




Extended Vector-Based EB-ESPRIT Method

Byeongho Jo , *Student Member, IEEE*, Franz Zotter , and Jung-Woo Choi , *Member, IEEE*

Abstract—The estimation of direction of arrivals (DoAs) from spherical microphone array data is one of the key issues in extracting source information from all-around audio recordings. One such technique is the eigenbeam estimation of signal parameters via the rotational invariance technique (EB-ESPRIT), which separates the signal subspace related to the stationary sound field and then directly estimates DoAs of multiple sound sources. EB-ESPRIT has been evolved in many different ways by involving different types of recurrence relations of spherical harmonics, all of which are able to identify DoAs of a limited number of sources that are noticeably smaller than the number of finite-order spherical harmonic coefficients recorded. In this work, we report that it is possible to go beyond the known limits of detectable sources. The proposed formula is also based on conventional recurrence relations and probably permits to reach the ultimate limit by additional constraints of the signal parameters that can better exploit the highest-order coefficients. Monte-Carlo simulations conducted with various source positions and signal-to-noise ratios (SNRs) reveal that the proposed technique can detect more sources with insignificant loss in estimation performance and robustness.

Index Terms—Direction-of-arrival estimation, EB-ESPRIT, recurrence relations, spherical harmonics.

NOMENCLATURE

$(\cdot)^*$	Complex conjugate.
$(\cdot)^T$	Transpose of a matrix.
$(\cdot)^H$	Conjugate transpose of a matrix.
$\ \cdot\ _F$	Frobenius norm of a matrix.
$Z_{\text{diag}}(\cdot)$	Off-diagonal part of a matrix.
k	Wavenumber (radian / m).
r	Radius of spherical microphone array.
ϑ_q, φ_q	Zenith and azimuth of q^{th} plane wave.
$Y_n^m(\vartheta_q, \varphi_q)$	Complex spherical harmonics.
$b_n(kr)$	Far-field mode-strength.
Q	Number of plane waves.
Q_{max}	Maximum number of detectable sources.
N	Maximum order of spherical harmonics.
L	Total Number of spherical harmonics up to the N^{th} order, $= (N + 1)^2$.

L_1	Number of spherical harmonics up to the $(N - 1)^{\text{th}}$ order, $= N^2$.
s_q	Complex amplitude of q^{th} plane wave.
s	Complex amplitude of Q plane waves.
\mathbf{n}	Measurement noises.
\mathbf{a}_q	Spherical harmonic coeffs. q^{th} plane wave.
$\tilde{\mathbf{a}}$	Observed spherical harmonic coefficients.
$\mathbf{y}(\Omega_q)$	Spherical harmonic manifold of q^{th} plane wave.
\mathbf{Y}	Spherical harmonics manifold matrix.
\mathbf{R}	Covariance matrix of spherical harmonic coefficients.
\mathbf{R}_s	Source signal covariance matrix.
\mathbf{R}_{n_s}	Noise covariance matrix.
\mathbf{U}_s	Signal subspace eigenvectors.
\mathbf{U}_T	Block diagonal matrix of \mathbf{U}_s .
\mathbf{M}	Binary mask matrix.
$\mathbf{M}_{(r,s)}$	Order-reducing binary matrix shifting harmonic indices by (r, s) .
\mathbf{M}_T	Block diagonal matrix of \mathbf{M} .
$\mathbf{W}_{(r,s)}, \mathbf{V}_{(r,s)}$	Diagonal matrices of recurrence coefficients.
$\mathbf{D}_{\{xy^*, xy, z\}}$	Coefficients of recurrence relations for expressing directional waves.
\mathbf{D}_T	Mixture of recurrence coefficients.
$\theta_{q, \{xy^*, xy, z\}}$	Directional parameters of q^{th} plane wave.
$\Theta_{\{xy^*, xy, z\}}$	Diagonal parameter matrices.
$\Psi_{\{xy^*, xy, z\}}$	Transformed parameter matrices.
\mathbf{T}	Transformation matrix of directional parameters.

I. INTRODUCTION

ESTIMATION of direction of arrivals (DoAs) is an important topic in sound source localization (SSL) problems. For example, DoAs of noise sources are important information in noise control problems. In the speech recognition task, locations of speakers are essential prerequisites for the source separation and noise suppression [1], [2]. For spatial audio coding and parameterization, DoAs are primary cues for separating the directional audio component from ambient signals, which enable the compression of multichannel data or optimal mixing to playback signals for various loudspeaker layouts [3]. Recently, there have been attempts to identify wall locations using echoes, and the DoA estimation can provide directions of image sources producing echoes inside a room [4]–[6].

Techniques for the DoA estimation can be largely categorized as parametric and non-parametric techniques [7]. The non-parametric technique generates a map of steered response power or detection probability, and peaks of high power are

Manuscript received September 2, 2019; revised March 17, 2020 and April 23, 2020; accepted May 9, 2020. Date of publication May 20, 2020; date of current version June 5, 2020. This work was supported in part by the National Research Foundation of Korea grant funded by the Korea government (MSIT) (1711091575) and in part by the BK21 Plus program through the National Research Foundation funded by the Ministry of Education of Korea. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jun Du. (*Corresponding author: Jung-Woo Choi.*)

Byeongho Jo and Jung-Woo Choi are with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea (e-mail: byeongho@kaist.ac.kr; jwoo@kaist.ac.kr).

Franz Zotter is with the Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, 8010 Graz, Austria (e-mail: zotter@iem.at).

Digital Object Identifier 10.1109/TASLP.2020.2996090

searched and chosen as DoAs. Well known techniques such as the delay and sum (DAS) beamformer [8], minimum variance distortionless response (MVDR) [9], and multiple signal classification (MUSIC) [10] techniques fall into this category. On the other hand, the parametric technique tries to directly identify DoAs in form of directional parameters, which negates the needs for time-intensive grid search. Root-MUSIC [11] and estimation of signal parameters via the rotational invariance technique (ESPRIT) [12] are popular ones that directly calculate such parameters by finding the roots of a polynomial equation defined from the covariance matrix or by setting up a relation between two subarrays shifted in space. The other popular parametric technique incorporates the intensity-based parameter, which directly provides DoAs from the mean active intensity approximated from the spherical harmonic recording (pseudo-intensity vector) [13], [14]. The pseudo-intensity based technique provides only one vector component per each time-frequency (TF) bin, so it can be problematic for multiple sources. For this reason, it is usually combined with the direct-path dominance (DPD) test [15], [16]. Intensity vectors are estimated in the time-frequency bins of strong direct field dominance and clustered to find the directions of multiple sound sources. Some of these techniques also utilize the subspace filtering to reduce the influence of noises but cannot handle multiple sources as much as the subspace-based technique when steady sources or noises are activated simultaneously or the signals involved are likely to cause concurrency within bins of the targeted short-term spectral analysis [17]–[19].

Although the attempt to detect the largest-possible number of multiple simultaneous sources could increase the computational cost and could therefore counteract real-time implementation on low-cost integrated hardware, there are many applications in which the benefits would outweigh the additional effort. For instance, the analysis of room impulse responses with high-order spherical microphone arrays, e.g. to analyze reverberation [20], [21], can benefit from the largest-possible number of detectable sources, in particular in time segments in which the echo density causes multiple, interfering image sources in the regarded time frame. With only slightly increased effort, the parametric higher-order directional audio coding and room impulse response rendering techniques (HO-DirAC [22], HO-SIRR [23]) efficiently extract multiple pseudo-intensity vectors. However, to cope with multiple sources, intensity vectors are extracted in directional sectors, therefore, sound field parameters are only obtained within those sectors.

This work deals with the parametric DoA estimation problem when spherical microphone array recordings are available. For spherical array recordings, various parametric estimation techniques have been developed, mainly stemming from the Eigenbeam ESPRIT (EB-ESPRIT) [24], [25] method. The EB-ESPRIT techniques use a spherical Fourier transform to handle spherical array recordings. When plane waves from different DoAs are analyzed, the spherical Fourier transform yields spherical harmonic (SH) coefficient signals that are first analyzed in terms of their signal covariance matrix. Unlike the other subspace-based technique such as EB-MUSIC [26], EB-ESPRIT first determines the signal plus noise subspace. The

signal subspace, denotes the vector space spanned by the SH signal covariance matrix of sound fields from multiple stably localized sources, whereas the noise subspace represents the subspace occupied only by noises or diffuse sounds [27]. Estimation of the signal subspace is an essential step in EB-ESPRIT that isolates the subspace in which stable source signal extraction and DoA analysis is possible. Once the eigenvectors that span the signal subspace are obtained, then one can pose several constraints to identify directions of sparsely located sound sources. As EB-ESPRIT algorithms deal with the signal subspace of SH coefficients, their key goal is relating the directional parameters to the SH coefficient signals. This is done by utilizing the recurrence relations of the spherical harmonic functions that are implicitly involved. These recurrences manage to factor out directional parameters as linear, diagonal factors of the signal subspace. The original EB-ESPRIT technique [24] uses a single recurrence relation that expresses the directional parameter as a multiplication of tangent and exponential functions (Table I).

Despite of its advantage for multisource situations compared to the intensity vector-based techniques, there are some practical issues involved with the EB-ESPRIT technique. The first has to do with the limited number of detectable sources. Owing to the given number of microphones, SH coefficients can only be measured up to a finite order N . Depending on the type of recurrence relations used for the DoA estimation, the number of detectable sources can vary (Table I). The original EB-ESPRIT is able to detect $\lfloor \frac{N^2}{2} \rfloor$ sources, which is outperformed by more recent variants of the EB-ESPRIT technique [28]–[32]. To widen the benefit and range of applications of the original EB-ESPRIT, research has been trying to reach the largest-possible number of detectable sources, as one of its foremost goals.

Other issues, such as robustness against the measurement noises, singularity or ambiguity problems for certain DoAs [33], are also troublesome in conventional EB-ESPRIT techniques.

To circumvent these issues, various modifications have been made. For example, the sine-based EB-ESPRIT [28], [29] uses different recurrence relations that take a sine function as the directional parameter for the zenith angle. By using the sine function, this method attempts to avoid the singularity issue arising from the original EB-ESPRIT's tangent function that diverges at the zenith angle $\vartheta \approx 90^\circ$. The use of the new recurrences functions also brought a dramatic increase in the number of detectable sources to $N^2 + \lfloor \frac{N}{2} \rfloor$. The sine function, however, is a slowly varying function close to the horizon $\vartheta = 90^\circ$ that inevitably degrades the DoA estimation accuracy there. In addition, the sine function is symmetric with regard to the horizon ($\sin \vartheta = \sin(180^\circ - \vartheta)$), so extra post-processing is required to resolve this up-down ambiguity issue.

