Periphony-Lattice Mixed-Order Ambisonic Scheme for Spherical Microphone Arrays

Chang, Jiho; Marschall, Marton

Published in:
IEEE/ACM Transactions on Audio, Speech, and Language Processing

Link to article, DOI:
10.1109/TASLP.2018.2800290

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Periphony-Lattice Mixed-Order Ambisonic Scheme for Spherical Microphone Arrays

Jiho Chang, *Member, IEEE*, and Márton Marschall

***Abstract*—Most methods for sound field reconstruction and spherical beamforming with spherical microphone arrays are mathematically based on the spherical harmonics expansion. In many cases, this expansion is truncated at a certain order as in higher order ambisonics (HOA). This truncation leads to performance that is independent of the incident direction of the sound waves. On the other hand, mixed-order ambisonic (MOA) schemes that select an appropriate subset of spherical harmonics can improve the performance for horizontal directions at the expense of other directions. This paper proposes an MOA scheme called Periphony-Lattice to improve sound field reconstruction performance for horizontally incident sound waves. The proposed scheme is compared with the previously introduced MOA and HOA schemes in terms of theoretical truncation error and performance in sound field reconstruction and spherical beamforming. Computer simulations and measurements are conducted with a spherical array of 52 microphones with a nonuniform layout. The results show that the proposed MOA scheme has better performance in sound field reconstruction and spherical beamforming for horizontal sound waves than the other schemes for a given number of microphones. This scheme can be applied to other spherical array layouts if the number of microphones is greater than that of the required spherical harmonics coefficients, and may improve the horizontal performance.**

***Index Terms*—Higher order ambisonics (HOA), mixed-order ambisonics (MOA), spherical beamforming, sound field reconstruction, spherical array of microphones.**

I. Introduction

ANY solution of the wave equation in spherical coordinates can be expressed with spherical Bessel functions and spherical harmonics functions [2]. This expression can be referred to as the spherical harmonics expansion. This expansion has been widely used for sound field reconstruction [3]–[5], spherical beamforming [6]–[9], sound field encoding with spherical microphone arrays [10], [11], and also for sound field reproduction and beamforming with loudspeaker arrays [12]–[16]. The terminology, ambisonics [17], has been used mostly in relation to spatial audio, but in this paper, it refers to all those techniques based on the spherical harmonics expansion.

This expansion includes an infinite number of terms, but in practice the number must be limited. Conventional ambisonics uses four terms up to the first order: a term for the zero-th order represents the omni-directional component, and three terms for the first-order directional components with respect to the x, y, and z axes. Higher-order ambisonics (HOA) uses more terms up to a higher order N, where the total number of terms becomes $(N+1)^2$ because the number of terms at the n-th order is $2n+1$ [18]. This particular truncation results in performance being independent of the direction of the incident sound waves. However, in many cases, performance for horizontally incident waves is more important than that for other directions because relevant sound sources are often placed close to the horizontal plane in real situations, as the human auditory system has better resolution in horizontal directions than in vertical [19]. To improve performance near the horizontal plane, several schemes of mixed-order ambisonics (MOA) have been proposed that select terms relevant to horizontal performance instead of using all terms up to a certain order [21]–[27]. The present study is concerned with MOA schemes. A MOA scheme termed Periphony-Lattice (PL in what follows) is proposed, and compared with other schemes. Periphony refers to sound reproduction both in the horizontal and vertical directions around a listener [20]. The periphonic order is used to indicate the order up to which all terms are considered (i.e, up to which performance is direction independent).

In the literature, the first MOA scheme was introduced by Daniel [21]. Travis categorized MOA schemes into #H, #H#V, #H#P, #H#V#P schemes where H, V, and P stand for horizontal, vertical, and periphony, respectively, and # is an integer [22]. These mixed-order schemes are described in more detail in Section II-B. Several studies have used the #H#P scheme to improve performance measures for horizontal sound waves [23]–[27]. It was shown that this scheme improves the directivity of the spherical beamforming power and the robustness [25]. Sound field reconstruction has also been investigated for the #H#P scheme, but the results were not directly compared with HOA [26]. In fact, it is unclear if the #H#P scheme improves reconstruction performance in the horizontal plane due to the omission of several terms that have considerable contribution in the horizontal plane. Other approaches to favoring

Manuscript received May 18, 2017; revised September 30, 2017 and November 6, 2017; accepted January 16, 2018. Date of publication January 31, 2018; date of current version March 15, 2018. This work was supported by the Korea Research Institute of Standards and Science under Grant 17011008. This paper was presented in part at the 5th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, Honolulu, HI, USA, November/December 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Augusto Sarti. *(Corresponding author: Jiho Chang.)*

J. Chang is with the Korea Research Institute of Standards and Science, Daejeon 34113, South Korea (e-mail: chang.jiho@gmail.com).

M. Marschall is with the Technical University of Denmark, Kongens Lyngby DK-2800, Denmark (e-mail: mm@elektro.dtu.dk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASLP.2018.2800290

or limiting the reconstruction to certain directions with the particular aim of supporting non-uniform loudspeaker layouts include the derivation of new basis functions [28]–[30], and combining ambisonics and panning techniques [31].

This paper proposes a new MOA scheme that contains all relevant terms for expressing sound fields generated by sound sources located in the horizontal plane, in order to improve reconstruction performance. In other words, unlike previously introduced MOA schemes, the proposed subset preserves all information available for horizontal sources. This arrangement differs from the categories introduced by Travis [22] and represents a new category of MOA schemes. Since the selected terms appear in a lattice, the proposed scheme is termed the PL (Periphony-Lattice) scheme. In this paper, the proposed MOA scheme is compared with several alternative MOA schemes, as well as with HOA.A comprehensive set of metrics is applied, including both analytically derived measures such as truncation error, as well as performance measures related to spherical beamforming and sound field reconstruction.

The paper is organized as follows. Section II introduces the considered MOA schemes, and proposes the PL scheme. Section III compares these schemes in terms of the truncation effect and spatial resolution in general. Section IV describes simulations of the performance of an example spherical microphone array of microphones while applying the considered schemes. Section V validates the proposed method with experiments. Section VI discusses relevant issues, and Section VII concludes this study.

II. Theoretical Background

A. Higher-Order Ambisonics (HOA)

An arbitrary incident sound field in a finite region can be expressed in spherical harmonics expansion as [2]

$$P(r,\theta,\phi;\omega)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}A_{mn}(\omega)\,j_n(kr)\,Y_n^m(\theta,\phi), \quad (1)$$

where ω is angular frequency (omitted in what follows for simplicity), k is the wave number, $A_{mn}(\omega)$ is a complex coefficient, θ and ϕ are the declination angle and the azimuth angle, respectively, and $Y_n^m(\theta,\phi)$ are the spherical harmonics,

$$Y_n^m(\theta,\phi)=\sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\cos\theta)\,e^{im\phi}, \quad (2)$$

where P_n^m is the associated Legendre function. A time- harmonic function $e^{-i\omega t}$ is omitted for simplicity. The coefficients $A_{mn}(\omega)$ define the incident sound field. For example, the coefficients for a plane wave propagating in (θ_0,ϕ_0) with unit amplitude are given by $4\pi i^n Y_n^m(\theta_0,\phi_0)^*$ [7], where $*$ indicates the conjugate operation.

Spherical harmonics can be categorized into zonal, sectoral, and tesseral harmonics. Zonal harmonics are components where the index $m=0$, and sectoral harmonics are components where m is equal to $\pm n$. The remaining components are the tesseral harmonics. It is known that sectoral harmonics are dominant for horizontal representation, whereas more tesseral components contribute to better vertical performance [20].

In the HOA approach, all three harmonic types up to a certain order N are included,

$$\widehat{P}(r,\theta,\phi)\cong\sum_{n=0}^{N}\sum_{m=-n}^{n}A_{mn}j_n(kr)\,Y_n^m(\theta,\phi). \quad (3)$$

Thus, HOA has direction independent performance properties. The total number of the terms in this case is $(N+1)^2$.

B. Mixed-Order Ambisonics (MOA)-PHV Scheme

Travis categorized MOA schemes based on three parameters that indicate the order for periphonic, horizontal, and vertical terms [20]. To clarify each parameter, let us define N_{P}, N_{H}, and N_{V}, respectively. N_{P} is an order up to which all terms are chosen so that the performance is periphonic (independent of direction) with the selected terms. N_{H} is a higher order than N_{P}, and only some terms are selected at the orders from $N_{\mathrm{P}}+1$ to N_{H} depending on another order, N_{V}. N_{V} is an integer from zero to N_{P}. If N_{V} is zero, only sectoral terms are selected, and horizontal performance is improved because sectoral harmonics are dominant in the horizontal plane. If N_{V} is not zero, tesseral terms of $|m|\geq n-N_V$ are added, which improves vertical performance, and provides a smoother transition between performance for horizontal and elevated sources [20].

A MOA scheme that has non-zero N_{P}, N_{H}, and N_{V} is called a PHV (Periphonic-Horizontal-Vertical) scheme, and labelled as $N_{\mathrm{P}}^{(\mathrm{P})}N_{\mathrm{H}}^{(\mathrm{H})}N_{\mathrm{V}}^{(\mathrm{V})}$ in this study (similarly to the #H#V#P notation used in [20]). A sound field can be expressed with this scheme as

$$\begin{aligned}\widehat{P}_{N_P{}^{(\mathrm{P})}N_H{}^{(\mathrm{H})}N_V{}^{(\mathrm{V})}}(r,\theta,\phi)&=\sum_{n=0}^{N_P}\sum_{m=-n}^{n}A_{mn}j_n(kr)Y_n^m(\theta,\phi)\\&+\sum_{n=N_P+1}^{N_H}\sum_{m=n-N_V}^{n}\big[A_{mn}j_n(kr)\,Y_n^n(\theta,\phi)\\&+A_{-m,n}j_n(kr)\,Y_n^{-n}(\theta,\phi)\big].\end{aligned} \quad (4)$$

The total number of the terms is $(N_{\mathrm{P}}+1)^2+2(N_{\mathrm{V}}+1)(N_{\mathrm{H}}-N_{\mathrm{P}})$. For example, Fig. 1 illustrates the $3^{(\mathrm{P})}5^{(\mathrm{H})}1^{(\mathrm{V})}$ scheme. Each block stands for a spherical harmonics component of order $(m,\,n)$. The selected components in this scheme are shaded. Only terms for $m\geq 0$ are shown, since all schemes are symmetric around m = 0. That is, in this figure, 14 components are shown, but the total number is 24 including the components with $m<0$.

