle: 8th-order EMA with 17 microphones. Bottom: 8th-order eXMA with 18 microphones.

responses deviate by less than 2 dB below the spatial aliasing frequency f_A of the arrays, which is approx. 5 kHz for the setups under consideration. The array responses deviate at higher frequencies whereby the eXMA shows a pronounced attenuation with increasing frequency.

Fig. 6 depicts the same data but for sound incidence from straight ahead from different elevations. As expected, the EMA and eXMA BTFs differ for non-horizontal sound incidence as this scenario violates the assumption of height invariance of the incident sound field. EMA and eXMA BTFs change in a very similar manner and gradually with the elevation. It is worth highlighting that the effect of the violation on the BTFs is of moderate magnitude. Remarkably, both EMA and eXMA preserve the interaural time difference (ITD) very well even for non-horizontal sound incidence, and they preserve the interaural level difference (ILD) reasonably well as depicted in Fig. 7. This is also evident from the binaural audio examples that accompany this paper [12]. In simple words, sound incidence from straight above produces no interaural differences, and sound incidence from a lateral direction and, say, 45° elevation produces a smaller interaural differences

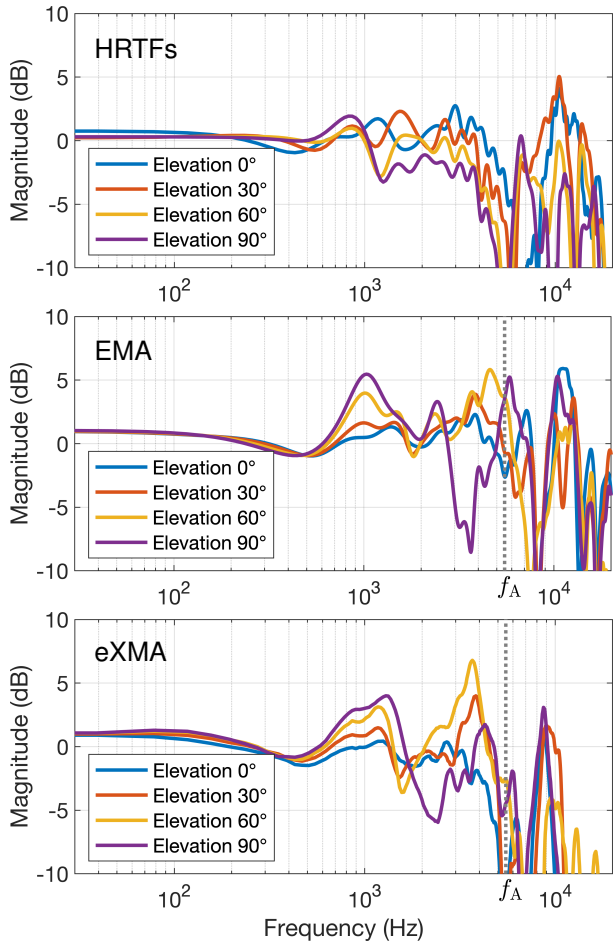


Fig. 6. Same as Fig. 5 but for different elevations of sound incidence in the median plane

compared to sound incidence from the same lateral angle and 0° elevation.

Fig. 8 depicts similar data like Fig. 5-6 but this time for point sources straight ahead of the listener at different distances. All data are normalized with respect to the source distance. We do not present the corresponding HRTF data because the employed HRTFs are only available for a source distance of 3 m.

Capturing sound fields from point sources with EMAs and eXMA requires some attention as the underlying assumption of height-invariance of the captured sound field is violated stronger the closer the point source is to the array. A small amplification at very low frequencies as well as an attenuation of mid frequencies is apparent in the EMA BTFs (cf. Fig. 8 (top)) 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 [16]. The BTFs of SMAs show the exact same dependency on the source distance [1]. Note that this phenomenon is not caused by changes in the way the incident sound field is scattered on the array's body. It rather reflects the range dependency of the HRTFs that are used in the rendering. The eXMA