Another method called two-step spherical harmonics ESPRIT (TS-SHESPRIT) separately estimates zenith and azimuth angles using decoupled recurrence relations to avoid the ambiguity issue [34]. The decoupled estimation process, however, requires extra pair-matching of zenith and azimuth angles. Moreover, the cosine function used to estimate the zenith angles also changes slowly near zenith $\vartheta = 0$, and nadir 180° , introducing a degraded estimation accuracy there. The most significant disadvantage

TABLE I
COMPARISON OF EB-ESPRIT TECHNIQUES

Algorithm	Directional parameters		Number of detectable sources	Free from ambiguity	Estimation performance	Disadvantage
	Zenith	Azimuth				
EB-ESPRIT [24]	$\tan \vartheta e^{i\varphi}$		$\lfloor N^2/2 \rfloor$	\times	Moderate	Singularity near $\vartheta = 90^\circ$
TS-SHESPRIT [34]	$\cos \vartheta$	$\sin \vartheta e^{-i\varphi}$	$(N-1)^2$	\circ	Low	Pair-matching is needed, Performance degradation near $\vartheta = 0^\circ, 180^\circ$
Sine-based [28], [29]	$\sin \vartheta e^{i\varphi}$		$N^2 + \lfloor N/2 \rfloor$	\times	Low	Performance degradation near $\vartheta = 90^\circ$
Vector-based [30], [31]	$\sin \vartheta e^{i\varphi}, \sin \vartheta e^{-i\varphi}, \cos \vartheta$		N^2	\circ	Highest	Joint eigenvalue decomposition is needed
Proposed	$\sin \vartheta e^{i\varphi}, \sin \vartheta e^{-i\varphi}, \cos \vartheta$		$N^2 + N + \lfloor N/3 \rfloor$	\circ	High	

owing to the decoupled estimation is the reduction to $(N-1)^2$ detectable sources.

Recently, the more comprehensive vector-based EB-ESPRIT proposed the use of multiple simultaneous recurrence relations [30], [31]. Subsequent studies considered the reduction of the computational complexity of the vector-based EB-ESPRIT [35] and the mathematical relationship between vector-based EB-ESPRIT and the pseudo-intensity vector [36]. Vector-based EB-ESPRIT overcomes the ambiguity issues and features isotropic estimation accuracy by using three recurrence relations. What distinguishes this approach is the joint estimation of directional parameters at the final stage. The directional parameters are jointly estimated from a generalized eigenvalue decomposition (GEVD) [30], selection of the best result from multiple eigenvalue decompositions (EVDs) [31], or through a joint eigenvalue decomposition (JEVD) algorithm [32]. Compared to the original EB-ESPRIT, this joint estimation with multiple recurrence relations can double the number of detectable sources to N^2 with improved accuracy [30], [31].

In the evolution of EB-ESPRIT, the constraints introduced by recurrence relations turned out to be crucial to the DoA estimation performance as they determine the number of detectable sources. For instance, although up-down ambiguous, the sine-based method clearly demonstrates that there should be a number of detectable sources at least as high as $N^2 + \lfloor \frac{N}{2} \rfloor$. While the state-of-the-art vector-based EB-ESPRIT technique additionally contains the cosine recurrence to resolve up-down ambiguity, it has only been reported to detect up to N^2 sources, so far. We take this as a strong indication that the development of the formalism has yet to be taken to its ultimate boundary. Our strongest indication exists for $N=1$, for which the HARPEX method [37] is able to detect 2 sources, which exceeds both numbers, N^2 and $N^2 + \lfloor \frac{N}{2} \rfloor$.

This study is dedicated to pushing the number of detectable sources in vector-based EB-ESPRIT to its upper boundaries. This is done by exploiting expressions that the current vector-based EB-ESPRIT considered to be indeterminable. Hereby, the number of detectable sources is raised to $N^2 + N + \lfloor \frac{N}{3} \rfloor$, at maintained numerical accuracy. It is likely to be the ultimate limit for the number of detectable sources, at least higher than reported for any conventional EB-ESPRIT technique, and close to the unreachable limit $(N+1)^2$.

II. EB-ESPRIT PROBLEM

A. Description of Sound Fields

Suppose that a sound field consists of Q plane waves (directional waves). The SH coefficients of the q^{th} plane wave of the complex amplitude s_q ($q=1, \dots, Q$) at a single frequency are given by [38]

$$\mathbf{a}_q = \mathbf{y}^*(\Omega_q) s_q, \quad (1)$$

for the propagating direction $\Omega_q = (\vartheta_q, \varphi_q)$. The vector \mathbf{y}^* is composed of conjugated spherical harmonics of order n and degree m such that

$$\mathbf{y}^*(\Omega_q) = \left[\underbrace{Y_0^0(\Omega_q)}_{n=0}, \underbrace{Y_1^{-1}(\Omega_q), Y_1^0(\Omega_q), Y_1^1(\Omega_q)}_{n=1}, \dots, \underbrace{Y_N^{-N}, \dots, Y_N^N(\Omega_q)}_{n=N} \right]^H, \quad (2)$$

where $(\cdot)^H$ is the conjugate transpose, and the complex spherical harmonics are defined as [39]

$$Y_n^m(\Omega_q) = \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos \vartheta_q) e^{im\varphi_q}, \quad (3)$$

where P_n^m denotes the associated Legendre functions. In the vector \mathbf{y}^* , order- n components of length $2n+1$ are stacked vertically, so the vector $\mathbf{a}_q \in \mathbb{C}^{L \times 1}$ measured up to the N^{th} order spherical harmonics has a size of $L = (N+1)^2$.

The measurement is usually contaminated by noise components, and EB-ESPRIT attempts to separate the subspace spanned by the actual sound field (signal subspace) from those spanned by the noise (noise subspace) using an EVD. The total observed sound field consisting of Q plane waves contaminated by noises can be modeled as

$$\tilde{\mathbf{a}} = \left[\mathbf{y}^*(\Omega_1) \quad \dots \quad \mathbf{y}^*(\Omega_Q) \right] \begin{bmatrix} s_1 \\ \vdots \\ s_Q \end{bmatrix} + \mathbf{n} = \mathbf{Y}\mathbf{s} + \mathbf{n}, \quad (4)$$

where $\mathbf{s} = [s_1, \dots, s_Q]^T$ contains complex amplitudes of Q plane waves, and the spherical harmonics manifold $\mathbf{Y} \in \mathbb{C}^{L \times Q}$ contains SH coefficients of Q plane waves. The noise included

in the measurement is denoted by the vector \mathbf{n} . Note that the manifold \mathbf{Y} is only of rank Q if the number of plane waves is bounded by the total number of harmonics ($Q \leq L$).

To extract the signal subspace, subspace-based techniques consider a modeled covariance of $\tilde{\mathbf{a}}$ given by

$$\begin{aligned} \mathbf{R} &= E \{ \tilde{\mathbf{a}}\tilde{\mathbf{a}}^H \} = \mathbf{Y}E \{ \mathbf{s}\mathbf{s}^H \} \mathbf{Y}^H + E \{ \mathbf{n}\mathbf{n}^H \} \\ &= \mathbf{Y}\mathbf{R}_s\mathbf{Y}^H + \mathbf{R}_{ns} \end{aligned} \quad (5)$$

for noises \mathbf{n} uncorrelated with the short-time stationary sound field $\mathbf{Y}\mathbf{s}$. Here, $\mathbf{R}_s = E\{\mathbf{s}\mathbf{s}^H\}$ and $\mathbf{R}_{ns} = E\{\mathbf{n}\mathbf{n}^H\}$ represent the source and noise covariance matrices, respectively, and $E\{\cdot\}$ denotes the expectation operator. Through the eigenvalue decomposition, the observed covariance \mathbf{R} can be rewritten in terms of the eigenvalues and eigenvectors as

$$\mathbf{R} = \mathbf{U}_s\mathbf{\Sigma}_s\mathbf{U}_s^H + \mathbf{U}_{ns}\mathbf{\Sigma}_{ns}\mathbf{U}_{ns}^H, \quad (6)$$

where columns of \mathbf{U}_s , \mathbf{U}_{ns} are eigenvectors of signal and noise subspaces, respectively, and $\mathbf{\Sigma}_s$ are the high eigenvalues distinguishing the signal subspace from the noise subspace. The signal subspace eigenvectors \mathbf{U}_s are the basis expressing the observed sound field. A transformation $\mathbf{T} \in \mathbb{C}^{Q \times Q}$ exists relating the model $\mathbf{Y}\mathbf{R}_s\mathbf{Y}^H$ to the observation $\mathbf{U}_s\mathbf{\Sigma}_s\mathbf{U}_s^H$

$$\mathbf{U}_s = \mathbf{Y}\mathbf{T}. \quad (7)$$

Note that unique factorization of \mathbf{U}_s into \mathbf{Y} times \mathbf{T} requires the spherical harmonic coefficients observed to outnumber the plane waves of the model $Q < L$. Otherwise, if $Q = L$, a matrix $\mathbf{T} = \mathbf{Y}^{-1}\tilde{\mathbf{Y}}\tilde{\mathbf{T}}$ exists to replace the set of plane-wave directions in \mathbf{Y} by another set $\tilde{\mathbf{Y}}$ able to span the subspace.

B. Vector-Based EB-ESPRIT

The EB-ESPRIT methods make use of the internal structure of the SH matrix \mathbf{Y} implied in the signal subspace of (7), and in particular, the EB-ESPRIT technique proposed in [30], [31] is explained here. This technique utilizes three recurrence relations of spherical harmonics corresponding to three linear factors related to the Cartesian coordinates [40], [41], so we may denote it as the *vector-based EB-ESPRIT*. These three recurrence relations [42] related to the complex-valued directional parameters $\theta_{xy} = \sin \vartheta e^{i\varphi}$, $\theta_{xy}^* = \sin \vartheta e^{-i\varphi}$, and $\theta_z = \cos \vartheta$ as linear factors are:

$$\begin{aligned} \theta_{xy}^* Y_n^{m*} &= w_n^{-m} Y_{n-1}^{m+1*} - w_{n+1}^{m+1} Y_{n+1}^{m+1*} \\ \theta_{xy} Y_n^m &= -w_n^m Y_{n-1}^{m-1} + w_{n+1}^{-m+1} Y_{n+1}^{m-1} \\ \theta_z Y_n^m &= v_n^m Y_{n-1}^{m*} + v_{n+1}^m Y_{n+1}^{m*}, \end{aligned} \quad (8)$$

where the recurrence coefficients are defined as

$$w_n^m = \sqrt{\frac{(n+m-1)(n+m)}{(2n-1)(2n+1)}}, \quad v_n^m = \sqrt{\frac{(n-m)(n+m)}{(2n-1)(2n+1)}}, \quad (9)$$

and Ω is omitted for simplicity.

