Ambisonics Capture using Microphones on Head-worn Device of Arbitrary Geometry

Amy Bastine, Lachlan Birnie, Thushara D. Abhayapala, Prasanga Samarasinghe
Audio & Acoustic Signal Processing Group, The Australian National University
Canberra, Australia
{amy.bastine, lachlan.birnie, thushara.abhayapala, prasanga.samarasinghe}@anu.edu.au

Vladimir Tourbabin
Meta (Facebook) Reality Labs
Redmond, Washington, USA
vtourbabin@fb.com

***Abstract*—3D auditory scenes for Extended Reality (XR) environments are usually recorded using spherical microphone arrays and then binaurally rendered through head-wearable devices. Considerable research is ongoing to integrate spatial audio recording capabilities into these devices. Plausible microphone array designs mountable on head-devices will have arbitrary geometry and unknown scattering effects. In this case, we can not apply conventional eigenbeam processing methods of regular arrays. In this paper, we propose an alternative method using Wearable-Device-Related Transfer Functions (WDRTFs) to extract higher-order ambisonics from the recordings of the head-device. WDRTFs are directional responses incorporating the sound wave transformations caused by the device's and wearer's body. The inherent diversity of these functions is exploited to estimate the ambisonics of the recorded soundfield. This paper presents a preliminary evaluation of this method using two irregular microphone array designs worn on a spherical head and the results are promising in comparison to an equivalent SMA. Therefore, we can further adapt this method to achieve perceptually accurate binaural rendering using the microphone recordings of the head-device.**

***Index Terms*—Arbitrary microphone array, head devices, spherical harmonics, extended reality, device response**

I. Introduction

Synthesized reality based on head-wearable devices is an emerging technology and major industrial research focuses on achieving high-fidelity audio-visual immersive experience using these devices. While there are significant developments in the visual component using 360° cameras and head-mounted displays [1], the spatial audio capabilities of such devices are still limited. In practice, 3D audio is captured using microphone arrays and then rendered binaurally through Virtual Reality (VR) headsets using Head-Related Transfer Functions (HRTFs) [2]. Spherical Microphone Arrays (SMAs) are extensively used for this purpose since they allow convenient rendering in the Spherical Harmonics (SH) domain [3]–[6]. Other geometries like hemispherical microphone arrays [7], parallel circular arrays [8], double-sided conical sensor arrays [9], planar hybrid microphone arrays [10], and equatorial microphone arrays [11] have also been developed for SH decomposition of the sound field. However, these spatial audio capturing techniques can not be applied on head-devices due to arbitrary geometry, limited number of microphones, and complex scattering effects.

This research was supported by Meta (Facebook) Reality Labs.

The problem of scattering in microphone arrays mounted on non-spherical devices is addressed in [12]–[15] by means of direction-dependent responses or Device Related Transfer Functions (DRTFs). This approach was adapted in [16] for head-mounted equatorial microphone arrays. Our paper aims to further develop this concept for higher-order ambisonics extraction using head-devices with arbitrary microphone placements. The proposed method will incorporate the scattering, reflection and diffraction effects caused by the device's and the wearer's body. This formulation allows us to take advantage of the additional diversity in the acoustic transfer functions to solve the sound field decomposition problem. Improved diversity is supposed to reduce the microphone requirements and achieve reliable estimation over a wider frequency range [12]. In this paper, we use this method to show the viability of head-devices for spatial audio capture.

II. Problem Formulation

In this section, we review the conventional sound field analysis method using SH decomposition and discuss its limitations in adapting to irregular microphone arrays positioned on non-spherical rigid bodies.