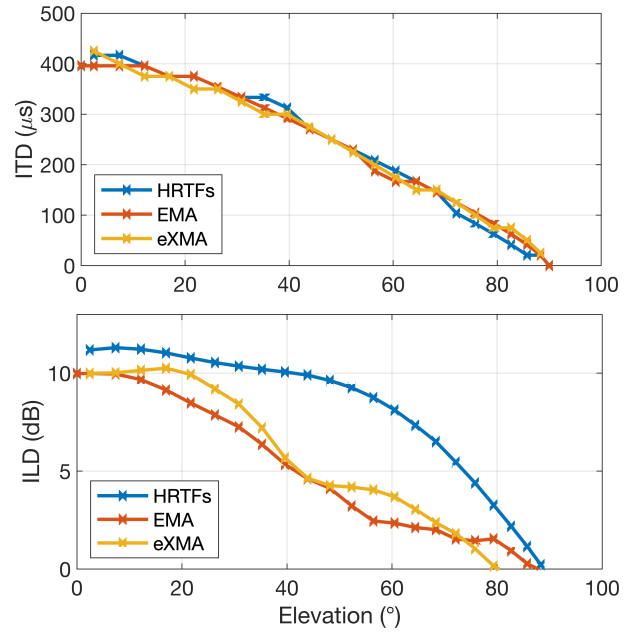


Fig. 7. Sound incidence from 45° azimuth and different elevations for the HRTFs and arrays from Fig. 5-6. Top: ITD defined here as the lag at which the maximum value of the interaural cross-correlation occurs. Bottom: broadband ILD.

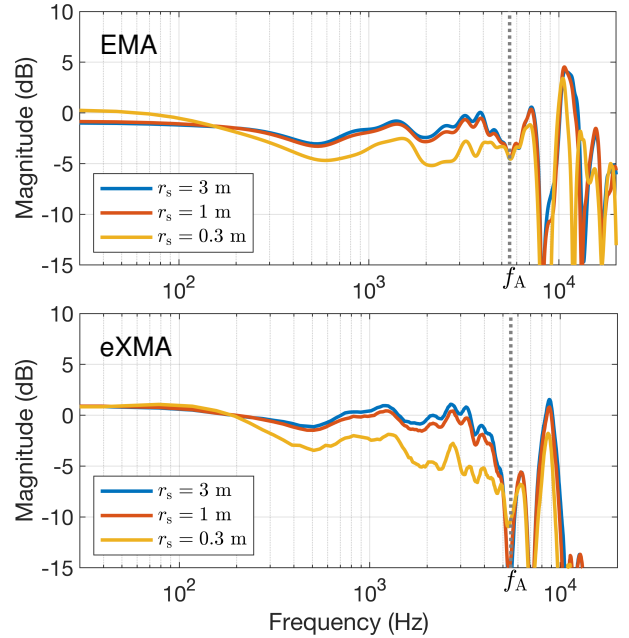


Fig. 8. Magnitude of the left-ear BTFs for point sources straight ahead at different distances r_s . Top: 8th-order EMA with 17 microphones. Bottom: 8th-order eXMA with 18 microphones.

exhibits a slightly larger attenuation of the mid frequencies by approx. 2 dB for the closest source distance of $r_s = 0.3$ m. This suggests that a small change of the timbre may occur.

B. Equalization

It is known from the SMA literature that SH order truncation leads to an attenuation of the binaural signals at high frequen-

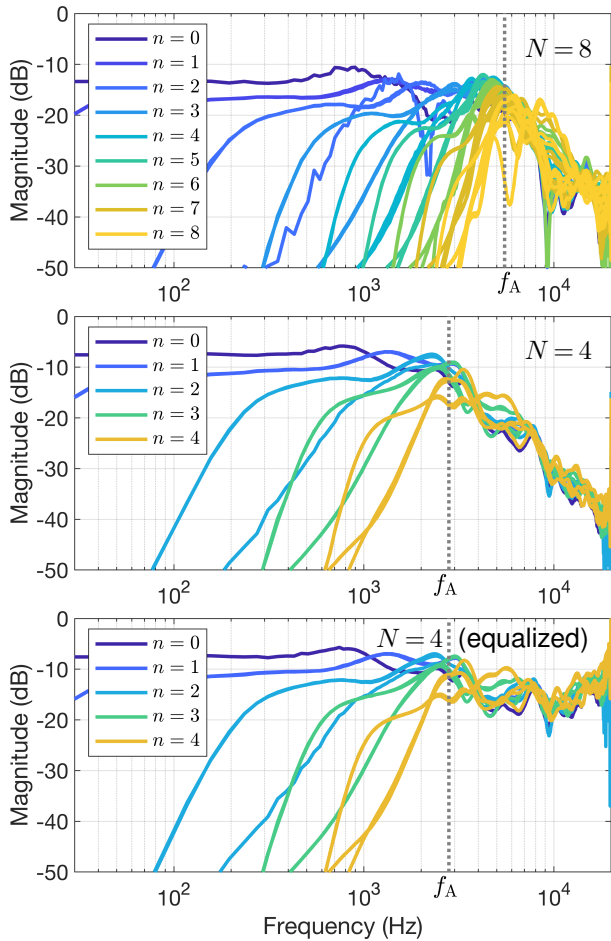


Fig. 9. $20 \log_{10} |\chi_{n,m}^{(q)}(\omega)|$ for the microphone located above the nose of the listener ($q=1$). Top: $N=8$ for an eXMA with 18 microphones. Middle: $N=4$ for an eXMA with 9 microphones. Bottom: same data like in the middle plot but equalized above $f_E = 2000$ Hz.

cies, and that spatial aliasing leads to an amplification at high frequencies [7]. A good number of equalization methods exist for binaural rendering of SMA signals most of which have been shown to be successful [17], [18]. It is evident from Fig. 5-8 that eXMAs require different equalization solutions as the attenuation of the high-frequency content is more pronounced than for SMAs and EMAs. The present section will shed some light on this.