In (8), these three recurrences are related to three directional parameters on the left-hand side. All of them can be used to extract a DoA $\Omega_q = (\vartheta_q, \varphi_q)$ from every column of the spherical harmonic matrix $\mathbf{Y} = \mathbf{U}_s\mathbf{T}^{-1}$ implied by the signal subspace (7). The recurrence relations involve spherical harmonics

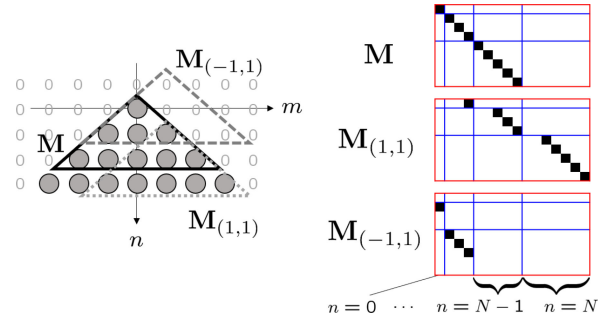


Fig. 1. Definition of masking matrices of the first equation in (8) for $N = 3$. (Left) triangles represent subarrays defined by masking matrices. Circles denote positions of SH coefficients. (Right) illustration of masking matrices (black cell = 1, white cell = 0).

of different orders and degrees. Since we need to relate the spherical harmonics of the low order to the higher order (e.g., Y_n^m and Y_{n+1}^m), it is convenient to define a subarray \mathbf{Y}_{N-1} with spherical harmonics of reduced order up to $N - 1$.

$$\mathbf{Y}_{N-1} = \begin{bmatrix} \mathbf{y}_{N-1}^*(\Omega_1) & \cdots & \mathbf{y}_{N-1}^*(\Omega_Q) \end{bmatrix} = \mathbf{M}\mathbf{Y}, \quad (10)$$

where $\mathbf{y}_{N-1}^*(\Omega_q) = [Y_0^0, Y_1^{-1}, Y_1^0, Y_1^1, \dots, Y_{N-1}^{N-1}]^H$. Note that the subarray \mathbf{Y}_{N-1} has reduced number of rows ($L_1 = N^2$), while its column size is unchanged (Q).

The binary mask matrix $\mathbf{M} = [\mathbf{I}, \mathbf{0}]$ is an $L_1 \times L$ identity matrix with zeros in its last $2N + 1$ columns defined for the highest-order components ($n = N$). This matrix extracts an order-reduced subarray of SH coefficients by truncating the highest order components.

Using the matrices defined, the three recurrence relations of (8) can be generically described in matrix form

$$\mathbf{M}\mathbf{Y}\mathbf{\Theta}_{\{xy^*, xy, z\}} = \mathbf{D}_{\{xy^*, xy, z\}}\mathbf{Y}, \quad (11)$$

where three parameter matrices $\mathbf{\Theta}_{\{xy^*, xy, z\}} \in \mathbb{C}^{Q \times Q}$ are diagonal matrices, each of which containing one of the three directional parameters $\{\theta_{xy^*, q}^*, \theta_{xy, q}, \theta_{z, q}\}$ for all of the Q sources. The recurrence matrices $\mathbf{D}_{\{xy^*, xy, z\}}$ are given by

$$\begin{aligned} \mathbf{D}_{xy^*} &= \overline{\mathbf{W}}_{(0,0)}\mathbf{M}_{(-1,1)} - \mathbf{W}_{(1,1)}\mathbf{M}_{(1,1)} \\ \mathbf{D}_{xy} &= -\mathbf{W}_{(0,0)}\mathbf{M}_{(-1,-1)} + \overline{\mathbf{W}}_{(1,1)}\mathbf{M}_{(1,-1)} \\ \mathbf{D}_z &= \mathbf{V}_{(1,0)}\mathbf{M}_{(-1,0)} + \mathbf{V}_{(0,0)}\mathbf{M}_{(1,0)}. \end{aligned} \quad (12)$$

The recurrence coefficients are now expressed by the diagonal, order-limited matrices that are $\in \mathbb{R}^{L_1 \times L_1}$,

$$\begin{aligned} \mathbf{W}_{(r,s)} &= \text{diag}\{[w_{n+r}^{m+s}]_{nm}\}, \\ \overline{\mathbf{W}}_{(r,s)} &= \text{diag}\{[\overline{w}_{n+r}^{-m+s}]_{nm}\}, \\ \mathbf{V}_{(r,s)} &= \text{diag}\{[v_{n+r}^{m+s}]_{nm}\}. \end{aligned} \quad (13)$$

Any shift in the spherical harmonic indices from (n, m) to $(n + r, m + s)$ is described in terms of order-reducing, binary matrices $\mathbf{M}_{(r,s)} \in \mathbb{R}^{L_1 \times L}$ (Fig. 1). In matrix multiplication with $\mathbf{M}_{(r,s)}$, the shifted harmonics Y_{n+r}^{m+s*} are extracted from the

harmonics $Y_{n'}^{m'}$ by summation over a Kronecker-Delta $\delta_{n,n-r}^{m,m-s'}$ that selects $n = n - r'$ and $m = m - s'$, hence

$$\left[\delta_{n,n-r}^{m,m-s'} \right]_{nm}^{n'm'} \mathbf{Y} = \mathbf{M}_{(r,s)} \mathbf{Y}. \quad (14)$$

Now, the three recurrence relations in (11) are applied to the observed signal subspace \mathbf{U}_s using (7):

$$\begin{aligned} \mathbf{M}\mathbf{U}_s\mathbf{T}^{-1}\Theta_{\{xy^*,xy,z\}} &= \mathbf{D}_{\{xy^*,xy,z\}}\mathbf{U}_s\mathbf{T}^{-1} \\ \Rightarrow \mathbf{M}\mathbf{U}_s\mathbf{\Psi}_{\{xy^*,xy,z\}} &= \mathbf{D}_{\{xy^*,xy,z\}}\mathbf{U}_s. \end{aligned} \quad (15)$$

Here, $\mathbf{\Psi} \in \mathbb{C}^{Q \times Q}$ denote modified direction parameter matrices related to their respective diagonal counterpart Θ by

$$\mathbf{\Psi}_{\{xy^*,xy,z\}} = \mathbf{T}^{-1}\Theta_{\{xy^*,xy,z\}}\mathbf{T} \quad (16)$$

through eigendecomposition by an unknown transformation matrix \mathbf{T} . Hence, EB-ESPRIT first uses the recurrence relations to estimate $\mathbf{\Psi}$, from which it subsequently extracts the direction parameters on the diagonal of the eigenvalue matrix Θ . It will be beneficial later to express the three recurrence relations of (15) in a single equation by vertically stacking the parameter matrices,

$$\underbrace{\begin{bmatrix} \mathbf{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{M} \end{bmatrix}}_{\mathbf{M}_T} \underbrace{\begin{bmatrix} \mathbf{U}_s & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{U}_s \end{bmatrix}}_{\mathbf{U}_T} \underbrace{\begin{bmatrix} \mathbf{\Psi}_{xy^*} \\ \mathbf{\Psi}_{xy} \\ \mathbf{\Psi}_z \end{bmatrix}}_{\mathbf{\Psi}_T} = \underbrace{\begin{bmatrix} \mathbf{D}_{xy^*} \\ \mathbf{D}_{xy} \\ \mathbf{D}_z \end{bmatrix}}_{\mathbf{D}_T} \mathbf{U}_s. \quad (17)$$

Note that the zero matrices ($\mathbf{0}$) used to describe \mathbf{M}_T and \mathbf{U}_T are of different sizes ($L_1 \times L$ and $L \times Q$, respectively) but are denoted by the same symbol $\mathbf{0}$ for the simplicity of expression. In this study, the size of a zero matrix is not specified if it can be inferred from the size of adjacent matrices.

The product of two block diagonal matrices $\mathbf{M}_T \in \mathbb{R}^{3L_1 \times 3L}$, $\mathbf{U}_T \in \mathbb{C}^{3L \times 3Q}$ is denoted as the EB-ESPRIT matrix. If there is a suitable inverse $(\mathbf{M}_T\mathbf{U}_T)^+$ of the EB-ESPRIT matrix satisfying $(\mathbf{M}_T\mathbf{U}_T)^+(\mathbf{M}_T\mathbf{U}_T)\mathbf{\Psi}_T = \mathbf{\Psi}_T$ then we can use it to estimate the parameter matrix $\mathbf{\Psi}_T$ by

$$\begin{bmatrix} \hat{\mathbf{\Psi}}_{xy^*} \\ \hat{\mathbf{\Psi}}_{xy} \\ \hat{\mathbf{\Psi}}_z \end{bmatrix} = (\mathbf{M}_T\mathbf{U}_T)^+ \mathbf{D}_T \mathbf{U}_s, \quad (18)$$

where $(\cdot)^+$ denote the pseudo-inverse (Moore-Penrose inverse) operation. Due to the block diagonal property of the EB-ESPRIT matrix, the inversion is equivalent to

$$\begin{bmatrix} \hat{\mathbf{\Psi}}_{xy^*} \\ \hat{\mathbf{\Psi}}_{xy} \\ \hat{\mathbf{\Psi}}_z \end{bmatrix} = \begin{bmatrix} (\mathbf{M}\mathbf{U}_s)^+ & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{M}\mathbf{U}_s)^+ & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\mathbf{M}\mathbf{U}_s)^+ \end{bmatrix} \begin{bmatrix} \mathbf{D}_{xy^*} \mathbf{U}_s \\ \mathbf{D}_{xy} \mathbf{U}_s \\ \mathbf{D}_z \mathbf{U}_s \end{bmatrix}. \quad (19)$$

So, in the conventional vector-based ESPRIT, only the pseudo-inverse of a small matrix $(\mathbf{M}\mathbf{U}_s)^+$ is required. If there is a unique solution such that $(\mathbf{M}\mathbf{U}_s)^+(\mathbf{M}\mathbf{U}_s) = \mathbf{I}$, then the rows of the matrix on the right-hand side of (18) can be partitioned into the three desired transformed parameter matrices, cf. (15). From the property of (16), all three matrices $\hat{\mathbf{\Psi}}_{\{xy^*,xy,z\}}$ share the same

transformation matrix \mathbf{T} . Therefore, the joint eigenvalue decomposition (JEVD) of these matrices is used to extract the DoA parameters $\Theta_{\{xy^*,xy,z\}}$ as eigenvalue matrices and the transformation \mathbf{T} as a joint eigenvector matrix. Many iterative JEVD algorithms [43]–[46] can be used to solve the JEVD problem. However, as mentioned in [31], the JEVD algorithm itself does not bring significant performance improvement compared to the ad-hoc approach, which chooses one of the eigenvector matrices of $\hat{\mathbf{\Psi}}_{xy^*}$, $\hat{\mathbf{\Psi}}_{xy}$, and $\hat{\mathbf{\Psi}}_z$ that minimizes the JEVD criterion. For this reason, the simple ad-hoc approach is adopted in this study.