The sound pressure at an arbitrary point $\boldsymbol{x}$ in a source-free region can be expressed in the SH domain as [17]

$$P(k,\boldsymbol{x}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \alpha_{nm}(k) j_n(k|\boldsymbol{x}|) Y_{nm}(\hat{\boldsymbol{x}}) \qquad (1)$$

where $k = 2\pi f/c$ is the wavenumber, f is frequency, c is speed of sound, $\alpha_{nm}(k)$ are the SH coefficients of the sound field with $\{n,m\} \in \mathbb{Z}$, $j_n(\cdot)$ is the n^{th}-order spherical Bessel function, $Y_{nm}(\cdot)$ is the SH function of n^{th}-order and m^{th}-mode, $|\cdot| = r$ denotes the vector magnitude, and $\hat{(\cdot)} = (\theta,\phi)$ denotes the vector direction with elevation angle $\theta \in [0,\pi]$ downwards from the Z-axis and azimuth angle $\phi \in [0,2\pi)$ counterclockwise from the X-axis. For a region of interest with radius r, (1) can be safely truncated to an order $N = \lceil ker/2 \rceil$ [18], where $e \approx 2.71828$ is the Euler's number and $\lceil\cdot\rceil$ denotes ceiling operation.

When the $P(k,\boldsymbol{x})$ function is known over a spherical surface, we can calculate $\alpha_{nm}(k)$ coefficients as [17]

$$\alpha_{nm}(k) = \frac{1}{j_n(k|\boldsymbol{x}|)} \int_{\hat{\boldsymbol{x}}} P(k,\boldsymbol{x}) Y^*_{nm}(\hat{\boldsymbol{x}}) d\hat{\boldsymbol{x}} \qquad (2)$$

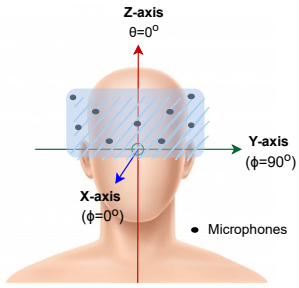


Fig. 1: Illustration of a microphone array of arbitrary geometry embedded on a head-worn device.

where $(\cdot)^*$ denotes complex conjugation operation and $j_n(k|x|) \neq 0$.

In typical practical realizations, (2) is approximated as a discrete sum using $P(k, \boldsymbol{x})$ samples recorded over a sphere by an SMA. Let $P_M(k, \boldsymbol{z}_v)$ be the incident sound pressure recorded by the v^{th} microphone element located at the spherical coordinate $\boldsymbol{z}_v = (R, \theta_v, \phi_v)$ for $v \in \{1, 2, \cdots, V\}$. The center of the microphone array is considered as the coordinate origin. Using $P_M(k, \boldsymbol{z}_v)$, (2) can be modified to [4]

$$\alpha_{nm}(k) = \frac{\sum_{v=1}^{V} P_M(k, \boldsymbol{z}_v) Y_{nm}^*(\hat{\boldsymbol{z}}_v)}{b_n(k, R, R_s)} \tag{3}$$

where

$$b_n(k, R, R_s) = \begin{cases} j_n(kR) & \text{, for an open array} \\ j_n(kR) - \frac{j_n'(kR_s)}{h_n^{(2)\prime}(kR_s)} h_n^{(2)}(kR) & \text{, for a rigid array} \end{cases} \tag{4}$$

is the mode strength [19], R_s is the radius of the rigid spherical scatterer, $h_n^{(2)}(\cdot)$ is the n^{th}-order spherical Hankel function of second kind, and $(\cdot)'$ indicates first-order derivative operation. For conventional rigid SMAs, $R_s = R$. Using an SMA of V microphones, we can solve (3) to a maximum order $\lfloor \sqrt{V} - 1 \rfloor$ using Least-Squares (LS) method [4], [19], where $\lfloor \cdot \rfloor$ denotes floor operation.

Note that the standard ambisonics derivation method using (3) is valid only for sufficiently uniform microphone placements that satisfy the SH orthogonality condition. Moreover, this method can be only applied when the microphones are mounted on a spherical rigid body with known scattering effects compensated through $b_n(k, R, R_s)$.

In this paper, we are interested in estimating the higher-order $\alpha_{nm}(k)$ coefficients using a microphone array positioned on an arbitrary-shaped device worn on a human head. In such a system, we need to consider the scattering and diffraction effects of the device as well as the wearer's head and torso. Considering these requirements, we propose an alternative ambisonic extraction method in the next section.