We have no confirmed explanation at this point for the pronounced attenuation of the high-frequency content. The attenuation turns out to be a property of the linear weights $\chi_{n,m}^{(q)}(\omega)$ from (10). Recall that $\chi_{n,m}^{(q)}(\omega)$ represent the filters that compute the SH coefficients of the incident sound field from the microphone signals. Fig. 9 (top) depicts $|\chi_{n,m}^{(q)}(\omega)|$ on a logarithmic scale for 8th-order capture based on the array from Fig. 2 (right), and Fig. 9 (middle) depicts the same data for 4th-order capture using every other microphone of the same array. It is evident that all $\chi_{n,m}^{(q)}(\omega)$ roll off above the spatial aliasing frequency with a rate of approx. 10 dB/oct.

Fig. 9 (bottom) depicts the same data like Fig. 9 (middle) but

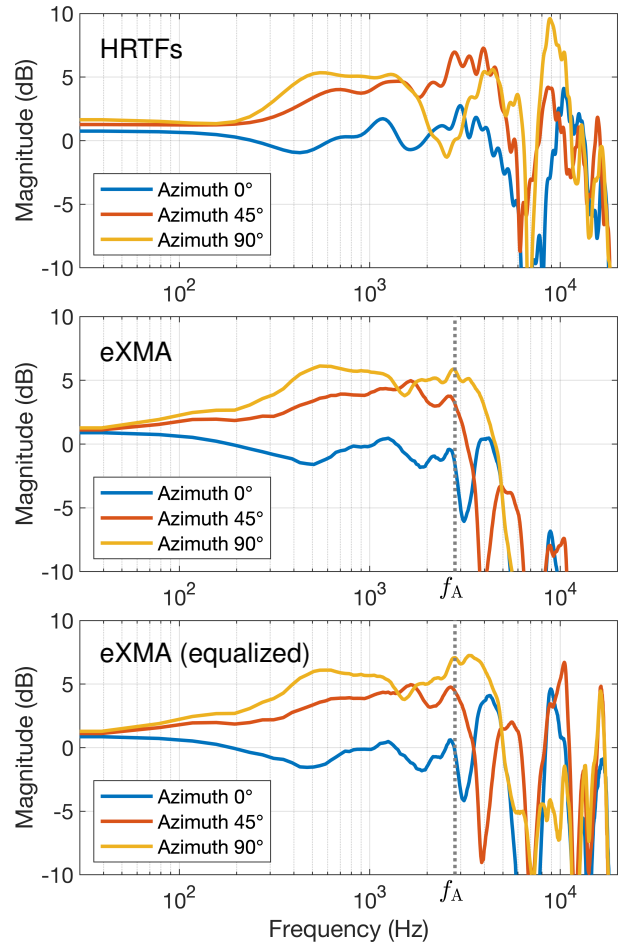


Fig. 10. Magnitude of the left-ear BTFs for different azimuth angles of horizontal sound incidence. Top: HRTFs (same as Fig. 5 (top)). Middle: 4th-order eXMA with 9 microphones. Bottom: 4th-order eXMA with 9 microphones (equalized).

with a manual compensation of the roll-off. The effect of this equalization on the binaural output signals becomes evident when comparing Fig. 10 (middle), which uses the unequalized $\chi_{n,m}^{(q)}(\omega)$ from Fig. 9 (middle), with Fig. 10 (bottom), which uses the equalized $\chi_{n,m}^{(q)}(\omega)$ from Fig. 9 (bottom). The spectral balance of the BTFs of the equalized eXMA is more even than without equalization. Significant deviations to the corresponding HRTFs in Fig. 10 (top) are still apparent at high frequencies. These are due to spatial aliasing and order truncation, and it is unclear at this stage in how far they can be overcome. End-to-end MagLS may be an option [19].

Finally, comparing the 4th-order eXMA in Fig. 10 (middle) with the 8th-order eXMA in Fig. 10 (bottom) suggests that higher orders can increase the accuracy of the BTFs also slightly below the spatial aliasing frequency. Note that the deviations of the 4th-order eXMA from the corresponding HRTFs are larger than those of the 8th-order eXMA at mid frequencies of, say, 1.5 kHz to 3.0 kHz.