If $N_{\mathrm{V}}^{(\mathrm{V})}$ is zero, this term can be omitted, and then the scheme can be called a PH (Periphonic-Horizontal, labelled as #H#P in [22]) scheme. In this paper, this scheme is designated as $N_{\mathrm{P}}^{(\mathrm{P})}N_{\mathrm{H}}^{(\mathrm{H})}$. The sound field in this case is given by

$$\begin{aligned}\widehat{P}_{N_P{}^{(\mathrm{P})}N_H{}^{(\mathrm{H})}}(r,\theta,\phi)&=\sum_{n=0}^{N_P}\sum_{m=-n}^{n}A_{mn}j_n(kr)\,Y_n^m(\theta,\phi)\\&+\sum_{n=N_P+1}^{N_H}\big[A_{nn}j_n(kr)\,Y_n^n(\theta,\phi)+A_{-n,n}j_n(kr)Y_n^{-n}(\theta,\phi)\big].\end{aligned} \quad (5)$$

The total number of terms becomes $N_{\mathrm{P}}^2+2N_{\mathrm{H}}+1$.

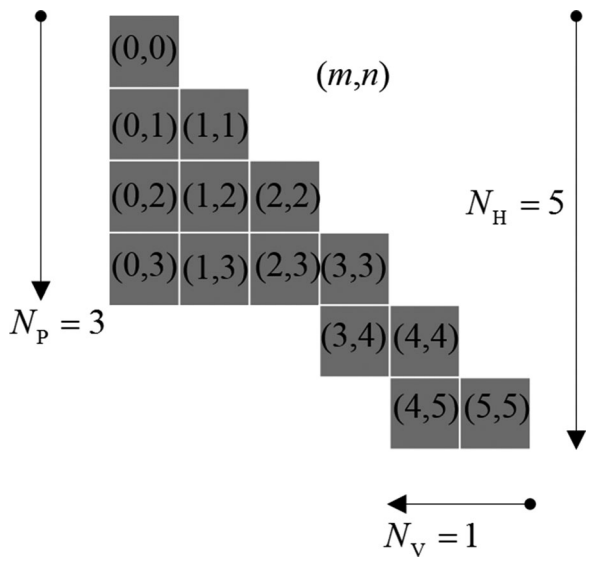

Fig. 1. PHV scheme (shown is $3^{(P)}5^{(H)}1^{(V)}$), left hand side (m < 0) is omitted.

If instead N_P is zero, the $N_P^{(P)}$ term could be omitted from the label, as in [22], which also defines H and HV schemes. However, N_P can also be considered to be unity. For example, $0^{(P)}2^{(H)}1^{(V)}$ is equivalent to $1^{(P)}2^{(H)}1^{(V)}$. This means that H and HV schemes are included in the PH and PHV categories. Here, for the sake of clarity, we define N_P as a non-zero integer $(N_P > 0)$.

In addition, N_v is defined to be equal to, or less than N_P, since otherwise, $1^{(P)}3^{(H)}2^{(V)}$ and $2^{(P)}3^{(H)}2^{(V)}$ for example, describe equivalent schemes.

If N_H and N_V are not defined, the PHV scheme is reduced to N_P-th order HOA. In other words, N_P-th order HOA can be labelled in the same system, as $N_P{}^{(P)}$

C. Mixed-Order Ambisonics (MOA)-PL Scheme

To improve the reconstruction performance, all terms relevant to the horizontally propagating waves need to be selected. The terms that are related to horizontal waves can be identified by inspecting the coefficients A_{mn} in (3), which define the incident sound waves. For example, the coefficient of a plane wave has the angular component $Y_n^m(\theta_0, \phi_0)^*$, where (θ_0, ϕ_0) is the propagating direction. A spherical wave contains $Y_n^m(\theta_s, \phi_s)^*$ as the angular component in the coefficients where (θ_s, ϕ_s) is the position of the source. Thus, with $\theta_0 = \pi/2$, i.e., a plane wave propagating or a monopole placed in the horizontal plane, the term $P_n^m(\cos(\pi/2))$ is included in the coefficients (cf. (2)).

Since $P_n^m(\cos(\pi/2))$ has a value of zero when $n + m$ is an odd number, the coefficients will also take a value of zero for these indices. As an example, Fig. 2 shows the magnitude of $Y_n^m(\pi/2, 0)^*$ with respect to the indices n and m. Note that grid points for $|m| > n$ do not have defined values. Non-zero values appear in lattice where $P_n^m(0) \neq 0$. This means that a sound field of horizontal waves can be expressed by just the non-zero terms without a loss of information. Hence, sound fields produced

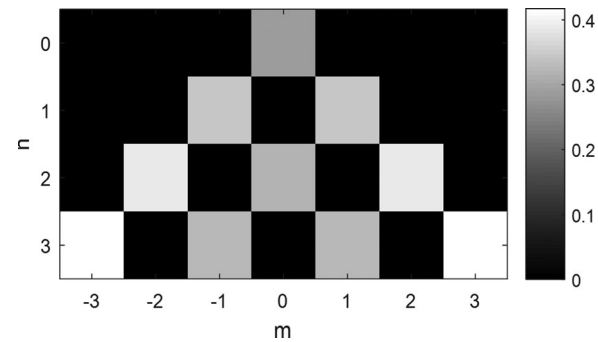

Fig. 2. Magnitude of $Y_n^m(\pi/2, 0)^*$.

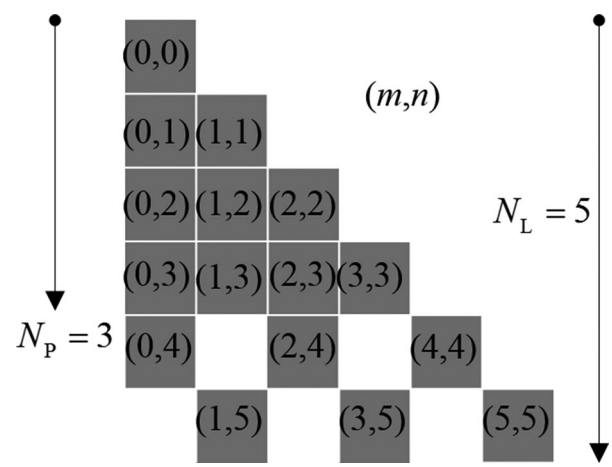

Fig. 3. PL scheme (shown in $3^{(P)}5^{(L)}$), left hand side $(m < 0)$ is omitted.

by sources in the horizontal plane can be efficiently described by using the above terms only. Based on these considerations, a new scheme, termed periphony-lattice (PL), can be defined where L stands for lattice, and labelled as $N_P^{(P)}N_L^{(L)}$. The sound field in this scheme can be expressed as

$$\widehat{P}_{N_P{}^{(P)}N_L{}^{(L)}}(r, \theta, \phi) = \sum_{n=0}^{N_P} \sum_{m=-n}^{n} A_{mn} j_n(kr) Y_n^m(\theta, \phi) + \sum_{n=N_P+1}^{N_L} \sum_{m=0}^{n} A_{2m-n,n} j_n(kr) Y_n^{2m-n}(\theta, \phi), \quad (6)$$

and the total number of terms is $(N_P + 1)^2 + (N_P + N_L + 3)(N_L - N_P)/2$. Fig. 3 shows an example of this scheme $(3^{(P)}5^{(L)})$.

Note that the sectoral harmonics have the largest values in general in Fig. 2, but other non-zero components also have significant contributions. By selecting only sectoral harmonics at orders above N_P, the PH scheme can reduce the reconstruction error to some extent, compared with the $N_P^{(P)}$ scheme. However, the PL scheme is expected to have lower reconstruction error than the PH scheme because of the inclusion of all relevant terms to horizontal sound waves.

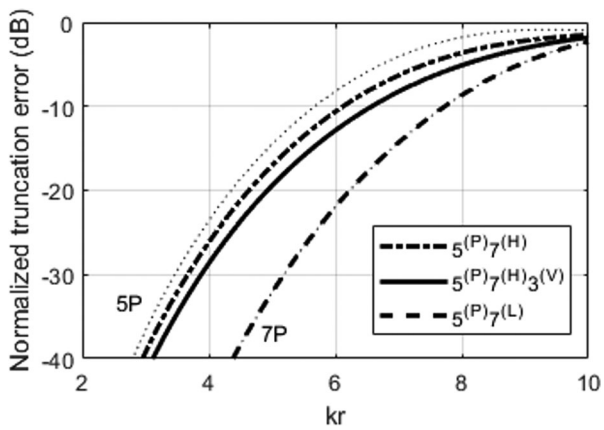

Fig. 4. Normalized truncation error of a plane wave with several schemes.

III. Comparison of Effect of Truncation

A. Truncation Error

The effect of truncation in the spherical harmonics expansion can be shown with the normalized truncation error and the normalized directivity. The normalized truncation error is defined as [13]

$$\bar{e}^2(kr) = \frac{\int_0^{\pi}\int_0^{2\pi}\left|P(r,\theta,\phi) - \widehat{P}(r,\theta,\phi)\right|^2 \sin\theta d\phi d\theta}{\int_0^{\pi}\int_0^{2\pi}\left|P(r,\theta,\phi)\right|^2 \sin\theta d\phi d\theta}, \quad (7)$$

and for a plane wave in HOA, the error is reduced to [13]

$$\bar{e}^2_{N^{(\mathrm{P})}}(kr) = 1 - \sum_{n=0}^{N}(2n+1)\, {j_n}^2(kr), \quad (8)$$

which is independent of the propagating direction. In contrast, the truncation error of the $N_{\mathrm{P}}^{(\mathrm{P})}N_{\mathrm{H}}^{(\mathrm{H})}N_{\mathrm{V}}^{(\mathrm{V})}$ scheme becomes

$$\bar{e}^2_{N_P{}^{(\mathrm{P})}N_H{}^{(\mathrm{H})}N_V{}^{(\mathrm{V})}}(kr) = 1 - \sum_{n=0}^{N_P}(2n+1)\, {j_n}^2(kr) + 4\pi \sum_{n=N_H+1}^{N_P}\sum_{m=-n+1+N_V}^{n-1-N_V}\left|Y_n^m(\theta_0,\phi_0)\right|^2 {j_n}^2(kr), \quad (9)$$

where θ_0 and ϕ_0 are the declination and the azimuth angles of the propagating direction of the plane wave, respectively. See Appendix for the detailed derivations. The truncation error of $N_{\mathrm{P}}^{(\mathrm{P})}N_{\mathrm{L}}^{(\mathrm{L})}$ scheme is

$$\bar{e}^2_{N_P{}^{(\mathrm{P})}N_L{}^{(\mathrm{L})}}(kr) = 1 - \sum_{n=0}^{N_L}(2n+1)\, {j_n}^2(kr) + 4\pi \sum_{n=N_P+1}^{N_L}\sum_{m=0}^{n-1}\left|Y_n^{2m-n+1}(\theta_0,\phi_0)\right|^2 {j_n}^2(kr). \quad (10)$$

If θ_0 is equal to $\pi/2$, this error is equal to the truncation error for N_{L}-th order HOA, $\bar{e}^2_{N_L{}^{(\mathrm{P})}}$, because the result of the second summation becomes zero.