III. Microphone Array on Head-worn Device

Consider a microphone array of arbitrary geometry mounted on a head-worn device as illustrated in Fig. 1 with the center of the head considered as the coordinate origin. The device has Q microphones embedded at different locations $\boldsymbol{x}_q = (R_q, \theta_q, \phi_q)$ for $q \in \{1, 2, \cdots, Q\}$. Assuming a linear-time-invariant system, we can define $D^{(q)}(k, \hat{\boldsymbol{y}}_l)$ as the acoustic transfer function between an incident plane-wave from direction $\hat{\boldsymbol{y}}_l$ and the recording by the microphone at $\boldsymbol{x}_q$. We term $D^{(q)}(k, \hat{\boldsymbol{y}}_l)$ as the Wearable-Device-Related Transfer Function (WDRTF) since it is the direction-dependent response of the microphone incorporating sound wave transformations caused by diffraction, scattering and reflection phenomena from the device's and the wearer's body.

Let $A(k, \hat{\boldsymbol{y}}_l)$ be the strength of the plane-wave arriving from $\hat{\boldsymbol{y}}_l$ as observed at the origin in the absence of the recording device and head. Since any incident sound field can be modeled as a superposition of plane-waves from all possible directions [20], we can express the pressure recorded by the q^{th} microphone on the head-device as a superposition of WDRTF-weighted plane-waves given by

$$P_H(k, \boldsymbol{x}_q) = \int_{\hat{\boldsymbol{y}}_l} A(k, \hat{\boldsymbol{y}}_l) D^{(q)}(k, \hat{\boldsymbol{y}}_l) d\hat{\boldsymbol{y}}_l. \tag{5}$$

In practice, $P_H(k, \boldsymbol{x}_q)$ are the microphone recordings of the head-device and $D^{(q)}(k, \hat{\boldsymbol{y}}_l)$ are the acoustic characteristics of the head-device including the contributions from the wearer's body. The plane-wave density functions $A(k, \hat{\boldsymbol{y}}_l)$ carry the sound field characteristics and are thereby related to the ambisonic coefficients. Hence, we can use the knowledge of $D^{(q)}(k, \hat{\boldsymbol{y}}_l)$ and $P_H(k, \boldsymbol{x}_q)$ to obtain the higher-order ambisonics of the recorded sound field via $A(k, \hat{\boldsymbol{y}}_l)$. This estimation requires the conversion of (5) into SH domain and the process is described in the next three subsections.

A. WDRTF Coefficients

As $D^{(q)}(k, \hat{\boldsymbol{y}}_l)$ is a function defined on a sphere, we can represent it using SH expansion [19] as

$$D^{(q)}(k, \hat{\boldsymbol{y}}_l) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} d_{nm}^{(q)}(k) Y_{nm}(\hat{\boldsymbol{y}}_l) \tag{6}$$

where $d_{nm}^{(q)}(k)$ are WDRTF coefficients of the q^{th} microphone.

Practically, WDRTF can be measured or simulated using similar methods as HRTF [21], but using the microphones on the head-device instead of the in-ear microphones. Considering finite WDRTF samples, we truncate (6) to an order N_d. With $D^{(q)}(k, \hat{\boldsymbol{y}}_l)$ signals for incident plane-wave directions from $l = 1, \cdots, L$, we can estimate $d_{nm}^{(q)}(k)$ to a maximum order $N_{d(max)} = \lfloor \sqrt{L} - 1 \rfloor$ using SH transform [19]. Likewise, we can compute $d_{nm}^{(q)}(k)\ \forall\ q \in \{1, 2, \cdots, Q\}$ to get WDRTF coefficients of all Q microphones on the head-device.

B. Plane-wave Density Coefficients

Similar to (6), since $A(k, \hat{\boldsymbol{y}}_l)$ is a spherical function, we can represent it in SH domain as

$$A(k, \hat{\boldsymbol{y}}_l) = \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} a_{n'm'}(k) Y_{n'm'}(\hat{\boldsymbol{y}}_l) \tag{7}$$

TABLE I: Microphone locations (x_q) of two irregular head-wearable 8-microphone arrays under test