C. Uniqueness of Solution and Number of Detectable Sources

In (18), the uniqueness of a solution is the most important prerequisite that limits the number of detectable sources. The total number of unknowns, i.e., the number of parameters to be estimated, is $3Q$. Since we assume that the number of sources is smaller than the number of SH coefficients ($Q < L$), the matrix \mathbf{U}_T has more rows ($3L$) than columns ($3Q$). Moreover, a single block \mathbf{U}_s consists of orthogonal eigenvectors so that the rank of \mathbf{U}_T is full ($3Q$).

Therefore, the uniqueness of the solution depends on the rank of the binary masking matrix \mathbf{M}_T . As already mentioned, its single block \mathbf{M} is an identity matrix for $n \leq N - 1$ providing N^2 linearly independent rows. Therefore, the rank of the matrix \mathbf{M}_T is given by $3 \times N^2$. From the following rank property of two matrices [47],

$$\text{rank}\{\mathbf{M}_T\mathbf{U}_T\} \leq \text{Min}[\text{rank}\{\mathbf{M}_T\}, \text{rank}\{\mathbf{U}_T\}] \quad (20)$$

we need $3Q \leq 3N^2$ to uniquely determine the desired $3Q$ parameters in $\hat{\mathbf{\Psi}}_{\{xy^*,xy,z\}}$. This condition sets the maximum bound on the number of detectable sources for the vector-based EB-ESPRIT: $Q_{max} = N^2$, which is the same number as discussed in the literature [30], [31].

The high resolution plane wave expansion (HARPEX) [37] is able to detect $Q = 2$ sources for $N = 1$. If this is not an exception for $N = 1$, there must be ways to increase the number of detectable sources beyond N^2 in vector-based EB-ESPRIT, in general. This requires to define suitable extra conditions capable of increasing the rank of \mathbf{M}_T .

III. PROPOSED METHOD

To come up with a proposed solution, we revisit the recurrences of vector-based EB-ESPRIT to see where it left out constraints. For mathematical brevity, the derivation starts with conjugation and the normalization of the harmonics omitted

$$\hat{Y}_n^m = P_n^m(\cos \vartheta) e^{im\varphi} = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} Y_n^m, \quad (21)$$

which modifies the three recurrences of (8), equivalently, to

$$\begin{aligned} (2n+1)\theta_{xy}\hat{Y}_n^m &= \hat{Y}_{n-1}^{m+1} - \hat{Y}_{n+1}^{m+1} \\ (2n+1)\theta_{xy}^*\hat{Y}_n^m &= -(n+m-1)(n+m)\hat{Y}_{n-1}^{m-1} \\ &\quad + (n-m+1)(n-m+2)\hat{Y}_{n+1}^{m-1} \\ (2n+1)\theta_z\hat{Y}_n^m &= (n+m)\hat{Y}_{n-1}^m + (n-m+1)\hat{Y}_{n+1}^m \end{aligned} \quad (22)$$

These recurrences clearly hold for any $0 \leq n \leq N$, and $|m| \leq n$. Without loss of information, the $n - 1$ order right-hand side harmonic of any recurrence vanishes whenever $n - 1 < 0$ or its particular degree $|m \pm 1|, |m|$ exceeds $|n - 1|$.

On the quest for the information discarded by vector-based EB-ESPRIT, it is rewarding to inspect the highest-order: vector-based EB-ESPRIT omitted the $n = N$ recurrences since their right-hand side harmonics \hat{Y}_{n+1}^μ for $\mu = \{m \pm 1, m\}$ were reasonably assumed to be outside of the observable range of an order- N spherical microphone array. However, we may eliminate any $Y_{n+1}^{m \pm 1}$ harmonic from the order $n = N$ recurrences for θ_{xy} and θ_{xy}^* by the Y_{n+1}^m harmonic of the θ_z recurrence suitably shifted in m .

The expression \hat{Y}_{n+1}^{m+1} is eliminated from the θ_{xy} recurrence after it is multiplied with $(n - m)$, summed with the θ_z recurrence shifted to $m \rightarrow m + 1$, and divided by $(2n + 1)$,

$$(n - m)\theta_{xy}\hat{Y}_n^m + \theta_z\hat{Y}_n^{m+1} = \hat{Y}_{n-1}^{m+1}. \quad (23)$$

This recurrence only exists for $-n \leq m \leq n - 1$ because of the shift in m .

The expression \hat{Y}_{n+1}^{m-1} is eliminated from the θ_{xy}^* recurrence by subtracting the θ_z recurrence shifted by $m \rightarrow m - 1$ and multiplied with $(n - m + 1)$, followed by division by $(2n + 1)$

$$\theta_{xy}^*\hat{Y}_n^m - (n - m + 1)\theta_z\hat{Y}_n^{m-1} = -(n + m - 1)\hat{Y}_{n-1}^{m-1}. \quad (24)$$

This recurrence only exists for $-n + 1 \leq m \leq n$ because of the shift in m .

The proposed solution is to use (23), (24) for $n = N$ to extend the set of recurrences in (22) to the highest order. With normalization re-inserted using (21) and complex conjugation that affects $\hat{Y}_n^m, \theta_{xy}, \theta_{xy}^*$, the extending relations become

$$\begin{aligned} \theta_{xy}^*\eta_n^{-m}Y_n^{m*} + \theta_z\eta_n^{m+1}Y_n^{m+1*} &= \eta_{n-1}^{-m}Y_{n-1}^{m+1*} \\ &\text{for } -n \leq m \leq n - 1, \\ \theta_{xy}\eta_n^mY_n^{m*} - \theta_z\eta_n^{-m+1}Y_n^{m-1*} &= -\eta_{n-1}^mY_{n-1}^{m-1*} \\ &\text{for } -n + 1 \leq m \leq n, \end{aligned}$$

$$\text{with } \eta_n^m = \sqrt{(n + m)/(2n + 1)}, \quad (25)$$

which is employed for $n = N$ in addition to (8).

A. Integration of Proposed Extension, Increase in Number of Detectable Sources

Similar as in (12), we define a matrix notation to integrate the pair of $2N$ extending equations in our system of equations

$$\underbrace{\begin{bmatrix} \bar{\mathbf{A}} & \mathbf{0} & \mathbf{A} \\ \mathbf{0} & \mathbf{A} & -\bar{\mathbf{A}} \end{bmatrix}}_{\mathbf{M}_L} \underbrace{\begin{bmatrix} \mathbf{U}_s & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{U}_s \end{bmatrix}}_{\mathbf{U}_T} \underbrace{\begin{bmatrix} \Psi_{xy^*} \\ \Psi_{xy} \\ \Psi_z \end{bmatrix}}_{\Psi_T} = \underbrace{\begin{bmatrix} \bar{\mathbf{C}} \\ -\mathbf{C} \end{bmatrix}}_{\mathbf{C}_L} \mathbf{U}_s, \quad (26)$$

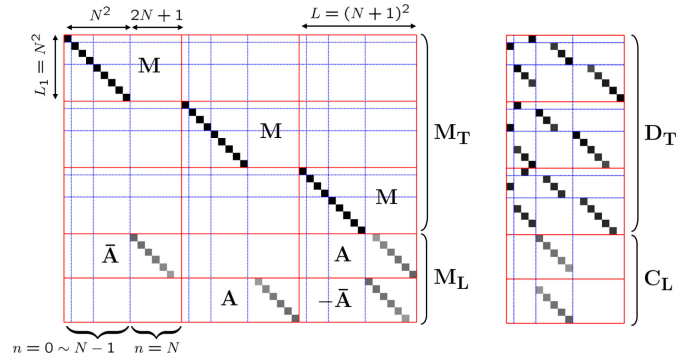


Fig. 2. Grayscale mapping of matrix entries of $[\mathbf{M}_T; \mathbf{M}_L]$ and $[\mathbf{D}_T; \mathbf{C}_L]$ for $N = 3$. (Red lines: boundaries between $L_1 \times L$ submatrices of \mathbf{M}_T and $2N \times L$ submatrices of \mathbf{M}_L , blue lines: boundaries between different orders).

using the matrices

$$\mathbf{A} = \mathbf{H}_{(0,0)} \mathbf{M}_{(0,0)}^{-N+1 \dots N} = \mathbf{H}_{(0,1)} \mathbf{M}_{(0,-1)}^{-N \dots N-1}$$

$$\bar{\mathbf{A}} = \bar{\mathbf{H}}_{(0,0)} \mathbf{M}_{(0,0)}^{-N \dots N-1} = \bar{\mathbf{H}}_{(0,1)} \mathbf{M}_{(0,1)}^{-N+1 \dots N}$$

$$\mathbf{C} = \mathbf{H}_{(-1,0)} \mathbf{M}_{(-1,-1)}^{-N+1 \dots N}$$

$$\bar{\mathbf{C}} = \bar{\mathbf{H}}_{(-1,0)} \mathbf{M}_{(-1,1)}^{-N \dots N-1}$$

$$\mathbf{H}_{(r,s)} = \text{diag}\{\{\eta_{N+r}^{m+s}\}_{m=-N+1 \dots N}\},$$

$$\bar{\mathbf{H}}_{(r,s)} = \text{diag}\{\{\eta_{N+r}^{-m+s}\}_{m=-N \dots N-1}\}. \quad (27)$$

Here, $\mathbf{M}_{(r,s)}^{-N+1 \dots N}$ and $\mathbf{M}_{(r,s)}^{-N \dots N-1}$ only extract the $2N$ rows of $n = N$ in the range of m specified in superscript. Equation (26) has the same form as the original EB-ESPRIT equation of (17), so we can utilize it to increase the number of independent rows. Stacking both matrix equations yields

$$\begin{bmatrix} \mathbf{M}_T \\ \mathbf{M}_L \end{bmatrix} \mathbf{U}_T \Psi_T = \begin{bmatrix} \mathbf{D}_T \\ \mathbf{C}_L \end{bmatrix} \mathbf{U}_s, \quad (28)$$

which augments the number of rows of the masking matrix, cf. Fig. 2. The added masking matrix \mathbf{M}_L no longer has binary-valued entries, but this is not important in view of invertibility.

The original masking matrix \mathbf{M}_T is a diagonal matrix and provides $3L_1$ unique row vectors for orders up to $N - 1$, and the proposed extending equations provide $4N$ extra rows that are depicted in Fig. 2. These rows are described by the matrix \mathbf{M}_L for the components of the highest order N . Fig. 2 clearly shows that these $4N$ rows are linearly independent and therefore increase the rank from $3L_1 = 3N^2$ to $3N^2 + 4N$. In principle, the rank required to fully decompose $L = (N + 1)^2$ directional parameters would be $3L = 3(N + 1)^2$.