Fig. 4 shows the truncation errors for two example schemes, $5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$ and $5^{(\mathrm{P})}7^{(\mathrm{L})}$ for a plane wave with $(\theta_0, \phi_0) = (\pi/2, 0)$. These schemes are chosen for a fair comparison as the number of the terms is similar (52 and 51, respectively). Truncation errors for $5^{(\mathrm{P})}, 7^{(\mathrm{P})}$ (dotted lines), and $5^{(\mathrm{P})}7^{(\mathrm{H})}$ are also shown for comparison. The truncation errors for $5^{(\mathrm{P})}7^{(\mathrm{H})}$ and $5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$ are smaller than that for $5^{(\mathrm{P})}$, but greater than for $7^{(\mathrm{P})}$. This shows that the more terms are added, the smaller the truncation error becomes. It is notable that the error for the PL scheme $5^{(\mathrm{P})}7^{(\mathrm{L})}$ is, however, equal to that of 7th order HOA despite having a smaller number of terms than either $5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$ or $7^{(\mathrm{P})}$. Table I shows these example schemes, which will be used in the following.

TABLE I
Example Schemes of HOA and MOA

Scheme	Number of terms	DI† (horizontal)	DI† (vertical)	Min. number of Mic. †
$5^{(\mathrm{P})}$ (HOA)	36	10.29	7.41	36
$5^{(\mathrm{P})}7^{(\mathrm{H})}$ (MOA)	40	11.66	7.29	40
$6^{(\mathrm{P})}$ (HOA)	49	11.00	8.10	51
$5^{(\mathrm{P})}7^{(\mathrm{L})}$ (MOA)	51	11.61	8.68	52
$5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$ (MOA)	52	11.69	7.76	54
$7^{(\mathrm{P})}$ (HOA)	64	11.61	8.68	64

†DI means directivity index. For details on DI values, see Section III-B. For minimum number of microphones (the last row), see Section V-B. Minimum number of microphones are calculated based on nearly-uniform layouts.

B. Normalized Directivity

Another measure to evaluate MOA schemes is the normalized directivity [6]–[9]. The spherical harmonics function has the completeness relation,

$$\sum_{n=0}^{\infty}\sum_{m=-n}^{n} Y_n^m(\theta,\phi)\, Y_n^m(\theta_0,\phi_0)^* = \delta(\phi-\phi_0)\,\delta(\cos\theta - \cos\theta_0)\,. \quad (11)$$

If the summation is truncated, it does not equate to a delta function in (11), but will exhibit a main lobe and side lobes. The width of the main lobe and the maximum level of the side lobes are widely used as performance indices of the directivity [6]–[9]. The directivity is considered higher when the width of the main lobe is narrower, and the maximum side lobe level is lower.

The directivity of the N-th order HOA can be defined as

$$w_{N^{(\mathrm{P})}}(\theta,\phi) = \sum_{n=0}^{N}\sum_{m=-n}^{n} Y_n^m(\theta,\phi)\, Y_n^m(\theta_0,\phi_0)^*. \quad (12)$$

Likewise, the directivity of MOA schemes can be defined as

$$w_{N_P{}^{(\mathrm{P})}N_H{}^{(\mathrm{H})}N_V{}^{(\mathrm{V})}}(\theta,\phi) = \sum_{n=0}^{N_P}\sum_{m=-n}^{n} Y_n^m(\theta,\phi)\, Y_n^m(\theta_0,\phi_0)^* + \sum_{n=N_P+1}^{N_H}\sum_{m=n-N_V}^{n}\left[Y_n^m(\theta,\phi)\, Y_n^m(\theta_0,\phi_0)^* + Y_n^{-m}(\theta,\phi)\, Y_n^{-m}(\theta_0,\phi_0)^*\right], \quad (13)$$

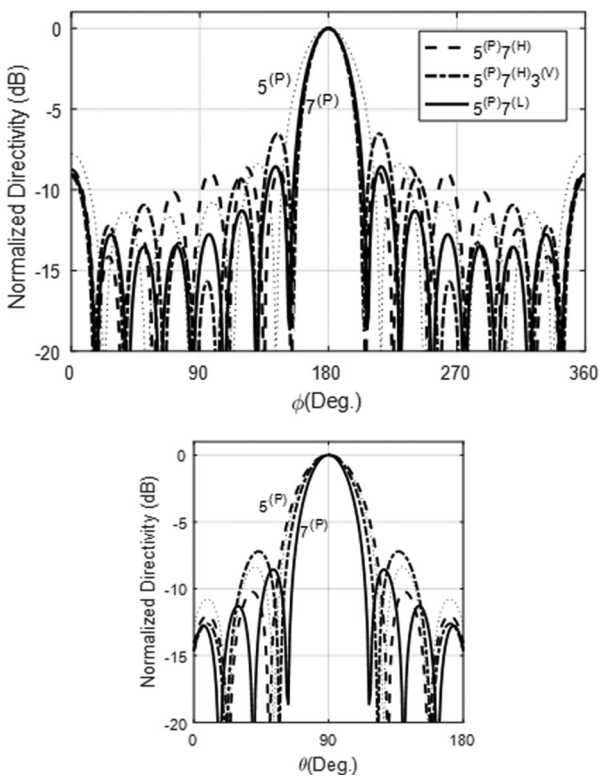

Fig. 5. Normalized directivities of different schemes with respect to azimuth angle (top) and declination angle (bottom).

$$w_{N_P{}^{(\mathrm{P})}N_L{}^{(\mathrm{L})}}(\theta,\phi) = \sum_{n=0}^{N_P}\sum_{m=-n}^{n} Y_n^m(\theta,\phi)\, Y_n^m(\theta_0,\phi_0)^* + \sum_{n=N_P+1}^{N_L}\sum_{m=0}^{n} Y_n^{2m-n}(\theta,\phi)\, Y_n^{2m-m}(\theta_0,\phi_0)^*. \quad (14)$$

Again, if θ_0 is equal to $\pi/2$, the directivity of $N_{\mathrm{P}}^{(\mathrm{P})}N_{\mathrm{L}}^{(\mathrm{L})}$ is equal to that of N_{L}-th order HOA because the result of the second summation in (14) is zero.

Fig. 5 shows the normalized directivities of three MOA schemes, $5^{(\mathrm{P})}7^{(\mathrm{H})}$, $5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$, and $5^{(\mathrm{P})}7^{(\mathrm{L})}$ with respect to the azimuth angle ϕ (top) and the declination angle θ (bottom), respectively. The reference direction $(\theta_0,\ \phi_0)$ is $(\pi/2, \pi)$. Fig. 5 (top) is the horizontal slice at $\theta = \theta_0$, and Fig. 5 (bottom) is the vertical slice of the directivity function at $\phi = \phi_0$. The directivities of HOA ($5^{(\mathrm{P})}$ and $7^{(\mathrm{P})}$) are indicated by the dotted lines.

In Fig. 5 (top), the widths of the main lobe of the MOA schemes are similar to that of HOA $7^{(\mathrm{P})}$. Unlike in terms of truncation error, $5^{(\mathrm{P})}7^{(\mathrm{H})}$ shows significant improvement compared with $5^{(\mathrm{P})}$ in terms of directivity. This can be explained by the maximum spatial frequency, which is related to the order $|m|$ $(=7)$ in this case. That is, the azimuthal component of each term is $e^{im(\phi-\phi_0)}$, and the greater the maximum $|m|$, the narrower the main lobe tends to be.

The directivity index (DI) is a widely used parameter to quantify the directivity. DI with respect to the azimuth angle can be defined as [18], [32]:

$$DI^{(\mathrm{H})} = 10\log_{10}\left[\frac{2\pi|w(\theta_0,\phi_0)|^2}{\int_0^{2\pi}|w(\theta_0,\phi)|^2 d\phi}\right]. \quad (15)$$

DI around the azimuth for $5^{(\mathrm{P})}$ and $7^{(\mathrm{P})}$ are 10.29 dB and 11.61 dB, respectively. $5^{(\mathrm{P})}7^{(\mathrm{L})}$ results in the same DI as $7^{(\mathrm{P})}$. DIs for $5^{(\mathrm{P})}7^{(\mathrm{H})}$ and $5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$ are 11.66 dB and 11.69 dB, respectively, as shown in Table I. DI around the azimuth does not vary with the looking direction, ϕ_0.

On the other hand, looking at directivity along the declination angle (Fig. 5, bottom), $5^{(\mathrm{P})}7^{(\mathrm{L})}$ has a narrower main lobe width than the other two schemes, showing that vertical directivity is better controlled with this scheme. In the declination direction, the maximum of $n-|m|$ determines the maximum spatial frequency. $5^{(\mathrm{P})}7^{(\mathrm{L})}$ has the same maximum value of $n-|m|$ $(=6)$ as $7^{(\mathrm{P})}$, while $5^{(\mathrm{P})}7^{(\mathrm{H})}$ and $5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$ have a smaller value, 5. Sectoral harmonics have zero value for $n-|m|$ by definition, and thus $N_{\mathrm{P}}^{(\mathrm{P})}N_{\mathrm{H}}^{(\mathrm{H})}$ always has the same maximum spatial frequency along the declination as $N_{\mathrm{P}}^{(\mathrm{P})}$. Vertical plane DIs can be defined as

$$DI^{(\mathrm{V})}(\theta_0) = 10\log_{10}\left[\frac{\pi|w(\theta_0,\phi_0)|^2}{\int_0^{\pi}|w(\theta,\phi_0)|^2 d\theta}\right], \quad (16)$$

where it depends on the looking direction θ_0, which is $\pi/2$ in this case. Those for $5^{(\mathrm{P})}$ and $7^{(\mathrm{P})}$ are 7.41 dB and 8.68 dB, respectively. The DI for $5^{(\mathrm{P})}7^{(\mathrm{H})}$ is lower than that of $5^{(\mathrm{P})}$ (7.29 dB). $5^{(\mathrm{P})}7^{(\mathrm{L})}$ has the same value as $7^{(\mathrm{P})}$, and $5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$ has 7.76 dB. These results are also shown in Table I.

IV. Ambisonic Encoding With Spherical Microphone Arrays

A. Spherical Microphone Array and Simulation Setup

In order to compare the MOA and HOA schemes, a spherical microphone array is employed, and the performance in terms of sound field reconstruction and spherical beamforming is compared between the schemes using simulations (Section IV) as well as experimental results (Section V). The simulation analysis is conducted in a similar way as in Section III; a plane wave is assumed for the reference field, and the reconstruction error and the normalized directivity are compared. However, the coefficients used are not the theoretical values, but estimated from the surface pressure on the spherical array of microphones. Hence, these simulations exhibit spatial aliasing at high frequencies, and the results are frequency dependent. The effect of noise is considered in Section V.

A rigid-sphere microphone array that was introduced in previous studies was considered [23], [24], which has a radius of 5 cm, and the total number of microphones is 52. The layout consists of rings of 2, 6, 10, 16, 10, 6 and 2 microphones at declination angles of 10, 35, 61, 90, 119, 145 and 170 degrees, respectively. The layout thus provides a somewhat denser arrangement of microphones around the equator.