Device	Microphone Locations
Device-1	(0.1436,65°,0°), (0.1440,65°,60°), (0.1401,65°,120°), (0.1441,65°,240°), (0.1427,65°,300°), (0.1100,85°,0°), (0.1109,85°,60°), (0.1122,85°,300°)
Device-2	(0.1127,65°,0°), (0.1100,65°,90°), (0.1109,65°,270°), (0.1036,75°,45°), (0.1040,75°,315°), (0.0943,85°,225°), (0.0943,85°,135°), (0.0903,85°,0°)

where $a_{n'm'}(k)$ are the plane-wave density coefficients that are directly related to the ambisonic coefficients. Substituting (7) and (6) in (5), and using the SH property of

$$\int_{\hat{\boldsymbol{y}}_l} Y_{nm}(\hat{\boldsymbol{y}}_l) Y_{n'm'}(\hat{\boldsymbol{y}}_l) d\hat{\boldsymbol{y}}_l = \begin{cases} (-1)^m & , m = -m'; n = n \\ 0 & , \text{otherwise} \end{cases}, \tag{8}$$

we can express $P_H(k, \boldsymbol{x}_q)$ in SH domain as

$$P_H(k, \boldsymbol{x}_q) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} a_{nm}(k) [\bar{d}_{nm}^{(q)}(k)]^* \tag{9}$$

where $[\bar{d}_{nm}^{(q)}(k)]^* = (-1)^m d_{n(-m)}^{(q)}(k)$. For implementation, we truncate (9) to an order N_a to get a closed-form expression modified into a matrix equation

$$\underbrace{\begin{bmatrix} P_H(k, \boldsymbol{x}_q) \\ \vdots \\ P_H(k, \boldsymbol{x}_Q) \end{bmatrix}}_{\boldsymbol{P}_H(k)_{[Q\times 1]}} = \underbrace{\begin{bmatrix} [\bar{d}_{00}^{(1)}(k)]^* & \cdots & [\bar{d}_{N_a N_a}^{(1)}(k)]^* \\ \vdots & \vdots & \vdots \\ [\bar{d}_{00}^{(Q)}(k)]^* & \cdots & [\bar{d}_{N_a N_a}^{(Q)}(k)]^* \end{bmatrix}}_{\boldsymbol{\Delta}(k)_{[Q\times(N_a+1)^2]}} \underbrace{\begin{bmatrix} a_{00}(k) \\ \vdots \\ a_{N_a N_a}(k) \end{bmatrix}}_{\boldsymbol{a}(k)_{[(N_a+1)^2\times 1]}}. \tag{10}$$

From the recordings of all the Q microphones on the head-device, we can construct $\boldsymbol{P}_H(k)$. Assuming $N_a \leq N_{d(max)}$, we can form $\boldsymbol{\Delta}(k)$ from $d_{nm}^{(q)}(k)$ coefficients of all the microphones which are measured beforehand. The diversity offered by the sound transformations due to the device's and wearer's body will make the $\boldsymbol{\Delta}(k)$ matrix of head-device better conditioned compared to conventional microphone arrays. Moreover, the microphone positions can be designed to optimize the condition of $\boldsymbol{\Delta}(k)$.

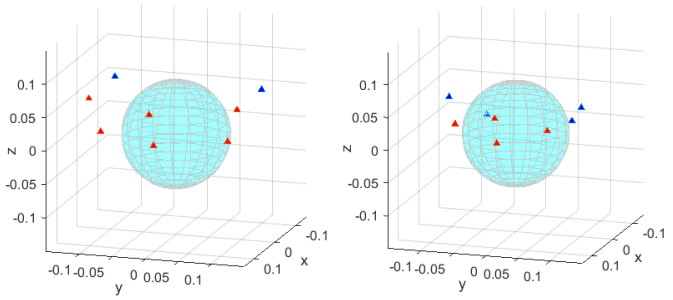
From (10), we can estimate the $a_{nm}(k)$ coefficients by

$$\boldsymbol{a}(k) = \boldsymbol{\Delta}(k)^{\dagger} \boldsymbol{P}_H(k) \tag{11}$$

where $[\cdot]^{\dagger}$ indicate the Moore-Penrose pseudo-inversion operation. For applications of binaural rendering, we will follow $N_a = \lceil keR_h/2 \rceil$ rule [18], where $R_h \approx$ head radius, to control the truncation error even though the human head is not a spherical structure. In addition, the condition of $\boldsymbol{\Delta}(k)$ should be monitored while deciding N_a for each frequency. Hence, for frequencies where $N_a \leq \lfloor \sqrt{Q} - 1 \rfloor$, we can solve (11) in the LS sense and for other frequencies, the solution will be in the Minimum-Norm Least-Squares (MNLS) sense [22]. The maximum N_a should be decided based on the tolerable error depending on the application.