Consequently, whenever the rank of \mathbf{U}_s is large enough, the 3 directional parameters can be retrieved for a maximum number of detectable sources

$$Q_{max} = N^2 + \lfloor \frac{4N}{3} \rfloor = N^2 + N + \lfloor \frac{N}{3} \rfloor. \quad (29)$$

For example, in the case of $N = 1$, we get 4 extending equations in addition to the 3 known ones. Divided by the three directional parameters, the resulting number of detectable independent

sources is $Q_{max} = \lfloor (3 + 4)/3 \rfloor = 2$. This finally corresponds to how many sources HARPEX [37], [48] is capable of detecting for $N = 1$.

Suggested by the exploitation of degrees of freedom in the equation, the new maximum number of detectable sources Q_{max} might be the ultimate one. We fully utilize the information of three independent directional-parameter recurrence relations of the highest order, which are reduced to two equations by eliminating the unobserved $\hat{Y}_{n+1}^{m\pm 1}$ terms. It is also rather close to the unreachable upper limit $(N + 1)^2$ by a narrow margin of only $\lceil \frac{2N}{3} \rceil + 1$ degrees of freedom.

Note that the number Q of signal-space eigenvalues of the covariance matrix still needs to be supported by distinctly larger signal-space eigenvalues than the ones of the noise and diffuse sound field. Whenever the signal subspace is smaller $Q < Q_{max}$, the increased maximum number of detectable sources may not be fully utilized.

B. Summary of Algorithm

The proposed method can be summarized as follows. First, the stacked transformed parameter matrices Ψ_T are obtained from the observed subspace matrix \mathbf{U}_s by

$$\begin{bmatrix} \hat{\Psi}_{xy^*} \\ \hat{\Psi}_{xy} \\ \hat{\Psi}_z \end{bmatrix} = \left(\begin{bmatrix} \mathbf{M}_T \\ \mathbf{M}_L \end{bmatrix} \begin{bmatrix} \mathbf{U}_s & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{U}_s \end{bmatrix} \right)^+ \begin{bmatrix} \mathbf{D}_{xy^*} \mathbf{U}_s \\ \mathbf{D}_{xy} \mathbf{U}_s \\ \mathbf{D}_z \mathbf{U}_s \\ \mathbf{C}_L \mathbf{U}_s \end{bmatrix}. \quad (30)$$

Note that the data-dependent left inverse recombines the blocks $\mathbf{D}_{xy^*} \mathbf{U}_s$, $\mathbf{D}_{xy} \mathbf{U}_s$, $\mathbf{D}_z \mathbf{U}_s$, and $\mathbf{C}_L \mathbf{U}_s$ of the right matrix. The resulting Ψ_T has the full rank of its Q columns, in general. Then, each of the EVD problems in its three partitions are solved separately to find the corresponding eigenvectors

$$\hat{\Psi}_{\{xy^*, xy, z\}} = \mathbf{T}_{\{xy^*, xy, z\}}^{-1} \Theta_{\{xy^*, xy, z\}} \mathbf{T}_{\{xy^*, xy, z\}}. \quad (31)$$

After separate EVDs of $\hat{\Psi}_{xy^*}$, $\hat{\Psi}_{xy}$, and $\hat{\Psi}_z$, the eigenvector matrix that minimizes the following JEVD criterion is selected as a common transformation matrix:

$$\mathcal{J}_{JEVD}(\mathbf{T}) = \sum_{\mu \in \{xy^*, xy, z\}} \left\| \text{Zdiag}(\mathbf{T} \hat{\Psi}_\mu \mathbf{T}^{-1}) \right\|_F^2, \quad (32)$$

where the $\text{Zdiag}(\cdot)$ operator extracts off-diagonal components, and $\|\cdot\|_F$ denotes the Frobenius norm. The corresponding diagonals of $\mathbf{T} \hat{\Psi}_{\{xy^*, xy, z\}} \mathbf{T}^{-1}$ yield the direction parameters.

C. Reduced Computational Complexity by Two Step Inversion

The computational complexity of the EB-ESPRIT is largely determined by (i) the identification of signal subspace eigenvectors \mathbf{U}_s , (ii) the Moore-Penrose inverse of the EB-ESPRIT matrix, and (iii) the JEVD of three parameter matrices. As (i) and (ii) are efforts common to both the vector-based and the proposed ESPRIT technique, their computational complexity only differs by the inversion of their EB-ESPRIT matrices. For subspace identification, the EVD of an $L \times L$ complex matrix costs $4O(L^3)$ operations [49], [50], making it the main source

of computational load [34]. Several iterative identification techniques were proposed to reduce this load [51].

The inversion of the $(3N^2 + 4N) \times 3Q$ EB-ESPRIT matrix of the proposed technique requires more computation time compared to the $N^2 \times Q$ matrix of vector-based EB-ESPRIT. In general, inversion of a $M \times N$ matrix requires $O(NM^2)$ operations for $M < N$, so that the roughly tripled dimensions increase the complexity by a noticeable factor of $27 + 36/N$.

To reduce the computational complexity, we propose a two step inversion (TSI) approach. To this end, we keep the vector-based EB-ESPRIT matrix with its block diagonal structure separated from its extension to save computational effort. In what follows, we assume that the number of sources Q is in range of $L_1 < Q \leq Q_{max}$. For $Q \leq L_1$, the vector-based EB-ESPRIT can be applied directly.

The augmented EB-ESPRIT equation of (28) is equivalent to solving the following two equations simultaneously:

$$\mathbf{F} \Psi_T = \mathbf{G}, \quad (33)$$

$$\mathbf{J} \Psi_T = \mathbf{K}, \quad (34)$$

where $\mathbf{F} = \mathbf{M}_T \mathbf{U}_T$, $\mathbf{G} = \mathbf{D}_T \mathbf{U}_s$, $\mathbf{J} = \mathbf{M}_L \mathbf{U}_T$, and $\mathbf{K} = \mathbf{C}_L \mathbf{U}_s$. The strategy adopted here is to find the particular solution, i.e., the least-norm solution of (33) first, and then fit its homogeneous solution to fulfill (34).

The first equation is exactly the same as that of the vector-based EB-ESPRIT. The multiplication with the masking matrix \mathbf{M} removes the order- N components, and we denote the order-reduced matrix as $\mathbf{U}_{N-1} = \mathbf{M} \mathbf{U}_s \in \mathbb{C}^{L_1 \times Q}$. Accordingly, the left matrix \mathbf{F} can be rewritten as a block diagonal matrix.

$$\mathbf{F} = \begin{bmatrix} \mathbf{U}_{N-1} & & \\ & \mathbf{U}_{N-1} & \\ & & \mathbf{U}_{N-1} \end{bmatrix}. \quad (35)$$

The pseudo-inverse of \mathbf{F} can be directly calculated by the pseudo-inverse of its single block as mentioned in (19). The full rank QR decomposition of the single block matrix \mathbf{U}_{N-1}^H after complex transposition can be written as

$$\mathbf{U}_{N-1}^H = \mathbf{Q} \mathbf{R}_T = [\mathbf{Q}_p \ \mathbf{Q}_0] \begin{bmatrix} \mathbf{R}_p \\ \mathbf{0} \end{bmatrix}, \quad (36)$$

where $\mathbf{R}_p \in \mathbb{C}^{L_1 \times L_1}$ is an upper triangular, nonsingular matrix. $\mathbf{Q} \in \mathbb{C}^{Q \times Q}$, $\mathbf{Q}_p \in \mathbb{C}^{Q \times L_1}$ and $\mathbf{Q}_0 \in \mathbb{C}^{Q \times (Q-L_1)}$ are orthogonal matrices satisfying $\mathbf{Q}^H \mathbf{Q} = \mathbf{I}$. Especially, \mathbf{Q}_0 contains the null space basis of \mathbf{U}_{N-1} , i.e., $\mathbf{U}_{N-1} \mathbf{Q}_0 = \mathbf{0}$.

The pseudo-inverse of \mathbf{U}_{N-1} is given by [52]

$$\begin{aligned} \mathbf{U}_{N-1}^+ &= \mathbf{U}_{N-1}^H (\mathbf{U}_{N-1} \mathbf{U}_{N-1}^H)^{-1} \\ &= \mathbf{Q} \mathbf{R}_T (\mathbf{R}_T^H \mathbf{Q}^H \mathbf{Q} \mathbf{R}_T)^{-1} = \mathbf{Q}_p (\mathbf{R}_p^H)^{-1}. \end{aligned} \quad (37)$$

In general, the complexity of the pseudo-inverse by QR decomposition is dominated by the $4O(\max[L_1, Q] \min[L_1, Q]^2)$ operations of the QR decomposition [50], [52]. From (35) to

TABLE II
COMPARISON OF COMPUTATIONAL COMPLEXITY

Algorithm	Subspace Identification	Main processing		JEVD (ad-hoc)
		$Q \leq N^2 = L_1$	$Q > N^2 = L_1$	
Vector-based	EVD: $4O(N^6)$	Pseudo-inverse of \mathbf{U}_{N-1}		$3 \times 4O(Q^3)$
		$4O(N^2Q^2)$	N/A	
Proposed (without TSI)	PASTd [35]: $4(N+1)^2Q + O(Q)$	Pseudo-inverse of EB-ESPRIT matrix		
		$27 \times 4O(N^2Q^2)$		
Proposed (with TSI)	$4(N+1)^2Q + O(Q)$	QR-decomposition of \mathbf{U}_{N-1}^H	QR-decomposition of \mathbf{U}_{N-1}^H + Extra inversion of $\mathbf{J}\mathbf{Q}_h$	
		$4O(N^2Q^2)$	$4O(N^4Q) + 4O(N^2(Q - N^2)^2)$	

(37), the particular solution can be rewritten as

$$\Psi_p = \begin{bmatrix} \mathbf{U}_{N-1}^+ & & \\ & \mathbf{U}_{N-1}^+ & \\ & & \mathbf{U}_{N-1}^+ \end{bmatrix} \mathbf{G}. \quad (38)$$

On the other hand, the homogeneous solution to (33) is derived from the null space of \mathbf{U}_{N-1} . If we construct a block diagonal matrix consisting of \mathbf{Q}_0 , then for an arbitrary nonsingular matrix $\mathbf{X}_h \in \mathbb{C}^{3(Q-L_1) \times Q}$, the homogeneous solution can be written as

$$\Psi_h = \begin{bmatrix} \mathbf{Q}_0 & & \\ & \mathbf{Q}_0 & \\ & & \mathbf{Q}_0 \end{bmatrix} \mathbf{X}_h = \mathbf{Q}_h \mathbf{X}_h. \quad (39)$$

which satisfies $\mathbf{F}\Psi_h = \mathbf{0}_{3L_1 \times 3Q}$. The next step is to determine the coefficients \mathbf{X}_h of the homogeneous solution, such that the extending equations (34) are satisfied by the total solution $\hat{\Psi}_T = \Psi_p + \mathbf{Q}_h \mathbf{X}_h$. That is,

$$\mathbf{J}(\Psi_p + \mathbf{Q}_h \mathbf{X}_h) = \mathbf{K}. \quad (40)$$

The solution to (40) can be found as

$$\mathbf{X}_h = (\mathbf{J}\mathbf{Q}_h)^+ (\mathbf{K} - \mathbf{J}\Psi_p), \quad (41)$$

which gives us the final solution

$$\hat{\Psi}_T = \Psi_p + \mathbf{Q}_h (\mathbf{J}\mathbf{Q}_h)^+ (\mathbf{K} - \mathbf{J}\Psi_p). \quad (42)$$

The extra computation required in comparison to the vector-based ESPRIT is only the second term of (42). Since its Moore-Penrose inverse poses the main computational load, the inversion of the small matrix $\mathbf{J}\mathbf{Q}_h \in \mathbb{C}^{4N \times 3(Q-L_1)}$ dominates the additional complexity. Hereby, the TSI approach brings huge saving in complexity, compared to inverting a matrix of the size $(3N^2 + 4N) \times 3Q$. For instance, with $Q = N^2 + 4N/3 \approx Q_{max}$, the sizes of the matrices subject to inversion are roughly $N^2 \times N^2$ for the vector-based EB-ESPRIT part plus $4N \times 4N$ for the extending part, compared to the one-step inversion with $(3N^2 + 4N) \times (3N^2 + 4N)$. The order of computational complexity is summarized in table II.