If a rigid sphere of radius a is located at the origin, the sound pressure on the surface due to a plane wave propagating in $(\theta_0,\ \phi_0)$ direction can be approximated with an effective

order N_{eff},

$$P_{tot}\left(r=a,\theta,\phi\right) \cong \sum_{n=0}^{N_{\text{eff}}} \sum_{m=-n}^{n} A_{mn} b_n\left(ka\right) Y_n^m\left(\theta,\phi\right), \quad (17)$$

where

$$b_n\left(ka\right) = j_n\left(ka\right) - \frac{j'_n\left(ka\right)}{h_n^{(1)\prime}\left(ka\right)} h_n^{(1)\prime}\left(ka\right) = \frac{i}{\left(ka\right)^2 h_n^{(1)\prime}\left(ka\right)}. \quad (18)$$

For an infinite N_{eff}, the equality is satisfied. For the simulation, a finite value that is large enough to neglect the truncation error should be chosen. N_{eff} was chosen as $\lceil 2ka+1 \rceil$ where $\lceil \cdot \rceil$ means the smallest integer that is greater than $2ka+1$. This gives less than 0.1% error at all frequencies according to (8). That is, P_{tot} with $N_{\text{eff}} = \lceil 2ka+1 \rceil$ can be assumed to be sufficiently close to the true value in the simulations.

B. Estimation of the Coefficients

The sound pressure on the surface and the coefficients A_{mn} have the following relation by the orthonormality of Y_n^m:

$$A_{mn} = \int_0^{2\pi} \int_0^{\pi} P_{tot}\left(r=a,\theta,\phi\right) \frac{Y_n^m\left(\theta,\phi\right)^*}{b_n\left(ka\right)} \sin\theta d\theta d\phi. \quad (19)$$

In practice, the sound pressure is sampled at a limited number of positions, and the coefficients can be estimated by discretizing (19). However, suboptimal sampling leads to orthonormality errors [8]. Instead, we introduce an unknown function for the q-th microphone position, $W_{mn}(\theta^{(q)},\phi^{(q)})$, and obtain it to minimize the orthonormality error after sampling:

$$\tilde{A}_{mn} = \sum_{q=1}^{Q} P_{tot}\left(r=a,\theta^{(q)},\phi^{(q)}\right) \frac{W_{mn}\left(\theta^{(q)},\phi^{(q)}\right)}{b_n\left(ka\right)}, \quad (20)$$

where Q is the total number of microphones. Substituting (17) leads to

$$\tilde{A}_{mn} = \sum_{q=1}^{Q} \sum_{n'=0}^{N_{\text{eff}}} \sum_{m'=-n'}^{n'} A_{m'n'} b_{n'}\left(ka\right) Y_{n'}^{m'}\left(\theta^{(q)},\phi^{(q)}\right) \times \frac{W_{mn}\left(\theta^{(q)},\phi^{(q)}\right)}{b_n\left(ka\right)}. \quad (21)$$

In order for $A_{m'n'}$ in this equation to be equal to $\tilde{A}_{mn}$ on the left-hand side, the following has to be satisfied:

$$\sum_{q=1}^{Q} \sum_{n'=0}^{N_{\text{eff}}} \sum_{m'=-n'}^{n'} \frac{b_{n'}\left(ka\right)}{b_n\left(ka\right)} Y_{n'}^{m'}\left(\theta^{(q)},\phi^{(q)}\right) W_{mn}\left(\theta^{(q)},\phi^{(q)}\right) = \frac{b_{n'}\left(ka\right)}{b_n\left(ka\right)}\left(\delta_{nn'}\delta_{mm'} + \varepsilon_{nn'}^{mm'}\right), \quad (22)$$

where the orthonormality error $\varepsilon_{nn'}^{mm'} \ll 1$. That is,

$$\tilde{A}_{mn} = A_{mn} + \sum_{q=1}^{Q} \sum_{n'=0}^{N_{\text{eff}}} \sum_{m'=-n'}^{n'} A_{m'n'} \frac{b_{n'}\left(ka\right)}{b_n\left(ka\right)} \varepsilon_{nn'}^{mm'}. \quad (23)$$

The second term on the right hand side is the orthonormality error noise (OEN) as defined in [8, eq. (23)]. To obtain $W_{mn}(\theta^{(q)},\phi^{(q)})$ that minimizes the OEN, these equations are expressed as a matrix equation. Equation (17) is expressed as

$$\mathbf{P} = \mathbf{Y}_{QK'} \cdot \mathbf{B}_{K'} \cdot \mathbf{A}_{K'} \quad (24)$$

where

$$\mathbf{P} = \left[P_{tot}\left(a,\theta^{(1)},\phi^{(1)}\right) \cdots P_{tot}\left(a,\theta^{(Q)},\phi^{(Q)}\right) \right]^T \quad (25)$$

$$\mathbf{Y}_{QK'} = \begin{bmatrix} Y_0^0\left(\theta^{(1)},\phi^{(1)}\right) & \cdots & Y_{N_{\text{eff}}}^{N_{\text{eff}}}\left(\theta^{(1)},\phi^{(1)}\right) \\ \vdots & \ddots & \vdots \\ Y_0^0\left(\theta^{(Q)},\phi^{(Q)}\right) & \cdots & Y_{N_{\text{eff}}}^{N_{\text{eff}}}\left(\theta^{(Q)},\phi^{(Q)}\right) \end{bmatrix} \quad (26)$$

$$\mathbf{B}_{K'} = \text{diag}\left[b_{n'}\left(ka\right)\right]_{K'} \quad (27)$$

$$\mathbf{A}_{K'} = \left[A_{00} \quad \cdots \quad A_{N_{\text{eff}},N_{\text{eff}}} \right]^T. \quad (28)$$

Each scheme estimates a different number of coefficients, K. Let the number of terms be K $(K \le Q)$, (20) is expressed as

$$\tilde{\mathbf{A}}_K = \mathbf{B}_K^{-1} \cdot \mathbf{W}_{KQ} \cdot \mathbf{P} \quad (29)$$

where $\tilde{\mathbf{A}}_K$ is a column vector of the estimated coefficients, $\mathbf{W}_{KQ}$ is a K by Q matrix, and $\mathbf{B}_K^{-1}$ is $\text{diag}\left[b_n(ka)^{-1}\right]_K$. To reduce the effect of noise, $\mathbf{B}_K^{-1}$ can be determined as

$$\mathbf{B}_K^{-1} = \text{diag}\left[\frac{b_n^*\left(ka\right)}{\left|b_n\left(ka\right)\right|^2 + \lambda^2}\right]_K, \quad (30)$$

instead of $\text{diag}\left[b_n(ka)^{-1}\right]_K$, where λ is the regularization parameter.

Inserting (24) into (29) yields

$$\tilde{\mathbf{A}}_K = \mathbf{B}_K^{-1} \cdot \mathbf{W}_{KQ} \cdot \mathbf{Y}_{QK'} \cdot \mathbf{B}_{K'} \cdot \mathbf{A}_{K'}. \quad (31)$$

Depending on K and K', (31) is reduced as follows.

1) $K' < K$ (Low Frequencies): In order to accurately estimate the coefficients the following relation has to be be satisfied:

$$\mathbf{B}_K^{-1} \cdot \mathbf{W}_{KQ} \cdot \mathbf{Y}_{QK'} \cdot \mathbf{B}_{K'} = \begin{bmatrix} \mathbf{I}_{K'} \\ \mathbf{0} \end{bmatrix}, \quad (32)$$

where $\mathbf{I}_{K'}$ is an identity matrix with the size of K' by K'. Since the size of $\mathbf{Y}_{QK'}$ is Q by K' $(K' < Q)$, this equation is underdetermined. Thus, there exists a matrix $\mathbf{W}_{KQ}$ that satisfies (32), which is

$$\mathbf{W}_{KQ} = \mathbf{B}_K \begin{bmatrix} \mathbf{I}_{K'} \\ \mathbf{0} \end{bmatrix} \mathbf{B}_{K'}^{-1} \mathbf{Y}_{QK'}^{+} = \begin{bmatrix} \mathbf{Y}_{QK'}^{+} \\ \mathbf{0} \end{bmatrix}. \quad (33)$$

Then, (31) is reduced to

$$\tilde{\mathbf{A}}_K = \mathbf{B}_K^{-1} \cdot \begin{bmatrix} \mathbf{Y}_{QK'}^{+} \\ \mathbf{0} \end{bmatrix} \cdot \mathbf{Y}_{QK'} \cdot \mathbf{B}_{K'} \cdot \mathbf{A}_{K'} = \begin{bmatrix} \mathbf{B}_{K'}^{-1} \mathbf{Y}_{QK'}^{+} \mathbf{Y}_{QK'} \cdot \mathbf{B}_{K'} \cdot \mathbf{A}_{K'} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{K'} \\ \mathbf{0} \end{bmatrix}. \quad (34)$$

2) $K' = K$ (Middle Frequencies): This condition leads to the relation:

$$\mathbf{B}_K^{-1}\mathbf{W}_{KQ}\mathbf{Y}_{QK'}\cdot\mathbf{B}_{K'} = \mathbf{I}_K. \tag{35}$$

Now, Q is equal to or greater than K'. This means that this equation is uniquely or under-determined, and there exists a matrix $\mathbf{W}_{KQ}$:

$$\mathbf{W}_{KQ} = \mathbf{B}_K\mathbf{B}_K^{-1}\mathbf{Y}_{QK}^{+} = \mathbf{Y}_{QK}^{+}. \tag{36}$$

The column vectors are assumed to be independent. In this case, the estimated coefficients are exact:

$$\tilde{\mathbf{A}}_K = \mathbf{A}_{K'}.$$

3) $K < K'$ (High Frequencies): The following condition has to be satisfied to estimate accurate coefficients:

$$\mathbf{B}_K^{-1}\mathbf{W}_{KQ}\mathbf{Y}_{QK'}\cdot\mathbf{B}_{K'} = \begin{bmatrix}\mathbf{I}_K & \mathbf{0}\end{bmatrix}. \tag{37}$$

This equation can be under-, uniquely, or over-determined depending on Q and K'. In the first two cases, $\mathbf{W}_{KQ}$ can be obtained that satisfies (37). On the other hand, in the over-determined case, there is no $\mathbf{W}_{KQ}$ that satisfies (37). Instead, a matrix that minimizes the error can be obtained, as in [8].

$$\tilde{\mathbf{W}}_{KQ} = \mathbf{B}_K\mathbf{I}_K\mathbf{B}_K^{-1}\mathbf{Y}_{QK}^{+} = \mathbf{Y}_{QK}^{+} \tag{38}$$

where $\mathbf{B}_K^{-1}\tilde{\mathbf{W}}_{KQ}\mathbf{Y}_{QK'}\cdot\mathbf{B}_{K'} = [\mathbf{I}_K + \varepsilon_K \;\; \varepsilon_1]$. ε_K and ε_1 are error matrices of size K by K and K by $(K' - K)$, respectively. The estimated coefficients are obtained as

$$\begin{aligned}\tilde{\mathbf{A}}_K &= \begin{bmatrix}\mathbf{I}_K + \varepsilon_K & \varepsilon_1\end{bmatrix}\mathbf{A}_{K'}\\ &\cong \begin{bmatrix}\mathbf{I}_K & \varepsilon_1\end{bmatrix}\begin{bmatrix}\mathbf{A}_K\\ \mathbf{A}_{K'-K}\end{bmatrix}\\ &= \mathbf{A}_K + \varepsilon_1\mathbf{A}_{K'-K}\end{aligned} \tag{39}$$

where $\mathbf{A}_{K'}$ is divided into two column vectors, $\mathbf{A}_K$ and $\mathbf{A}_{K'-K}$, and $\mathbf{A}_{K'-K}$ contains the terms with higher orders than N_{eff}. This means that the estimated coefficients have errors mainly from the higher orders, which is due to the phenomenon of spatial aliasing. This happens when the number of microphones is not enough to sample all the spherical harmonics components present in the sound field.