C. Higher-order Sound Field Coefficients

We can establish the relation between the desired $\alpha_{nm}(k)$ coefficients and the estimated plane-wave density coefficients



(a) Device-1 (b) Device-2

Fig. 2: Irregular head-wearable 8-microphone array designs around a spherical-shaped head of radius 8 cm. The y-axis runs along the ears and the positive x-axis is in the look direction. The triangles denote the microphone positions with red ones on the front-side ($+x$).

$a_{nm}(k)$ by representing the sound pressure $P(k, \boldsymbol{x})$ using equivalent plane-wave distribution [20] given by

$$P_{pw}(k, \boldsymbol{x}) = \int_{\hat{\boldsymbol{y}}_l} A(k, \hat{\boldsymbol{y}}_l) e^{ik\hat{\boldsymbol{y}}_l \cdot \boldsymbol{x}} d\hat{\boldsymbol{y}}_l \tag{12}$$

where i is the imaginary unit. Substituting (7) and the SH expansion [17]

$$e^{ik\hat{\boldsymbol{y}}_l \cdot \boldsymbol{x}} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} 4\pi i^n Y_{nm}^*(\hat{\boldsymbol{y}}_l) j_n(k|\boldsymbol{x}|) Y_{nm}(\hat{\boldsymbol{x}}), \tag{13}$$

in (12) and equating with (1), gives the relation

$$\alpha_{nm}(k) = 4\pi i^n a_{nm}(k). \tag{14}$$

In summary, we can obtain the higher-order ambisonics $\alpha_{nm}(k)$ using a head-device through (11) and (14) by knowing its WDRTF coefficients $d_{nm}^{(q)}(k)$ and recordings $P_H(k, \boldsymbol{x}_q)$.

IV. Simulation Study

In this section, we present a preliminary evaluation of the proposed method using two simulated irregular head-wearable 8-microphone arrays. For this study, we assume that the head is a rigid sphere of radius $R_h = 0.08$ m. The two sample head-devices under test, hereafter referred to as Device-1 and Device-2, are illustrated in Fig. 2. The device is mounted on the upper half of the head and the embedded microphone locations with $|\boldsymbol{x}_q| > R_h + 0.01$, as given in Table I, emulate a plausible wearable device structure. For the ease of simulation, we treat the devices as acoustically transparent, but this does not impair the significance of the evaluation. We consider an equivalent rigid SMA with 8 microphones and radius R_h as the