In addition, TSI can manifest better compatibility with the vector-based EB-ESPRIT. When the number of sources Q is less than or equal to N^2 , the null space does not exist. Accordingly, the computation of the homogeneous solution is not necessary, and the solution of the proposed method with TSI is identical to that of the vector-based EB-ESPRIT. The disadvantage of

the TSI, however, is that the solution satisfying the additional constraints is only searched among the solutions that fit the original three recurrence relations best. In contrast, the inversion of the total augmented EB-ESPRIT matrix (30) can balance the errors from the original and extending constraints.

IV. EVALUATION

The evaluation of the proposed method was conducted for different numbers of sources (Q), signal-to-noise ratios (SNR), and different room conditions. In the subsections A and B, we demonstrate that the proposed method is comparably accurate as the conventional technique for a small number of sources under free-field conditions. When $Q \leq L_1$, the proposed method with TSI is identical to the vector-based EB-ESPRIT. Therefore, only the proposed method without TSI was compared to the conventional technique. In the next steps, we increased Q to L_1 and Q_{max} . For $Q = Q_{max}$, no conventional technique can estimate this number of sources, so the proposed methods with and without TSI were compared. The last evaluation was done in a more realistic condition, i.e., in a closed room with two speech signals.

A. Free-Field Simulation: Setup

To evaluate the performance of the proposed method, numerical simulations were conducted. A rigid-sphere microphone array consisting with 32 microphones was simulated. The microphones were spatially arranged as spherical t-design [53]. Under this configuration, SH coefficients without spatial aliasing were directly computed using (4) with $N = 3$. The source signals of plane waves and microphone self-noises were simulated as uncorrelated white Gaussian noise. The signal-to-noise ratio (SNR) is defined as the ratio of the total power of source signals to the total power of noise. The spatial aliasing was simulated by generating the pressure data of each microphone on the rigid sphere using spherical harmonics up to a sufficient order ($N = 20$) and transforming microphone signals to the SHD up to the truncated order ($N = 3$) by using discrete spherical Fourier transform (SFT) and mode-strength compensation:

$$a_{nm} = \frac{1}{b_n(kr)} \sum_{\ell=1}^{L_{mic}} \alpha_\ell Y_n^m(\Omega_\ell)^* p(\Omega_\ell), \quad (43)$$

here a_{nm} denotes the transformed SH coefficients, ℓ is the microphone index ($1, 2, \dots, L_{mic}$), Ω_ℓ indicates the angular position of ℓ^{th} microphone, $p(\Omega_\ell)$ is the pressure data of the ℓ^{th} microphone, and α_ℓ is the quadrature weight depending on the sampling scheme [39]. For the spherical t-design sampling, the quadrature weight (α_ℓ) is given by $4\pi/L_{mic}$ for all microphones [53]. The division by $b_n(kr)$ represents the compensation of far-field mode strength of a rigid sphere, defined as [38]

$$b_n(kr) = 4\pi i^{n+1} / \left[(kr)^2 h_n^{(1)'}(kr) \right], \quad (44)$$

where $h_n^{(1)'}(kr)$ denotes the derivative of a spherical Hankel function of the first kind [38] at the wave number k and microphone radius r . The target frequency was chosen to satisfy $kr = 2$, such that the noise amplification due to the mode-strength compensation of higher-order components is not severe [39]. The covariance matrix was constructed from 300 independent temporal snapshots [54].

The estimation errors were examined as angular distances between true DoAs (ϑ_q, φ_q) and estimated DoAs ($\hat{\vartheta}_q, \hat{\varphi}_q$). The root-mean-square error (RMSE) is defined as

$$\text{RMSE}(\Omega) = \sqrt{\sum_{j=1}^J \sum_{q=1}^Q \left| \Delta\Omega_q^{(j)} \right|^2 / JQ}, \quad (45)$$

where $\Delta\Omega_q^{(j)}$ is the DoA difference of the q th source at the j^{th} trial, given by $\Delta\Omega_q^{(j)} = \cos^{-1} \{ \cos \vartheta_q^{(j)} \cos \hat{\vartheta}_q^{(j)} + \cos(\varphi_q^{(j)} - \hat{\varphi}_q^{(j)}) \sin \vartheta_q^{(j)} \sin \hat{\vartheta}_q^{(j)} \}$. The RMSEs were averaged for $J = 400$ independent trials. As a theoretical lower bound of RMSEs, the stochastic Cramér-Rao lower bound (CRB) for DoA estimation in the SHD was computed. The derivation of CRB is given in [55] and involved with covariance matrices (\mathbf{R}, \mathbf{R}_s), the manifold matrix in the SHD (\mathbf{Y}), and the partial derivatives of the manifold matrix with respect to zenith and azimuth angles. The calculated CRBs are shown in the following results. For the comparison with other techniques, the RMSEs of EB-ESPRIT [24], sine-based EB-ESPRIT [28], [29] and vector-based EB-ESPRIT [30], [31] were evaluated under the same conditions.

B. Free-Field Simulation: Results

To examine the performances for various SNRs with two sources, the RMSEs of the geodesic distance were computed. The sound sources were positioned at $(\vartheta_1, \varphi_1) = (33^\circ, 45^\circ)$, $(\vartheta_2, \varphi_2) = (57^\circ, 68^\circ)$. The SNRs were varied from 0 dB to 30 dB. Fig. 3(a) shows the performances of the case without aliasing. All methods show that RMSEs are inversely proportional to the SNR. Without spatial aliasing, the proposed method outperforms EB-ESPRIT by far and has a slightly larger RMSEs than the vector-based EB-ESPRIT, but the error differences are insignificant. In the case with aliasing (Fig. 3(b)), the performance ranking is the same for the low SNR conditions (SNR = 0, 5 dB), but it can be seen that the RMSEs of the proposed method are slightly higher than those of the EB-ESPRIT and vector-based EB-ESPRIT for very high SNRs greater than 10 dB. Nevertheless, the RMSE differences between vector-based EB-ESPRIT and the proposed method without TSI are still small (around 0.08 degrees). This suggests that the DoA estimation

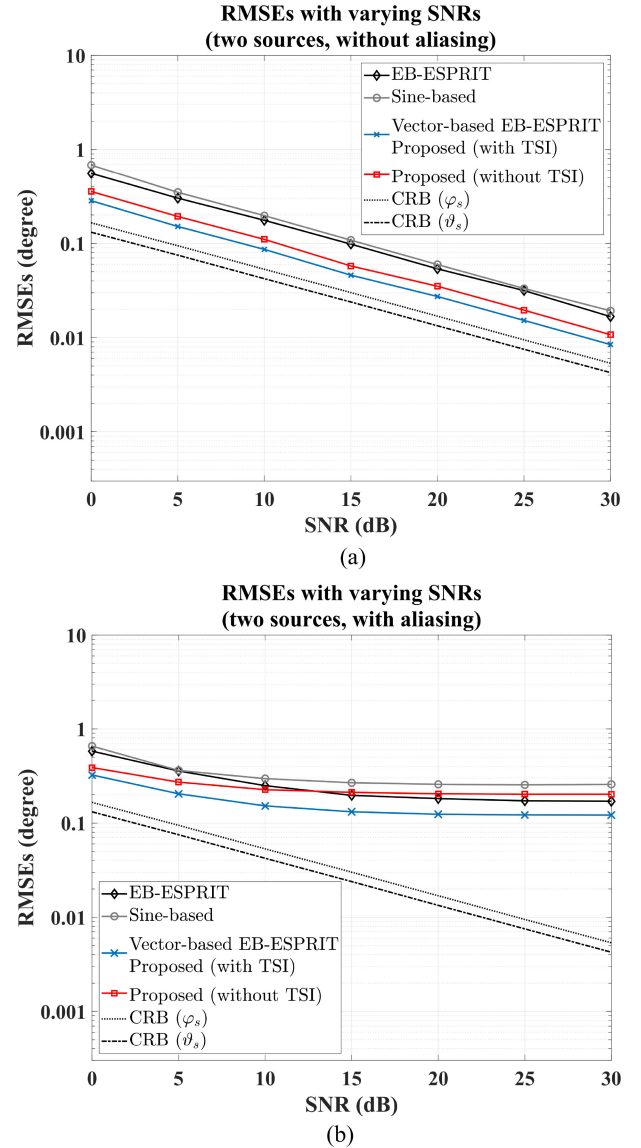


Fig. 3. RMSEs of two sound sources for various SNRs in a free-field condition (a) without spatial aliasing (b) with spatial aliasing.

performance of the proposed method is similar to those of the vector-based EB-ESPRIT but slightly less robust to model mismatches or spatial aliasing. This can be confirmed by comparing the 2-norm condition number of the EB-ESPRIT matrices used as a robustness measure of the EB-ESPRIT techniques [33]. Fig. 4 shows the condition number for each algorithm at different zenith angles ($0^\circ \leq \vartheta_1 \leq 180^\circ$) of a single sound source at azimuth $\varphi_1 = 45^\circ$. This simulation was done without aliasing and noises. The vector-based EB-ESPRIT has value 1 for all zenith angles, while the proposed method increased slightly to 1.118. However, this number is still not high, especially when taking the benefit of increasing the number of detectable sources into account.