As described above, the spherical harmonics coefficients can be obtained with each HOA/MOA scheme from the measured pressure values. As in the previous sections, the schemes $5^{(\mathrm{P})}7^{(\mathrm{H})}$, $5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$ and $5^{(\mathrm{P})}7^{(\mathrm{L})}$ are compared in what follows. However, $6^{(\mathrm{P})}$ is used instead of $5^{(\mathrm{P})}$ or $7^{(\mathrm{P})}$ because the number of the terms of $7^{(\mathrm{P})}$ is 64, which requires more microphones than are available, and $5^{(\mathrm{P})}$ shows obviously worse performance than all the chosen MOA schemes. The sound field measured with the microphone array is assumed to be a plane wave with unit magnitude.

C. Coefficient Error

The coefficients $\tilde{A}_{mn}$ obtained in the previous section can be compared with the theoretical values for a plane wave, $A_{mn} = 4\pi i^n Y_n^m(\theta_0, \phi_0)^*$. The coefficient error is defined for each order n as

$$\bar{e}_n^2 = \frac{\sum_{m=-n}^{n}\left|A_{mn} - \tilde{A}_{mn}\right|^2}{\sum_{m=-n}^{n}\left|A_{mn}\right|^2}. \tag{40}$$

If the denominator is zero, this error is not defined.

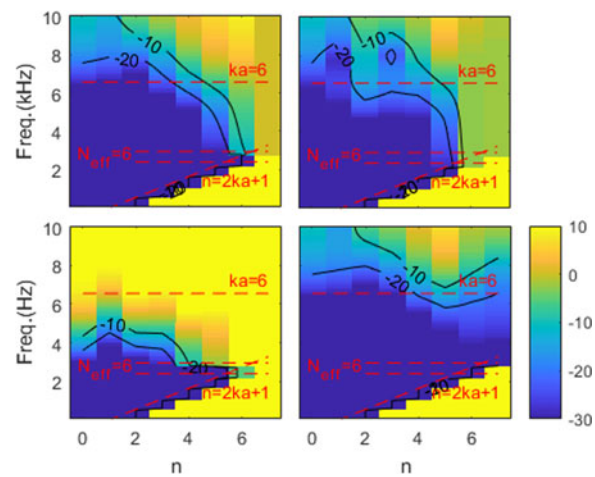

Fig. 6. Coefficient error with $6^{(\mathrm{P})}$ (top left), $5^{(\mathrm{P})}7^{(\mathrm{H})}$ (top right), $5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$ (bottom left), and $5^{(\mathrm{P})}7^{(\mathrm{L})}$ (bottom right).

The coefficient errors for a reference direction (θ_0, ϕ_0) of $(\pi/2, \pi)$ are shown in Fig. 6, with the schemes $6^{(\mathrm{P})}$ (top left), $5^{(\mathrm{P})}7^{(\mathrm{H})}$ (top right), $5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$ (bottom left), and $5^{(\mathrm{P})}7^{(\mathrm{L})}$ (bottom right). The color scale shows the error in a range of 40 dB; contours of -10 dB and -20 dB are also plotted. The order limit of $n = 2ka + 1$, up to which the surface pressure can be accurately obtained for a given frequency, is also shown. The coefficient error on the right hand side of the line is set to be infinity (indicated as yellow) because the values are not defined (the denominator is nearly zero, due to the low contribution of high orders at low frequencies).

At frequencies lower than 2100 Hz, N_{eff} is less than 6, and K for all these schemes is greater than K'. This means that the first condition (34) is satisfied for all the schemes. At frequencies higher than 2700 Hz, N_{eff} is greater than 6, and K for all the schemes is less than K', which belongs to the third condition (37). Spatial aliasing occurs in this case. From 2200 Hz to 2700 Hz, indicated by two dashed lines, $6^{(\mathrm{P})}$ belongs to the second condition, $5^{(\mathrm{P})}7^{(\mathrm{H})}$ to the third condition, and $5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$ and $5^{(\mathrm{P})}7^{(\mathrm{L})}$ to the first condition. The coefficient error of $5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$ at these frequencies is not small because it does not contain all the components of $n = 6$.

As the frequency increases, the error due to spatial aliasing increases. Especially above $ka = 6$ (dashed line around 7 kHz), the effect becomes severe because the number of microphones $(Q = 52)$ is too small to sample spherical harmonics components at higher orders than 6.

$5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$ in Fig. 6 (bottom left) has a larger error than the other schemes at frequencies above 2800 Hz. Some tesseral components reconstructed by $5^{(\mathrm{P})}7^{(\mathrm{H})}3^{(\mathrm{V})}$, e.g., $\tilde{A}_{\pm 6,7}$ and $\tilde{A}_{\pm 5,6}$, are excessively large and the condition number of the matrix

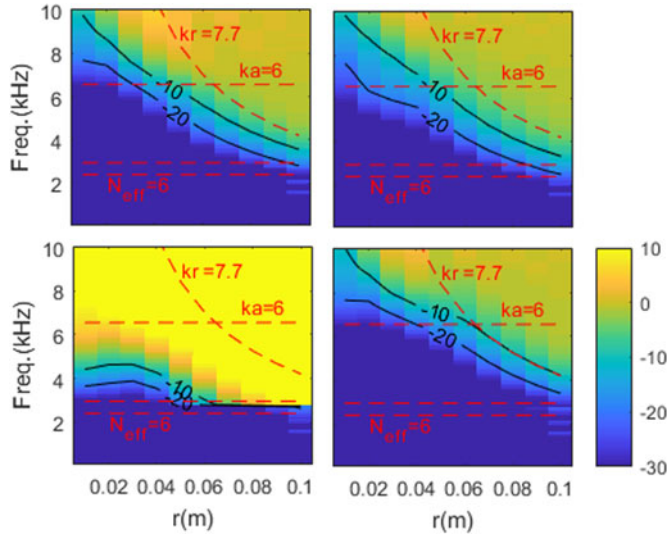

Fig. 7. Reconstruction error with $6^{(P)}$ (top left), $5^{(P)}7^{(H)}$ (top right), $5^{(P)}7^{(H)}3^{(V)}$ (bottom left), and $5^{(P)}7^{(L)}$ (bottom right).

$\mathbf{W}_{KQ}$ is much greater than for the other schemes. This means that with this layout the number of microphones is not sufficient to sample the spherical harmonics components of this scheme. This issue is further discussed in Section V-A.

$5^{(P)}7^{(L)}$ in Fig. 6 (bottom right), on the other hand, shows the smallest errors in overall. For horizontal waves, this scheme works as well as $7^{(P)}$, and the coefficients can be accurately obtained below $ka = 6$. It is worth noting that $7^{(P)}$ is not feasible with the given number of microphones in this case.

D. Reconstruction Error

By using the estimated coefficients $\tilde{A}_{mn}$, the original sound field can be reconstructed. For example, the reconstructed field by $N^{(P)}$ HOA scheme is

$$\tilde{P}_{N^{(P)}}(r,\theta,\phi) = \sum_{n=0}^{N}\sum_{m=-n}^{n} \tilde{A}_{mn}^{\left(N^{(P)}\right)} j_n(kr)\, Y_n^m(\theta,\phi). \quad (41)$$

Then, the reconstruction error can be calculated as

$$\bar{e}_{NP}^2(r) = \frac{\int_0^{\pi}\int_0^{2\pi}\left|P(r,\theta,\phi) - \tilde{P}(r,\theta,\phi)\right|^2 \sin\theta d\phi d\theta}{\int_0^{\pi}\int_0^{2\pi}\left|P(r,\theta,\phi)\right|^2 \sin\theta d\phi d\theta}. \quad (42)$$

In the simulations, the integral is replaced by summation, with an angular spacing of 10 degrees for both the azimuth and the declination angle. This spacing allows describing sound fields up to order 17 ($2N + 2 \leq Q$ [23]). Since the reconstructed fields are limited to lower orders, this spacing was sufficient to capture the changes in pressure with angle.

Note that although the paper focuses on optimizing performance for sound sources in the horizontal plane, performance is not only regarded in the horizontal plane, but in a volume around the origin.

Fig. 7 shows the reconstruction error with different schemes for radii from 0.01 to 0.1 m. Solid lines are contours at -10 dB and -20 dB of the reconstruction error. Since the $7^{(P)}$ scheme is not feasible with the considered microphone array, only the truncation error calculated based on (8) can be shown for comparison. The dashed line indicates where the truncation error has a level of -10 dB for $7^{(P)}$. Equation (8) yields -10 dB error at about $kr = 7.7$. Since the truncation error is included in the reconstruction error, the reconstruction error is always larger.

The reference direction (θ_0, ϕ_0) is $(\pi/2, \pi)$ as in the previous section. The $5^{(P)}7^{(L)}$ scheme has a lower error than the other schemes in general. The -10 dB contour line of the error of this scheme almost agrees with the -10 dB contour of truncation error ($kr = 7.7$) below $ka = 6$ as shown in Fig. 7 (top right).

$5^{(P)}7^{(H)}$ has slightly larger errors than $6^{(P)}$. The $5^{(P)}7^{(H)}3^{(V)}$ scheme is effective only up to 3 kHz, and displays very large errors at higher frequencies due to the excessively large values of some of the tesseral components (Fig. 6).

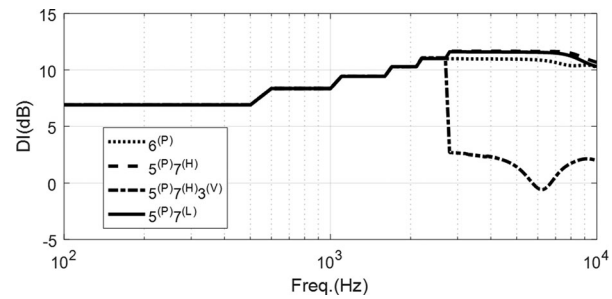

Fig. 8. Directivity index on the horizontal direction with four schemes (in dB).