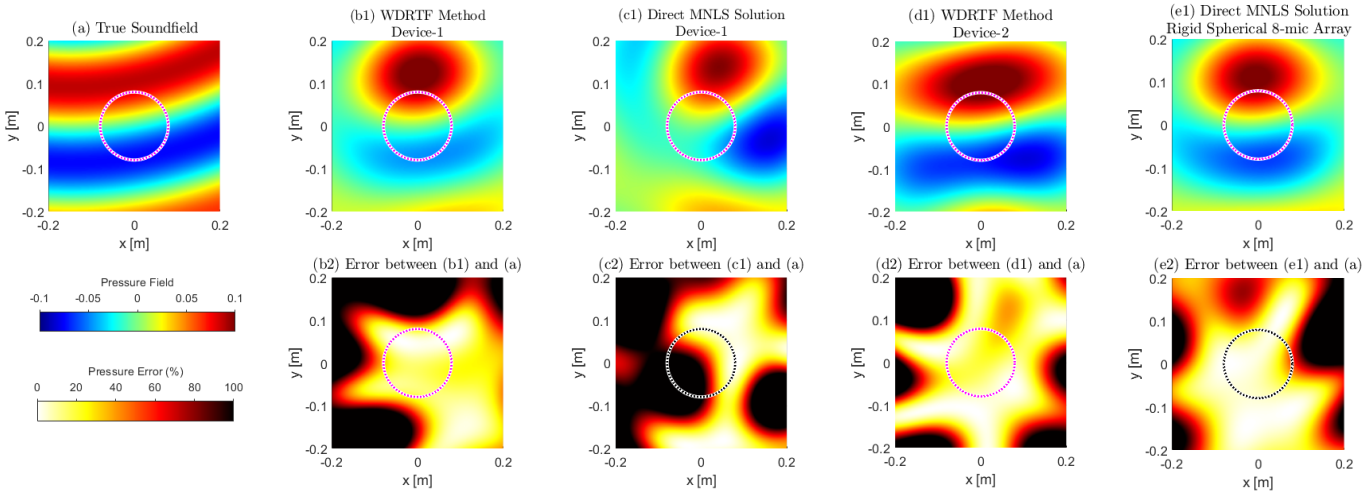


Fig. 3: Sound field reconstruction at 1125 Hz in the x-y plane for a target point-source located at $(-0.15, 0.85, 0.5)$ using 3rd order $\alpha_{nm}(k)$. The circle inside the plots indicate the spherical head region with radius $R_h = 8$ cm.

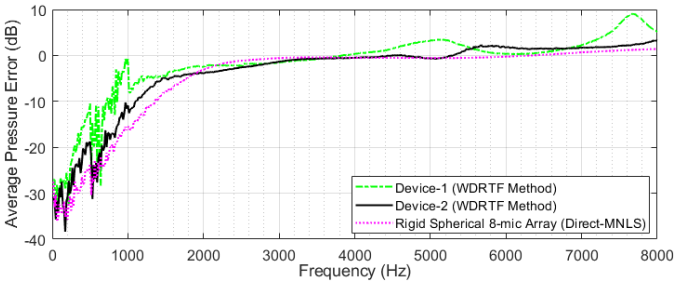


Fig. 4: Average pressure error across frequencies for a point-source located at $\boldsymbol{y}_o = (-0.15, 0.85, 0.5)$.

benchmark. The microphone positions of the SMA are chosen using a uniform spiral-based sphere sampling technique [23].

We simulate WDRTF $D^{(q)}(k, \hat{\boldsymbol{y}}_l)$ as the sound pressure measured by the q^{th} microphone due to a plane-wave from $\hat{\boldsymbol{y}}_l$ direction with sphere scattering consideration using [17]

$$D^{(q)}(k, \hat{\boldsymbol{y}}_l) = \sum_{n=0}^{19} \sum_{m=-n}^{n} \left[4\pi i^n Y^*_{nm}(\hat{\boldsymbol{y}}_l)\right] b_n(k, |\boldsymbol{x}_q|, R_h) Y_{nm}(\hat{\boldsymbol{x}}_q). \tag{15}$$

Similarly, we generate $D^{(q)}(k, \hat{\boldsymbol{y}}_l)\ \forall\ q$ and 400 $\hat{\boldsymbol{y}}_l$ directions distributed across Fliege nodes [24].

The device recordings $P_H(k, \boldsymbol{x}_q)$ are simulated as the incident pressure due to a point-source located at $\boldsymbol{y}_o$ with sphere scattering consideration as [17]

$$P_H(k, \boldsymbol{x}_q) = \sum_{n=0}^{40} \sum_{m=-n}^{n} \left[-ikh_n^{(2)}(k|\boldsymbol{y}_o|) Y^*_{nm}(\hat{\boldsymbol{y}}_o)\right] b_n(k, |\boldsymbol{x}_q|, R_h) Y_{nm}(\hat{\boldsymbol{x}}_q). \tag{16}$$

This point-source is the target source whose sound field is recorded and to be reconstructed from the estimated ambisonics. To account for measurement noise, white Gaussian noise of 30dB SNR is added to both the signals.

For performance evaluation, we use the error metric

$$E(k, x) = \frac{|P_{true}(k, x) - P_{est}(k, x)|^2}{|P_{true}(k, x)|^2} \tag{17}$$

where $P_{true}(k, x)$ is the true free-field sound pressure due to the target point-source and $P_{est}(k, x)$ is the sound pressure reconstructed by substituting the estimated $\alpha_{nm}(k)$ in (1) truncated to corresponding order. We emphasize that the proposed method is not aimed at a perfect sound field reconstruction, but to obtain higher-order ambisonics for a perceptually accurate binaural rendering. Therefore, we use $E(k, x)$ to show the feasibility of irregular head-wearable microphone arrays in ambisonics extraction compared to commonly used SMAs.