To verify this benefit in terms of a noticeable improvement of DoA estimation performance with more than two sources, the RMSEs for 9 ($Q = L_1$) and 13 ($Q = Q_{max}$) sources were

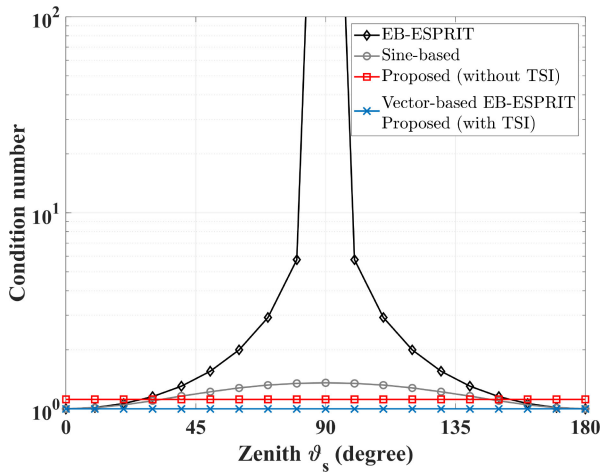
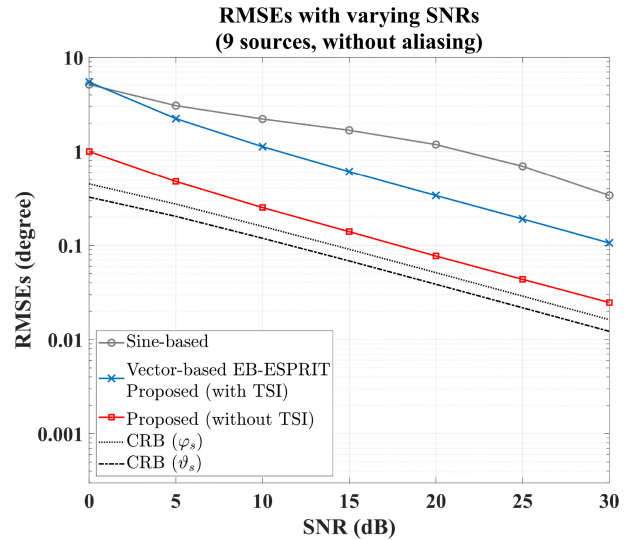


Fig. 4. 2-norm condition numbers of EB-ESPRIT matrices for various zenith angles ($0^\circ \leq \vartheta_s \leq 180^\circ$) in a free-field condition (without spatial aliasing and noises).

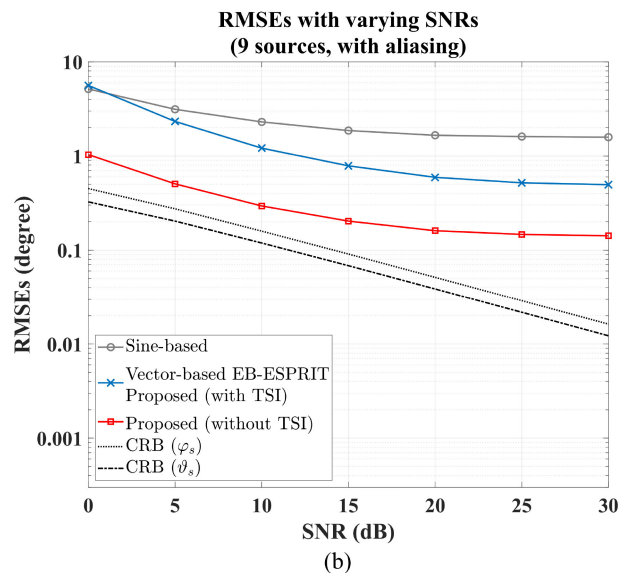
computed under different SNR conditions. As the EB-ESPRIT technique cannot detect these numbers of sources, it is excluded from the simulation. The DoAs of sources were randomly selected from a set of 48 positions determined by a spherical t-design [53]. Fig. 5(a) shows the RMSE results for 9 sources without spatial aliasing. Similar to the results for two sources (Fig. 3(a)), the RMSEs of three different methods are inversely proportional to the SNR. However, unlike the results for two sources, the proposed method now outperforms both the other methods. The RMSEs of the proposed method also stay smaller than those of the other methods under the presence of spatial aliasing (Fig. 5(b)). Apparently, the additional highest-order recurrence relations (25) increase the estimation performance when several concurrent sources are active. With a reduced number of sources of L_1 or slightly smaller, a trade off between computational cost and estimation accuracy becomes obvious. In this case, higher accuracy is achieved by the proposed method without TSI, while vector-based EB-ESPRIT (or the proposed method with TSI) provides faster computation speeds.

Fig. 6 shows the RMSE results for 13 sources ($Q = Q_{max}$). Since no conventional method can detect this number of sources, RMSEs were computed only for two types of the proposed method (with and without TSI). For both cases with and without aliasing (Fig. 6(a), (b)), the estimation performances of two methods are almost the same. However, as mentioned in Section III-C, the computational cost with TSI is much smaller than without TSI. For a large number of sources ($Q > L_1$), therefore, the TSI becomes more efficient without noticeable degradation in the estimation performance.

To analyze the DoA estimation accuracy with respect to the DoA of a sound source, RMSEs were evaluated for various DoAs (Fig. 7). The DoA of a single sound source was selected from all directions of 48 spherical t-design sampling [53]. The SNR was set to 0 dB. As can be expected from the results in condition number (Fig. 4), the original EB-ESPRIT exhibits a significant singularity near the horizon ($\vartheta = 90^\circ$). By contrast, the two types of the proposed method (with and without TSI) are both free



(a)



(b)

Fig. 5. RMSEs of 9 sound sources for various SNRs in a free-field condition (a) without spatial aliasing (b) with spatial aliasing.

from any singularities or degradation depending on DoAs. As shown in Fig. 3(a), the proposed method without TSI has slightly larger RMSEs than the vector-based EB-ESPRIT by about 0.2 degrees, which is still insignificant.

The maximum number of detectable sources is also verified in terms of RMSEs. In the simulations shown in Figs. 8 and 9, the proposed method was compared with all methods (EB-ESPRIT, sine-based EB-ESPRIT, and vector-based EB-ESPRIT). Among those alternatives, sine-based EB-ESPRIT was known to provide the highest number of detectable sources but with decreased accuracy near the horizon $\vartheta = 90^\circ$. For both simulations, the source DoAs were randomly generated and the covariance matrix was ideally constructed, disregarding noise and spatial aliasing. In Fig. 8, RMSEs were calculated when varying the number of sources from 2 to 25 using an expansion order $N = 4$. The fundamental ambiguities of both the original EB-ESPRIT and sine-based EB-ESPRIT [33] were resolved by choosing one of

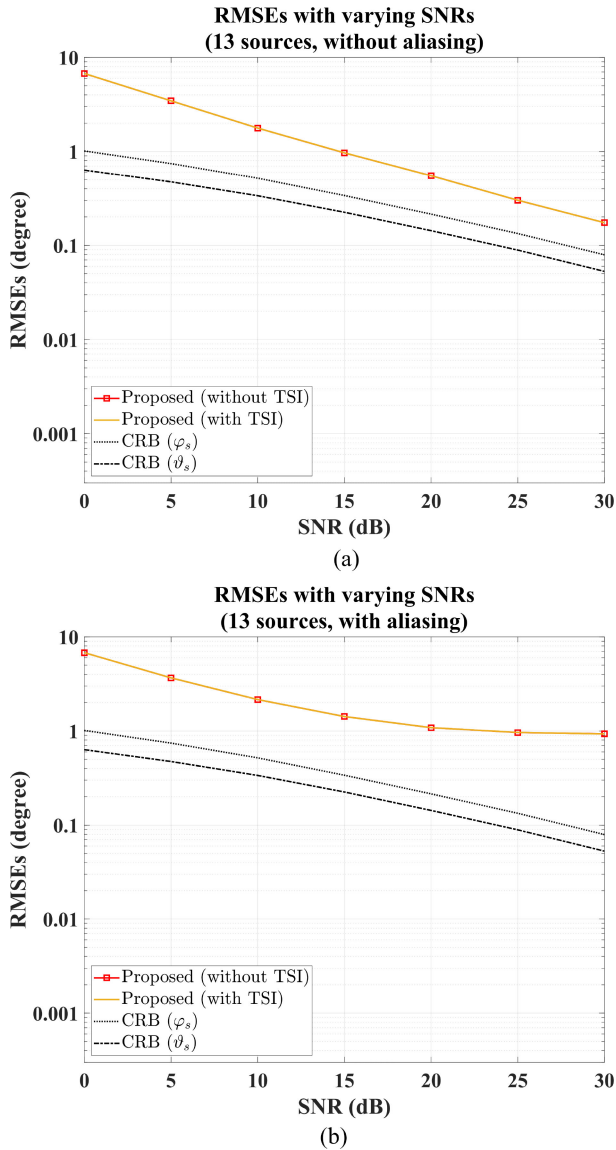


Fig. 6. RMSEs of 13 sound sources for various SNRs in a free-field condition (a) without spatial aliasing (b) with spatial aliasing.

two ambiguous DoAs based on the EB-MUSIC beampower [26]. As proven in Section III, the proposed method can estimate up to $Q_{max} = 21$ sources for $N = 4$, which is the highest number of detectable sources among the known techniques. To validate the method for various SH orders, the RMSEs for the proposed method were calculated for SH expansion orders (N) increasing from 1 to 8, and the number of sound sources was varied from 1 to 81. The RMSEs are displayed as a 2D gray-scale image in Fig. 9 such that the image becomes brighter as the RMSE increases. It can be seen that the proposed method can estimate up to $\lfloor N^2 + 4N/3 \rfloor$ with negligible amount of errors. These results show that the number of detectable sources using the proposed EB-ESPRIT doubles for the first-order Ambisonics signal ($N = 1$), and it increases from 9 to 13 in the case of third-order Ambisonics ($N = 3$), which is currently popular in studies using spherical arrays.