E. Beamforming Power

The MOA schemes can also be compared in terms of the spatial resolution as in Section III. Instead of the normalized directivity however, the beamforming power is used. For example, the beamforming power of the $N^{(P)}$ scheme can be defined as

$$\tilde{w}_{N^{(P)}}(\theta,\phi) = \sum_{n=0}^{N}\sum_{m=-n}^{n} \frac{Y_n^m(\theta,\phi)\,\tilde{A}_{mn}}{4\pi i^n}. \quad (43)$$

This value is equivalent to the normalized directivity (12) if the theoretical coefficient $4\pi i^n Y_n^m(\theta_0,\phi_0)^*$ is used instead of the estimated $\tilde{A}_{mn}$. Because of the use of the estimated coefficients $\tilde{A}_{mn}$, this beamforming power is also frequency dependent. When spatial aliasing occurs, the error $\varepsilon_1 \mathbf{A}_{K'-K}$ from the high orders in (39) is included in the beamforming power.

Directivity index values with respect to the azimuth angle of each scheme are shown in Fig. 8. Up to 2700 Hz, all the schemes have the same directivity. The DI has a step-like behavior because of the effective order N_{eff}, which depends on the frequency. Above 2700 Hz, $5^{(P)}7^{(L)}$ and $5^{(P)}7^{(H)}$ reach slightly higher directivity than $6^{(P)}$. Above 7 kHz, the DI decreases because of spatial aliasing. The $5^{(P)}7^{(H)}3^{(V)}$ scheme exhibits a large drop in directivity due to its inability to estimate some tesseral components at higher frequencies, which was also reflected in the previously considered measures.

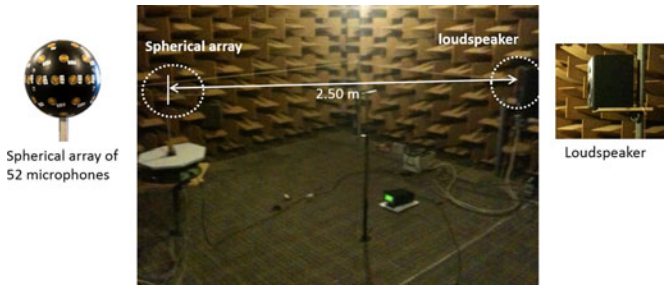

Fig. 9. Experimental setup for validation measurements.

V. Experiments

A. Experimental Setup

For a validation of the proposed method, measurements were conducted in an anechoic chamber at the Technical University of Denmark (4.8 m $\times$ 4.1 m $\times$ 2.9 m, cutoff frequency: about 100 Hz). The experimental setup is shown in Fig. 9. Sound waves are generated by a loudspeaker, and measured with the spherical array of microphones (B&K 4959). The distance to the source was 2.5 m, which is greater than the wavelength above around 140 Hz. The sound waves are assumed to be plane waves. The loudspeaker and the microphone array were set to be at the same height.

Transfer functions between the input signal fed into the loudspeaker and the measured signals at the microphones were estimated using white noise. The sampling frequency was 32768 Hz. The transfer function values were then used for the estimation of the expansion coefficients.

B. Experimental Results

In order to calculate the coefficient error in (40), true values for the coefficient A_{mn} have to be determined. Since the sound waves were assumed to be plane waves, A_{mn} can be regarded as $C \cdot 4\pi i^n Y_n^m(\theta_0, \phi_0)^*$, where C is a constant. This constant was determined by the estimated coefficient from the measurement $\tilde{A}_{00}$:

$$C \cdot 4\pi i^0 Y_0^0(\theta_0, \phi_0)^* = C \cdot \sqrt{4\pi} = \tilde{A}_{00}. \tag{44}$$

The regularization parameter λ in (30) was empirically set to 0.01 to reduce the effect of experimental noise.

Fig. 10 shows the coefficient error with $6^{(P)}$ (top left), $5^{(P)}7^{(H)}$ (top right), $5^{(P)}7^{(H)}3^{(V)}$ (bottom left), and $5^{(P)}7^{(L)}$ (bottom right) now based on measured data. The error for the zero-th order $(n = 0)$ component is always zero because it was used to match the overall levels of the coefficients, as described by (44). Compared with the simulation data in Fig. 6, error values are in general greater than for the simulations. In particular, the error at low frequencies is larger. This is most likely due to the effect of measurement noise, as discussed in Section VI-C. Despite this, it can be observed that $5^{(P)}7^{(L)}$ has the smallest errors overall among the schemes as in the simulation results (Fig. 6).

Fig. 11 shows the reconstruction error with the same schemes. As in Fig. 7, the errors with $6^{(P)}$ and $5^{(P)}7^{(H)}$ do not show large differences, but $5^{(P)}7^{(L)}$ scheme again results in the smallest errors overall. The -10 dB contour line for $5^{(P)}7^{(L)}$ is closer to the line for $kr = 7.7$ (the reference contour for a truncation error of -10 dB with $7^{(P)}$) than for the other schemes.

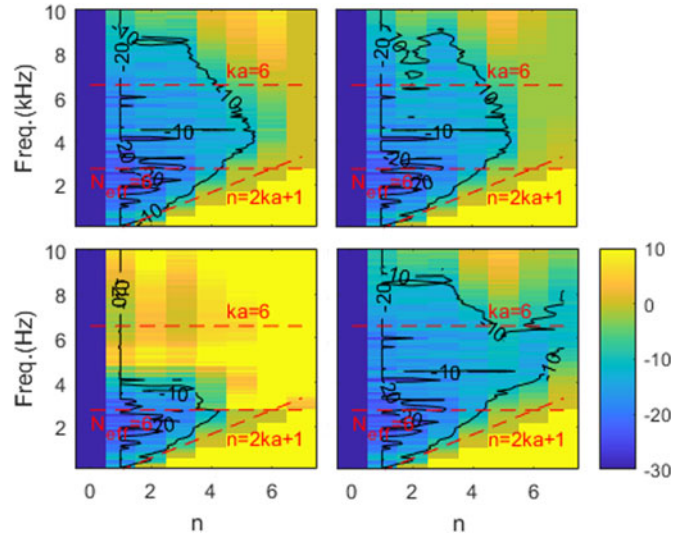

Fig. 10. Coefficient error with $6^{(P)}$ (top left), $5^{(P)}7^{(H)}$ (top right), $5^{(P)}7^{(H)}3^{(V)}$ (bottom left), and $5^{(P)}7^{(L)}$ (bottom right) (experimental data).

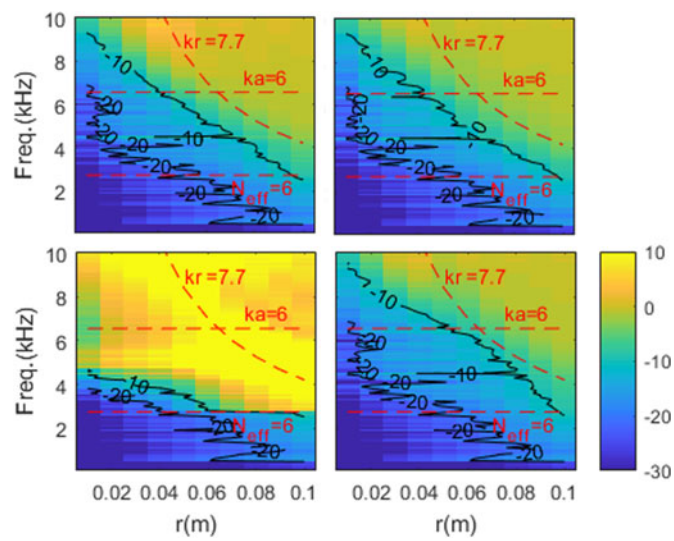

Fig. 11. Reconstruction error with $6^{(P)}$ (top left), $5^{(P)}7^{(H)}$ (top right), $5^{(P)}7^{(H)}3^{(V)}$ (bottom left), and $5^{(P)}7^{(L)}$ (bottom right) (experimental data).

Fig. 12 shows the directivity index based on the measurements. The step-like behavior is not apparent here, and the measured directivity at low frequencies is lower, but the results are otherwise similar to those obtained from the simulation (Fig. 8).

VI. Discussions

A. Sound Sources Out of the Horizontal Plane

In previous sections, only sound sources in the horizontal plane were regarded as the reference sound fields because the proposed schemes have advantages in those cases. However, it is worth investigating the performance for sound sources out of the horizontal plane.

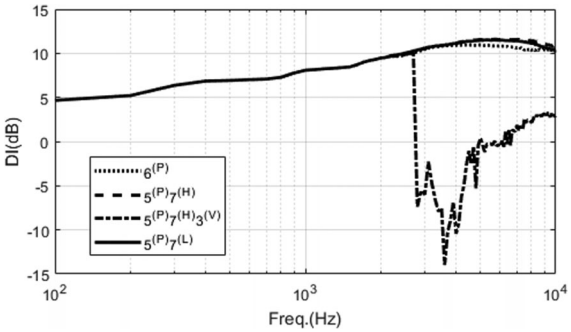

Fig. 12. Directivity index on the horizontal direction (experimental data).

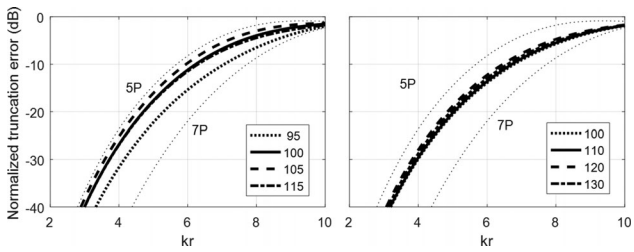

Fig. 13. Normalized truncation error of a plane wave of several declination angles (left: $5^{(P)}7^{(L)}$, right: $5^{(P)}7^{(H)}3^{(V)}$).

Fig. 13 shows the truncation error with four declination angles (95, 100, 105, 115 degrees) when the scheme $5^{(P)}7^{(L)}$ (left) and $5^{(P)}7^{(H)}3^{(V)}$ (right) are used. The truncation error for these source angles takes values between those of $5^{(P)}$ and $7^{(P)}$, as expected. For the $5^{(P)}7^{(L)}$ scheme it was previously shown that for a horizontal source (declination angle of 90 degrees) the truncation error corresponds to that of the periphonic scheme $7^{(P)}$. As the source angle increases, the truncation error also gradually increases up to 105 degrees, and decreases again for higher angles. In general, the truncation error fluctuates with the declination angle depending on the dominant spherical harmonics components. That is, the value of each component of spherical harmonics function differs with respect to the declination angle. If a scheme does not contain dominant components for the angle of the sources, this will result in a larger truncation error for those particular directions.

The truncation error of $5^{(P)}7^{(H)}3^{(V)}$, however, does not appear to be strongly dependent on the declination angle, as shown in Fig. 13 (right). Hence, this scheme may be preferred if a relatively uniform performance along the vertical direction is required.

The scheme $5^{(P)}7^{(H)}$ has only the sectoral harmonics in the 6th and 7th order, which monotonically decreases as the declination angle increases (from $\pi/2$ to π). Thus, this scheme has the smallest error for $\pi/2$, and the largest error for π although it is not shown.