Fig. 3 shows the sound field reconstruction performance of the test devices and the SMA in the x-y plane at 1125 Hz for a single point-source located at $(x = -0.15, y = 0.85, z = 0.5)$. An accurate reconstruction at 1125 Hz within a region of radius R_h requires 3rd-order $\alpha_{nm}(k)$ according to the $N = \lceil keR_h/2 \rceil$ rule [18]. Ideally, an SMA of 8-microphones and radius R_h is limited to 1st-order with an upper-frequency limit of ≈ 515 Hz to avoid errors due to spatial aliasing and truncation [4], [18]. However, for a fair comparative study on higher-order ambisonics extraction, we have performed a 3rd-order reconstruction using the test devices and the SMA. In this figure, the WDRTF method refers to the proposed method in Section III and the Direct MNLS solution refers to solving the conventional method in (3) in MNLS sense to estimate $\alpha_{nm}(k)$ beyond the ideal order. In reference to the true incident sound field in Fig. 3a, all the reconstructed sound fields in Fig. 3b - Fig. 3e have wavefront distortions due to truncation and this error deteriorates beyond the array radius. By the case of Device-1, it is apparent that the $\alpha_{nm}(k)$ from the WDRTF method (Fig. 3b) provides better sound field reconstruction than the Direct MNLS solution (Fig. 3c) when dealing with irregular microphone arrays. From Fig. 3b and Fig. 3d, we can observe that Device-2 has more reproduction accuracy and a larger sweet-spot in the x-y plane compared to Device-1, with both using the proposed WDRTF method. Even though both these devices have close condition numbers of $\boldsymbol{\Delta}(k)$, the more number of microphones around the source direction in Device-2 might be the reason for this observation.

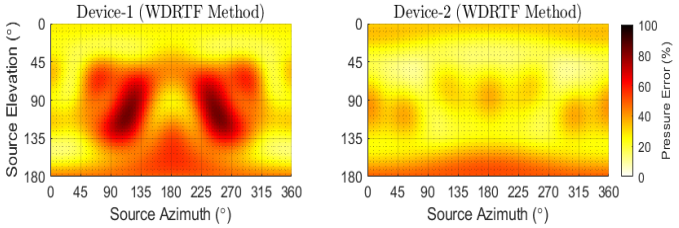


Fig. 5: Average pressure error of 3[rd]-order sound field reconstruction at 1125 Hz for different source positions at 1 m radial distance.

Above all, the performance of the test devices in Fig. 3b and Fig. 3d is acceptable compared to the performance of the SMA in Fig. 3e.

The above observation is further corroborated by Fig. 4 which shows $E(k, x)$ averaged over the spherical head region for different frequencies using maximum 6[th]-order ambisonics. A 6[th]-order reconstruction over a region of radius R_h will have an upper frequency limit of $\approx$ 3 kHz in an ideal case [18]. Apart from the anticipated error increase at high frequencies due to spatial aliasing, the error difference between SMA and Device-2 in Fig. 4 is tolerable. In real scenario, this error is expected to reduce by the advantage of more distinct diffraction and scattering phenomena [12] occurring around the head-device compared to the rigid sphere scattering considered in this simulation study.

The performance of the test devices for different source positions at 1125 Hz is shown in Fig. 5 based on $E(k, x)$ averaged over the spherical head region. The source directions with less number of accessible microphones exhibit more errors. It should be noted that both these devices (Fig. 2) do not cover the full head circumference. The more distributed microphone placements in Device-2 (Fig. 2b) makes it more effective in 3D sound capture than Device-1 (Fig. 2a). Since both these devices are placed on the upper half of the head, source positions around the bottom of the head are difficult to capture. But this is not a probable scenario in the intended Extended Reality (XR) applications. Therefore, we can design such head-devices to attain the optimum spatial coverage within the hardware constraints by utilizing their WDRTF characteristics.