VI. CONCLUSIONS AND FUTURE WORK

We presented a summary of recent advancements on microphone arrays on scattering objects with a sphere-like shape that allow for a spherical harmonic decomposition of the impinging sound field. We focused on an array that was placed along a circumferential contour on a human head and used the application example of binaural rendering of the captured sound fields for the evaluation. Binaural audio examples are available at [12].

Future work includes determining the requirements for the inevitable calibration of the array. It will be interesting to understand if it requires to calibrate the array for each user separately or if, for example, a generic calibration can be used that is adapted to the specific user.

The circumferential arrays exhibit the limitation that they cannot produce monaural elevation cues but always output a representation of a horizontally propagating sound field, yet with the correct interaural time difference for all angles of sound incidence. It is unclear at this point in how far this affects the perceptual result. It is also not fully clear what role the individualization of the HRTFs has that may be used in the rendering [20]. Remarkably, it has not been proven that conventional spherical microphone arrays are actually able to preserve such monaural elevation cues correctly even if individual or otherwise favorable HRTFs are used. 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 [21].

The absence of monaural elevation cues has the potential to also alter the spatial impression even if all sound sources are located in the horizontal plane because room reflections may impinge from elevated directions and may therefore be altered. A preliminary study to determine the perceptual importance of elevation information in acoustic room responses is available in [22]. It was found that projecting strong ceiling reflections onto the horizontal plane can cause an audible difference. Room responses with less strong ceiling reflections can sound identical with and without elevation information. These observations are in line with results from the perceptual evaluation of pseudobinaural rendering in [21] where the absence of monaural elevation cues did not lead to a reduction of the plausibility of room perception. Studies on the dependency of listener preference on elevation information are available, for example, in [23].

REFERENCES

- [1] J. Ahrens, H. Helmholtz, D. Alon, and S. 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).
- [2] 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.
- [3] D. N. Zotkin, N. A. Gumerov, and R. Duraiswami, "Incident field recovery for an arbitrary-shaped scatterer," in *IEEE ICASSP*, 2017, pp. 451–455.
- [4] J. Ahrens, H. Helmholtz, D. L. Alon, and S. V. Amengual Garí, "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.
- [5] N. Gumerov and R. Duraiswami, *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*. Amsterdam: Elsevier, 2005.
- [6] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. on Speech and Audio Proc.*, vol. 13, no. 1, pp. 135–143, 2005.
- [7] B. Bernschütz, "Microphone arrays and sound field decomposition for dynamic binaural recording," PhD thesis, Technische Universität Berlin, 2016.
- [8] 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.
- [9] R. Rabenstein, S. Spors, and J. Ahrens, "Sound field synthesis," in *Academic Press Library in Signal Processing: Volume 4*, Joel Trussell et al., Ed. Oxford, UK: Academic Press, 2014, ch. 32, pp. 915–979.
- [10] T. Weller, S. Favrot, and J. M. Buchholz, "Application of a circular 2D hard-sphere microphone array for higher-order ambisonics auralization," in *Proc. of Forum Acusticum*, Aalborg, Denmark, 2011, pp. 2269–2274.
- [11] J. Ahrens, H. Helmholtz, D. Alon, and S. Amengual Garí, "The far-field equatorial array for binaural rendering," in *IEEE ICASSP*, 2021.
- [12] "Audio examples, I3DA 2021," <http://www.ta.chalmers.se/research/audio-technology-group/audio-examples/i3da-2021/>, accessed: 2021-06-15.
- [13] 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.
- [14] 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.
- [15] 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.
- [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] T. Lübeck, H. Helmholtz, 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.
- [18] T. Lübeck, J. M. Arend, C. Pörschmann, H. Helmholtz, and J. Ahrens, "Perceptual Evaluation of Mitigation Approaches of Impairments Due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data: Dry Acoustic Environments," in *23rd International Conference on Digital Audio Effects (DAFx)*, 2020, pp. 250–257.
- [19] T. Deppisch, H. Helmholtz, and J. Ahrens, "End-to-end magnitude least squares binaural rendering of spherical microphone array signals," in *Int. 3 D Audio Conference (I3DA)*, 2021.
- [20] C. Armstrong, L. Thresh, D. Murphy, and G. Kearney, "A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database," *Applied Sciences*, vol. 8, no. 11, 2018.
- [21] 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, Jul. 2020.
- [22] L. Müller, "Perceptual differences caused by altering the elevation of early room reflections," MSc thesis, Chalmers University of Technology, 2021.
- [23] T. Robotham, M. Stephenson, and H. Lee, "The effect of a vertical reflection on the relationship between preference and perceived change in timbral and spatial attributes," in *140th Convention of the AES*, Paris, France, May 2016.