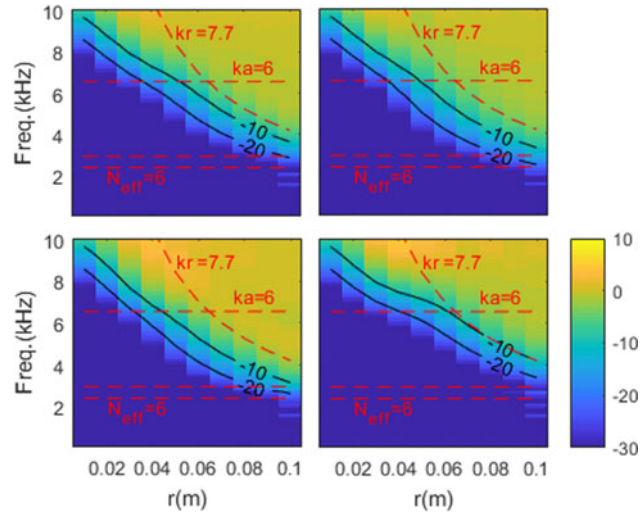

Fig. 14. Reconstruction error with nearly uniform sampling of 54 microphones with $6^{(P)}$ (top left), $5^{(P)}7^{(H)}$ (top right), $5^{(P)}7^{(H)}3^{(V)}$ (bottom left), and $5^{(P)}7^{(L)}$ (bottom right).

B. Microphone Array Layout and Coefficient Error

As shown in Fig. 7 (bottom left), the coefficient error with $5^{(P)}7^{(H)}3^{(V)}$ scheme was larger than with the other schemes because of inadequately captured components as mentioned in Section IV-B. The other considered schemes did not exhibit this behavior. This issue is related to the number and the layout of microphones on the considered spherical array.

According to [7], the required number of microphones for Nth order HOA is $4(N+1)^2$ and $2(N+1)^2$, for equi-angle sampling and the Gaussian sampling, respectively. For nearly uniform sampling, $(N+1)^2$ to $1.5(N+1)^2$ microphones are required. The layout used in this study is non-uniform and is different from these samplings. It is difficult to say exactly how many microphones are needed with such a layout for each scheme. Based on the simulation and measurement results however, it appears that the 52 microphones were only insufficient for the $5^{(P)}7^{(H)}3^{(V)}$ scheme.

To briefly investigate the effects of choosing a different layout, nearly uniform layouts for different numbers of points (as provided in [33]) were also considered. If the nearly uniform layout with the same number of microphones as the previously considered non-uniform array is used, the results do not show large differences (not shown). If the number of microphones is increased to 54 in the nearly uniform layout, the coefficient error with $5^{(P)}7^{(H)}3^{(V)}$ is considerably reduced. Fig. 14 shows the reconstruction error with 54 microphones for the nearly uniform layout. In this case, the reconstruction error for $5^{(P)}7^{(H)}3^{(V)}$ is as low as for $5^{(P)}7^{(H)}$. Using only 51 microphones in the nearly uniform layout results in large coefficient errors also for the $5^{(P)}7^{(L)}$ scheme. Similarly, the error with $6^{(P)}$ significantly increases when only 50 microphones are considered (not shown). For the

$$WNG\left(\theta_0,\phi_0\right)=10\log_{10}\left[\frac{\left|\mathbf{y}_{1K}\left(\theta_0,\phi_0\right)\mathbf{C}_K^{-1}\mathbf{B}_K^{-1}\mathbf{W}_{KQ}\cdot\mathbf{P}\right|^2}{\left\{\mathbf{y}_{1K}\left(\theta_0,\phi_0\right)\mathbf{C}_K^{-1}\mathbf{B}_K^{-1}\mathbf{W}_{KQ}\right\}\left\{\mathbf{y}_{1K}\left(\theta_0,\phi_0\right)\mathbf{C}_K^{-1}\mathbf{B}_K^{-1}\mathbf{W}_{KQ}\right\}^H}\right], \tag{45}$$

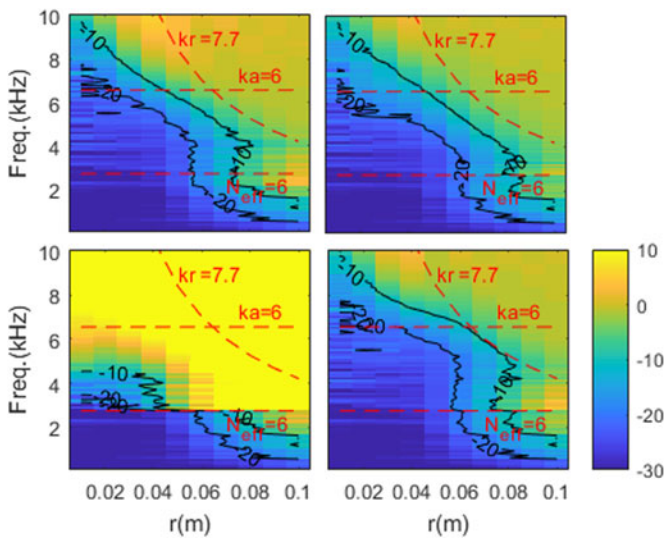

Fig. 15. Reconstruction error with noise, $6^{(P)}$ (top left), $5^{(P)}7^{(H)}$ (top right), $5^{(P)}7^{(H)}3^{(V)}$ (bottom left), and $5^{(P)}7^{(L)}$ (bottom right).

$5^{(P)}7^{(H)}$ scheme, the number of microphones can be reduced to 40 (equal to the number of components) without considerably increasing the coefficient errors (not shown). This indicates that about $1.1(N+1)^2$ microphones are required for these schemes when using nearly uniform layouts, although the exact number varies. The matrix inversion with PHV schemes appears to be more sensitive than with other schemes, and PH schemes tend to be more robust. The minimum required number of microphones for each scheme is shown in Table I.

C. Effect of Noise

No noise was assumed in the simulations, but measurement noise was present in the experimental validation, and degraded the performance to some extent. To investigate the effect of noise, the self-noise of a commercial microphone (B&K 4959) and a measurement noise of 20 dB SNR were added to the simulated sound pressure. The self-noise had a level of about 22 dB SPL at 100 Hz and decreased with frequency. Fig. 15 shows the reconstruction error from simulations considering the additive noise noises. Due to the effect of the noise, the reconstruction error increases for all schemes. At low frequencies, the effect was larger than at high frequencies, mirroring the measured results. This is a common issue in spherical array processing, due to the radial function $b_n(ka)$ being very small at low frequencies for high orders, which results in a high sensitivity to low-frequency noise.

This effect can be reduced by regularization to some extent as shown in Fig. 16, where the regularization parameter is set to be 0.01 in (29). These figures show similar behavior as the measured results (Fig. 11). However, in the experiments, the resulting error values were somewhat larger in general. Imperfection of the anechoic chamber, the plane wave assumption, and scattering from the experimental instruments may have also affected the measurements.

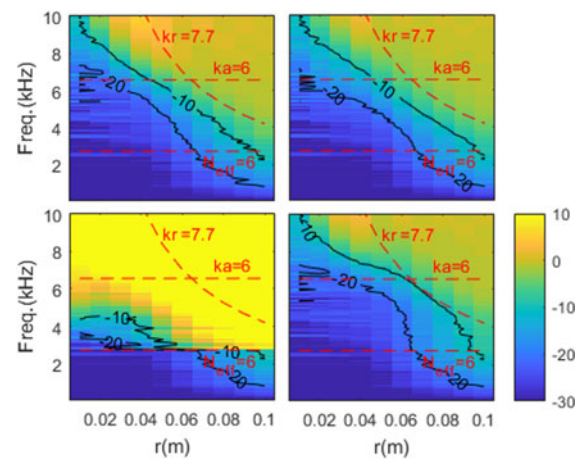

Fig. 16. Reconstruction error with noise, but regularized ($\lambda = 0.01$), $6^{(P)}$ (top left), $5^{(P)}7^{(H)}$ (top right), $5^{(P)}7^{(H)}3^{(V)}$ (bottom left), and $5^{(P)}7^{(L)}$ (bottom right).

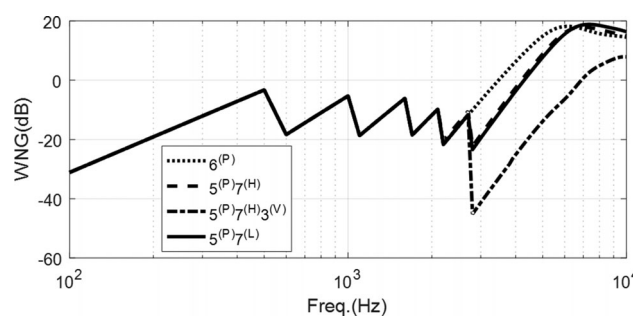

Fig. 17. White noise gain with the schemes in dB.

The robustness of an array processing scheme to noise can be shown with a performance index, the white noise gain (WNG) [32], which is given by eq. (45) as shown at the bottom of the previous page, where $\mathbf{C}_K = \mathrm{diag}\,[4\pi i^n]_K$ and $\mathbf{y}_{1K}(\theta_0, \phi_0)$ is a row vector of 1 by K that is composed of the spherical harmonics functions.

Fig. 17 shows WNG for each scheme, computed for the non-uniform 52-channel microphone array. Again, up to about 2700 Hz, all schemes behave identically, and the differences are apparent at higher frequencies. $6^{(P)}$ has the largest WNG, and thus is the most robust, at least up to about 6 kHz. $5^{(P)}7^{(H)}3^{(V)}$ comes out as the least robust with this particular configuration. The peak WNG for $5^{(P)}7^{(H)}$ and $5^{(P)}7^{(L)}$ are shifted towards higher frequencies due to the inclusion of 7-th order terms, which comes at a cost of somewhat reduced robustness at mid-high frequencies. However, it can be seen that the overall behavior of the latter two MOA schemes in terms of WNG is similar to what would be expected for $7^{(P)}$ (HOA).

VII. Conclusion

This study proposed a MOA scheme termed PL, which selects a subset of coefficients from the spherical harmonics expansion in a lattice pattern at orders higher than the periphonic order N_P. A particular feature of this scheme is that it preserves all information available for sources in the horizontal plane. This

scheme was compared with other previously introduced MOA schemes (PH and PHV), as well as HOA. The comparison was performed considering the effect of truncation, as well as that of spatial encoding with a spherical microphone array through simulations as well as measurements. The results show that the proposed $N_{\mathrm{P}}^{(\mathrm{P})} N_{\mathrm{L}}^{(\mathrm{L})}$ scheme is equivalent to $N_{\mathrm{L}}^{(\mathrm{P})}$ for sound waves propagating in horizontal directions, and performs as well as $N_{\mathrm{P}}^{(\mathrm{P})}$ for those propagating in elevated directions in the worst case. Analysis on the layout and the required number of microphones and the effect of noise show that this MOA scheme does not have any serious disadvantages compared with HOA schemes.