V. Conclusion

In this paper, we presented a sound field decomposition method using microphone array of arbitrary geometry embedded on a head-wearable device. The method exploits the diverse acoustic characteristics imparted by the device's and wearer's body to estimate the higher-order ambisonics. By using this method, properly designed head-wearable devices can replace SMAs without much performance degradation and can thus solely handle both spatial audio capture and binaural rendering. Future work will evaluate the binaural reproduction performance using perceptual listening tests.

References

[1] S. Greengard, *Virtual reality*. Mit Press, 2019.

[2] C. D. Salvador, S. Sakamoto, J. Trevino, and Y. Suzuki, "Design theory for binaural synthesis: Combining microphone array recordings and head-related transfer function datasets," *Acoustical Science and Technology*, vol. 38, no. 2, pp. 51–62, 2017.

[3] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2. IEEE, 2002, pp. II–1781.

[4] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, 2002, pp. 1949–1952.

[5] L. S. Davis, R. Duraiswami, E. Grassi, N. A. Gumerov, Z. Li, and D. N. Zotkin, "High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues," in *Audio Eng. Soc. Conv. 119*. Audio Eng. Soc., 2005.

[6] H. Helmholz, J. Ahrens, D. L. Alon, S. V. A. Garí, and R. Mehra, "Evaluation of sensor self-noise in binaural rendering of spherical microphone array signals," in *IEEE Int. Conf. Acoust., Speech, Signal Process.* IEEE, 2020, pp. 161–165.

[7] Z. Li and R. Ruraiswami, "Hemispherical microphone arrays for sound capture and beamforming," in *IEEE Workshop Applications Signal Process. Audio, Acoust.* IEEE, 2005, pp. 106–109.

[8] T. D. Abhayapala and A. Gupta, "Spherical harmonic analysis of wavefields using multiple circular sensor arrays," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 6, pp. 1655–1666, 2009.

[9] A. Gupta and T. D. Abhayapala, "Double sided cone array for spherical harmonic analysis of wavefields," in *IEEE Int. Conf. Acoust., Speech, Signal Process.* IEEE, 2010, pp. 77–80.

[10] H. Chen, T. D. Abhayapala, and W. Zhang, "Theory and design of compact hybrid microphone arrays on two-dimensional planes for three-dimensional soundfield analysis," *J. Acoust. Soc. Amer.*, vol. 138, no. 5, pp. 3081–3092, 2015.

[11] J. Ahrens, H. Helmholz, D. L. 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. Amer.*, vol. 150, no. 2, pp. 805–815, 2021.

[12] D. S. Talagala, W. Zhang, and T. D. Abhayapala, "Broadband doa estimation using sensor arrays on complex-shaped rigid bodies," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 8, pp. 1573–1585, 2013.

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

[14] H. Gamper, M. R. Thomas, L. Corbin, and I. Tashev, "Synthesis of device-independent noise corpora for realistic asr evaluation." in *INTERSPEECH*, 2016, pp. 2791–2795.

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

[16] J. Ahrens, H. Helmholz, D. L. Alon, and S. V. A. Garí, "A head-mounted microphone array for binaural rendering," in *Immersive 3D Audio: Architecture to Automotive*. IEEE, 2021, pp. 1–7.

[17] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. London, UK: Academic Press, 1999.

[18] R. A. Kennedy, P. Sadeghi, T. D. Abhayapala, and H. M. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2542–2556, 2007.

[19] D. P. Jarrett, E. A. Habets, and P. A. Naylor, *Theory and applications of spherical microphone array processing*. Springer, 2017, vol. 9.

[20] R. Duraiswami, Z. Li, D. Zotkin, E. Grassi, and N. Gumerov, "Plane-wave decomposition analysis for spherical microphone arrays," in *IEEE Workshop Applications Signal Process. Audio, Acoust.*, 2005, pp. 150–153.

[21] B. Xie, *Head-related transfer function and virtual auditory display*. J. Ross Publishing, 2013.

[22] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2013.

[23] A. Semechko, "Suite of functions to perform uniform sampling of a sphere," 2020. [Online]. Available: https://github.com/AntonSemechko/S2-Sampling-Toolbox

[24] J. Fliege and U. Maier, "The distribution of points on the sphere and corresponding cubature formulae," *IMA J. Numer. Anal.*, vol. 19, no. 2, pp. 317–334, 1999.