In particular, this scheme may be useful to improve performance for horizontal sources with a relatively fewer number of microphones than would be required for increasing the periphonic (HOA) order. As demonstrated by the examples schemes investigated in this study, $5^{(\mathrm{P})}7^{(\mathrm{L})}$ performs as well as $7^{(\mathrm{P})}$ with only 52 microphones, even though the number of microphones is insufficient for full periphonic order of 7.

The present MOA scheme could be used in various applications: sound field microphones for virtual reality or augmented reality (VR/AR) audio recording, sound field decomposition in room acoustics, reconstruction of intensity vector for noise source localization, and so on.

Appendix
The Truncation Error of PHV Schemes

This section proves (9). The truncation error is defined as

$$\bar{e}^2_{N_P{}^{(\mathrm{P})} N_H{}^{(\mathrm{H})} N_V{}^{(\mathrm{V})}}(kr) = \frac{\int \left| P(r,\theta,\phi) - \widehat{P}_{N_P{}^{(\mathrm{P})} N_H{}^{(\mathrm{H})} N_V{}^{(\mathrm{V})}}(r,\theta,\phi) \right|^2 d\Omega}{\int |P(r,\theta,\phi)|^2 d\Omega}, \tag{A1}$$

where $P(r,\theta,\phi)$ is a plane wave. If the magnitude of the plane wave is 1, the denominator is

$$\int |P(r,\theta,\phi)|^2 d\Omega = \int \left| e^{i\mathbf{k}\cdot\mathbf{r}} \right|^2 d\Omega = 4\pi. \tag{A2}$$

The nominator is expressed as a spherical harmonics expansion, and the orthogonality of the spherical harmonics is applied:

$$\begin{aligned}
&\int \left| P(r,\theta,\phi) - \widehat{P}_{N_P^{(\mathrm{P})} N_H^{(\mathrm{H})} N_V^{(\mathrm{V})}}(r,\theta,\phi) \right|^2 d\Omega \\
&= \int \left| \begin{array}{l} \sum\limits_{n=N_P}^{\infty} \sum\limits_{m=-n}^{n} A_{mn} j_n(kr) Y_n^m(\theta,\phi) \\ - \sum\limits_{n=N_H+1}^{N_P} \sum\limits_{m=-n+N_V+1}^{n-N_V-1} \left[A_{mn} j_n(kr) Y_n^m(\theta,\phi) \right] \end{array} \right|^2 d\Omega \\
&= \sum_{n=N_P}^{\infty} {j_n}^2(kr) \sum_{m=-n}^{n} |A_{mn}|^2 + \sum_{n=N_H+1}^{N_P} {j_n}^2(kr) \\
&\quad \times \sum_{m=-n+N_V+1}^{n-N_V-1} |A_{mn}|^2.
\end{aligned} \tag{A3}$$

The coefficient A_{mn} is replaced by $4\pi i^n Y_n^m(\theta_0,\phi_0)^*$, and the addition theorem is applied [34]:

$$\begin{aligned}
&\int \left| P(r,\theta,\phi) - \widehat{P}_{N_P^{(\mathrm{P})} N_H^{(\mathrm{H})} N_V^{(\mathrm{V})}}(r,\theta,\phi) \right|^2 d\Omega \\
&= (4\pi)^2 \sum_{n=N_P}^{\infty} {j_n}^2(kr) \sum_{m=-n}^{n} |Y_n^m(\theta_0,\phi_0)^*|^2 \\
&\quad + (4\pi)^2 \sum_{n=N_H+1}^{N_P} {j_n}^2(kr) \sum_{m=-n+N_V+1}^{n-N_V-1} |Y_n^m(\theta_0,\phi_0)^*|^2 \\
&= (4\pi)^2 \sum_{n=N_P}^{\infty} {j_n}^2(kr) \frac{2n+1}{4\pi} \\
&\quad + (4\pi)^2 \sum_{n=N_H+1}^{N_P} {j_n}^2(kr) \sum_{m=-n+N_V+1}^{n-N_V-1} |Y_n^m(\theta_0,\phi_0)^*|^2.
\end{aligned} \tag{A4}$$

Equation (9) is obtained by dividing (A4) by (A2). Equation (10) can be derived in the similar way.

References

[1] J.-H. Chang and M. Marschall, "A mixed-order ambisonic scheme to improve performance for sound sources on the horizontal plane," presented at the 5th Joint Meeting Acoustical Society Japan Acoustical Society America, Honolulu, HI, USA, Nov. 28–Dec. 2, 2016.

[2] E. G. Williams, *Fourier Acoustics*. Cambridge, U.K.: Academic, 1999, pp. 183–186.

[3] F. Jacobsen, G. Moreno-Pescador, E. Fernandez-Grande, and J. Hald, "Near field acoustic holography with microphones on a rigid sphere," *J. Acoust. Soc. Amer.*, vol. 129, no. 6, pp. 3461–3464, Jun. 2011.

[4] E. G. Williams and K. Takashima, "Vector intensity reconstructions in a volume surrounding a rigid spherical microphone array," *J. Acoust. Soc. Amer.*, vol. 127, no. 2, pp. 773–783, Feb. 2010.

[5] E. Fernandez-Grande, "Sound field reconstruction using a spherical microphone," *J. Acoust. Soc. Amer.*, vol. 139, no. 3, pp. 1168–1178, Mar. 2016.

[6] M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *J. Acoust. Soc. Amer.*, vol. 118, no. 5, pp. 3094–3103, Nov. 2005.

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

[8] Z. Li and R. Duraiswami, "Flexible and optimal design of spherical microphone arrays for beamforming," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 2, pp. 702–714, Feb. 2007.

[9] I. Balmages and B. Rafaely, "Open-sphere designs for spherical microphone array," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 2, pp. 727–732, Feb. 2007.

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

[11] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the sound field," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2002, vol. 2, pp. 1781–1784.

[12] J. Daniel and S. Moreau, "Further study of sound field coding with higher order ambisonics," in *Proc. 116th Conv. Audio Eng. Soc.*, Berlin, Germany, May 8–11, 2004, Paper 6017.

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

[14] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc.*, vol. 53, pp. 1004–1025, Nov. 2005.

[15] B. Rafaely and D. Khaykin, "Optimal model-based beamforming and independent steering for spherical loudspeaker arrays," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2234–2238, Sep. 2011.

[16] F. Zotter, M. Zaunschirm, M. Frank, and M. Kronlachner, "A beamformer to play with wall reflections: The icosahedral loudspeaker," *Comput. Music J.*, vol. 41, no. 3, pp. 50–68, 2017.
[17] P. B. Fellgett, "Ambisonic reproduction of directionality in surround sound systems," *Nature*, vol. 252, pp. 534–538, Dec. 1974.
[18] B. Rafaely, *Fundamentals of Spherical Array Processing*. Berlin, Germany: Springer-Verlag, 2015, pp. 38–42.
[19] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA, USA: MIT Press, 1997.
[20] M. A. Gerzon, "Periphony: With-height sound reproduction," *J. Audio Eng. Soc.*, vol. 21, pp. 2–10, 1973.
[21] J. Daniel, "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia," Doctoral dissertation, Univ. Paris, Paris, France, Jul. 2001.
[22] C. Travis, "A new mixed-order scheme for ambisonic signals," in *Proc. Ambisonics Symp.*, Jun. 2009, pp. 1–6.
[23] S. Favrot, M. Marschall., J. Kasbach, J. Buchholz, and T. Weller, "Mixed-order ambisonics recording and playback for improving horizontal directionality," in *Proc. 131st Conv. Audio Eng. Soc.*, New York, NY, USA, 2011, Paper 8528.
[24] J. Trevino, S. Koyama, S. Sakamoto, and Y. Suzuki, "Mixed-order ambisonics encoding of cylindrical microphone array signals," *Acoust. Sci. Technol.*, vol. 35, no. 3, pp. 174–177, 2014.
[25] M. Marschall, S. Favrot, and J. Buchholtz, "Robustness of a mixed-order ambisonics microphone array for sound field reproduction," in *Proc. 132nd Conv. Audio Eng. Soc.*, 2012, Paper 8645.
[26] M. Marschall and J.-H. Chang, "Sound field reconstruction performance of a mixed-order ambisonics microphone array," *Proc. Meetings Acoust.*, vol. 19, 2013, Art. no. 055007.
[27] T. Weller, J. Buchholz, and C. Oreinos, "Frequency dependent regularization of a mixed-order ambisoncis encoding system using psychoacoustically motivated metrics," in *Proc. 55th Int. Conf. Audio Eng. Soc.*, Helsinki, Finland, Aug. 2014, Paper 3-3.
[28] H. Pomberger and F. Zotter, "An ambisonics format for flexible playback layouts," in *Proc. Ambisonics Symp.*, Graz, Austria, Jun. 2009, pp. 1–8.
[29] F. Zotter, H. Pomberger, and M. Noisternig, "Energy-preserving ambisonic decoding," *Acta Acustica United Acustica*, vol. 98, pp. 37–47, Jan./Feb. 2012.
[30] C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi, "AMBIX-A suggested ambisonics format," in *Proc. Ambisonics Symp.*, Lexington, KY, USA, Jun. 2011, pp. 1–11.
[31] F. Zotter and M. Frank, "All-round ambisonic panning and decoding," *J. Audio Eng. Soc.*, vol. 60, no. 10, pp. 807–820, 2012.
[32] M. Brandstein and D. Ward, Eds., *Microphone Arrays*. New York, NY, USA: Springer-Verlag, 2001.
[33] N. J. A. Sloane, "Spherical codes: Nice arrangements of points on a sphere in various dimensions." 1996. [Online]. Available: http://neilsloane.com/packings/
[34] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, 2nd ed. New York, NY, USA: Springer, 1998, p. 27.

Jiho Chang (M'17) was born in Seoul, South Korea, in 1980. He received the B.S. degree from Seoul National University, Seoul, South Korea, in 2003, and the Ph.D. degree from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2011, both in mechanical engineering.

From 2011 to 2014, he was a Postdoctoral Researcher with the Electrical Engineering Department, Technical University of Denmark. From 2014 to 2016, he was a Senior Engineer with the Samsung Electronics. Since 2016, he has been a Senior Researcher with the Korea Research Institute of Standards and Science, Daejeon, South Korea. His research interests include acoustical array signal processing for loudspeaker arrays and microphone arrays, audio signal processing, spatial audio, and spatial hearing.

Dr. Chang is a member of the Acoustical Society of America, the Korean Society for Noise and Vibration Engineering, and the Korean Society of Mechanical Engineers.

Márton Marschall is a native of Budapest, Hungary. He received the M.Sc. degree in electrical engineering from Budapest University of Technology and Economics, Budapest, Hungary, in 2006, and the M.Sc. degree in engineering acoustics and the Ph.D. degree in electrical engineering from the Technical University of Denmark, Kongens Lyngby, Denmark, in 2008 and 2014, respectively.

Since 2014, he has been a Postdoctoral Researcher with the Centre for Applied Hearing Research, Technical University of Denmark. His research interests include spatial audio recording and reproduction, virtual environments, spatial hearing, and auditory modeling.