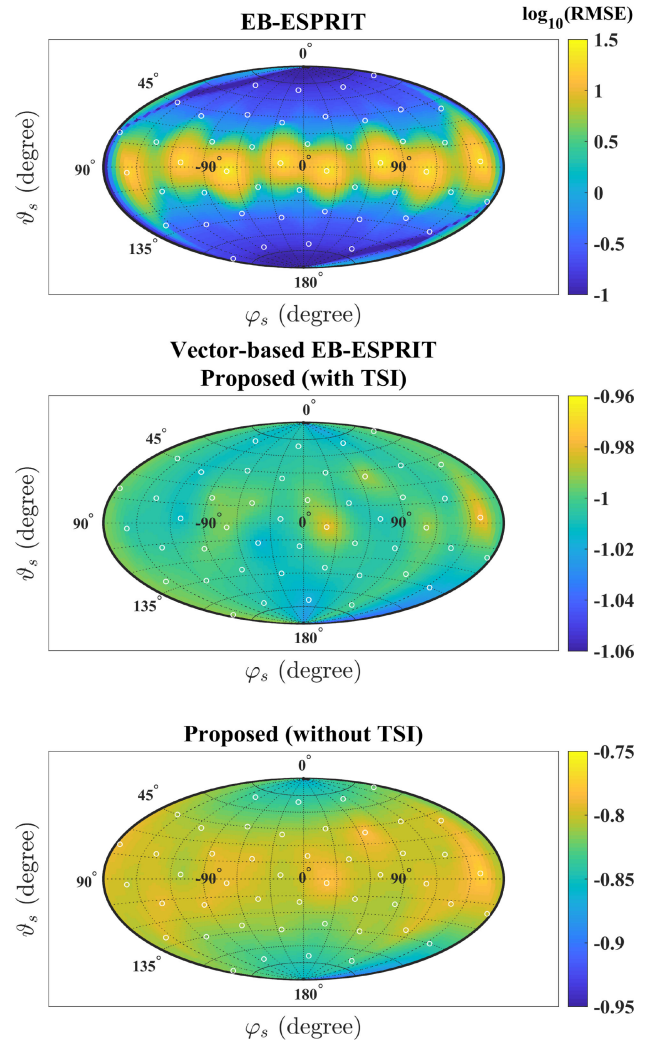


Fig. 7. Distribution of RMSEs for various source DoAs in a free-field condition (without spatial aliasing, SNR = 0 dB). (Top) EB-ESPRIT, (middle) vector-based EB-ESPRIT and proposed method with TSI, and (bottom) proposed method without TSI. Note that color range is much wider for figure of EB-ESPRIT.

C. Room Simulation: Setup

To evaluate the estimation performance for nonstationary source signals in a reflective environment, we simulated two speech sources in a reverberant room. The room impulse responses (RIRs) were generated using a spherical microphone array impulse response generator [56] based on the image source method [57]. The virtual spherical microphone array with 32 microphones uniformly distributed over a rigid sphere of 7 cm radius was positioned at [4.103 m, 3.471 m, 2.912 m] in a simulated room of size $8 \times 7 \times 6 \text{ m}^3$. The reverberation time of the room was set to $T_{60} = 0.3 \text{ s}$ and 0.6 s by changing the absorption coefficient of all walls. Sixteen speech signals with a sampling rate 16 kHz were randomly selected from the ASR corpus (5 seconds 8 male and 8 female English speech files) [58]. Eight sets of two source position pairs were randomly selected from 48 spherical t-design directions [53]. The distance between each source and the center of the spherical microphone

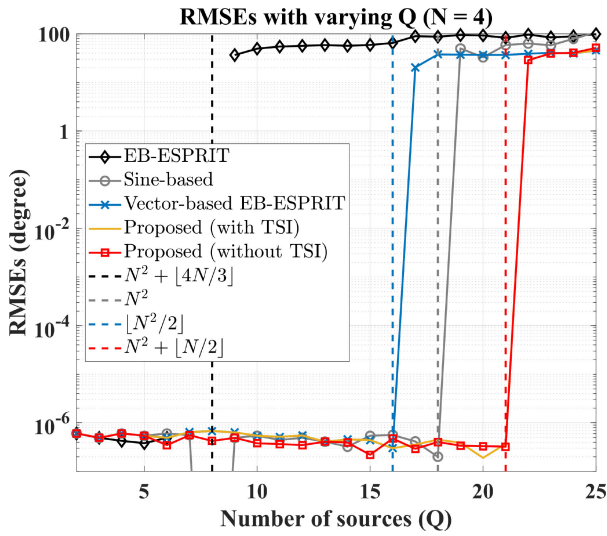


Fig. 8. RMSEs for varying the number of sources in a free-field condition without self-microphone noises and spatial aliasing ($N = 4$). Note that the truncated lines of EB-ESPRIT indicate zero RMSE.

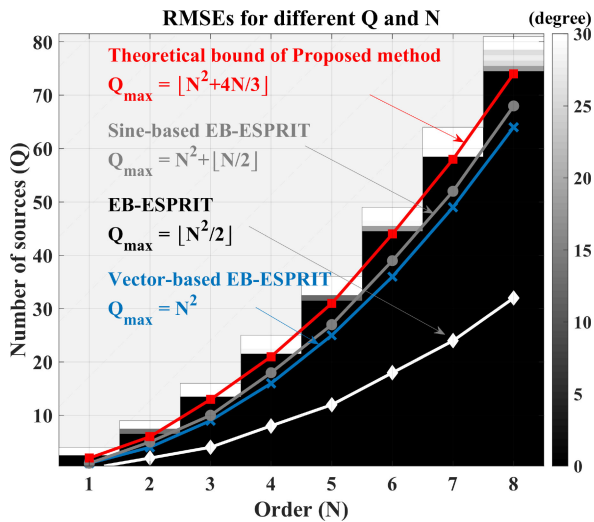


Fig. 9. RMSEs for various SH orders (N) and number of sources (Q) in a free-field condition without self-microphone noises and spatial aliasing. Note that the brightness of each bin indicates the magnitude of RMSE value for corresponding N and Q .

array was 2 m. The virtual microphone signals were synthesized by convolving simulated RIRs with speech signals and adding microphone self-noises of SNR = 10 dB. The time-frequency (TF) analysis of microphone signals was carried out by applying the Short-time Fourier transform (STFT). For STFT, a Hann window of 512 samples was used, which was shifted with 50 % overlap for each time bin, and 1024-points FFT was applied. Then the spherical Fourier transform, followed by the mode-strength compensation, is applied to calculate SH coefficients of each time-frequency bin. The mode-strength compensation filter was designed from the regularized least-squares solution of the theoretical SH model ($N = 3$), whose regularization parameter

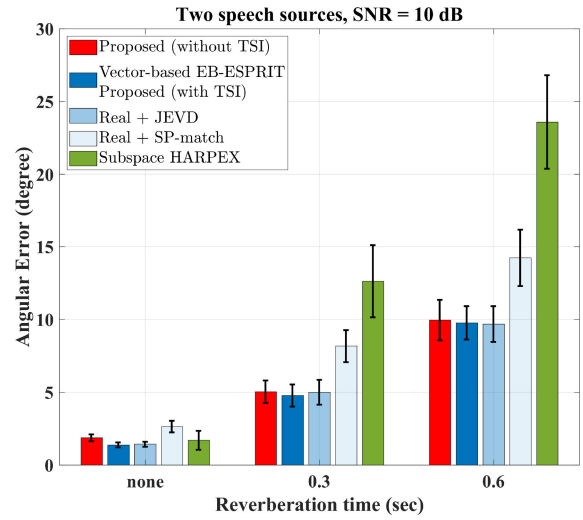


Fig. 10. Mean angular estimation errors of two speech sources for three different reverberation times ($T_{60} = 0$ s, 0.3 s, 0.6 s).

for each frequency was determined such that the maximum allowed amplification is 30 dB [59].

In this simulation, the proposed method was compared with the state-of-the-art methods: the joint diagonalization with real-valued SH [35], SP-match technique [35], robust B-format DoA estimator [48], and the vector-based EB-ESPRIT [31]. The comparison with the SP-match ('Real + SP-match') and robust B-format DoA estimator ('Subspace HARPEX') is meaningful in that they adopt the HARPEX-like estimation technique combined with the signal subspace filtering, although they are applicable only to two sources. The comparison with other methods (EB-ESPRIT and Sine-based) that were inferior in the free-field simulation were not considered here. The vector-based EB-ESPRIT using real-valued spherical harmonics ('Real + JEVD') was included as a comparison case for its low error performance reported in [35]. To handle the non-stationary speech signals, covariance matrices were updated online with the forgetting factor of 0.9. DoAs were estimated in the frequency range [766, 2328] Hz, and estimation errors ($\Delta\Omega_q(t, f)$) were calculated for each time-frequency bin. The mean angular error across TF bins is defined as

$$\overline{\Delta\Omega} = \frac{1}{2n_{\mathcal{S}_{TF}}} \sum_{q=1}^2 \sum_{t, f \in \mathcal{S}_{TF}} \Delta\Omega_q(t, f), \quad (46)$$

where \mathcal{S}_{TF} is a set of TF bins whose energy is higher than -30 dB from the maximum energy, and $n_{\mathcal{S}_{TF}}$ is the cardinality of \mathcal{S}_{TF} .

D. Room Simulation: Results

The mean angular errors are shown in Fig. 10. The result for each method is the average with respect to 8 different configurations (different positions of sources and speech signals). The standard deviations are presented as black bars. As a reference room, the anechoic case without any reflection is also shown ('none' in Fig. 10).

Without the reverberation case, mean angular errors of five methods except SP-match are indistinguishable. The relatively large errors compared to the stationary random noise case (Fig. 3) is suspected due to the time-varying and wide-band properties of speech signals. The errors of ‘Real + SP-match’ are larger than those of the other methods since the estimation of real SHs manifold matrix is less accurate in exchange for acquiring fewer computations [35]. For the reverberation time $T_{60} = 0.3$ s case, errors of the SP-match and robust B-format DoA estimator are higher than others. The highest mean angular errors of robust B-format DoA estimator come from the use of low order SH coefficients only. The mean angular errors of the other three methods are still similar. However, for the reverberation time $T_{60} = 0.6$ s case, ‘Real + JEVD’ technique of [35] yields less errors than complex-valued versions (proposed methods and vector-based EB-ESPRIT). This result can be elaborated that only early and late reflections contribute to the imaginary parts of SH coefficients so that methods using complex-valued covariance matrix are more vulnerable to the effect of reflections as discussed in [35]. Nonetheless, the proposed method can be easily converted to a real version, and even the present form has comparable estimation performance with the conventional methods for speech sound sources in a reverberant environment.

V. CONCLUSION

We proposed a vector-based EB-ESPRIT method whose three different recurrence relations were extended by inter-relations of their highest-order terms. Various numerical simulation results showed that the method has an estimation performance comparable with the state-of-the-art EB-ESPRIT technique [30], [31] and is free from any singularity or ambiguity problem. The proposed method is able to detect a larger number of sources $N^2 + N + \lfloor N/3 \rfloor$ than any other existing EB-ESPRIT technique. This is accomplished by the extension that increases the rank of the EB-ESPRIT matrix with the information from the highest-order coefficient signals. Moreover, the two-step solution scheme that we proposed offers to only make use of the extension if needed, at relatively small computational extra effort. Owing to its high estimation accuracy and large number of detectable sources, the proposed method has high potential to integrate well in acoustic measurement and parametric spatial audio processing applications [3], [37], [60], [61] even with first-order Ambisonics (FOA) signals.

REFERENCES

- [1] M. Wölfel and J. W. McDonough, *Distant Speech Recognition*. Hoboken, NJ, USA: Wiley, 2009.
- [2] D. Yu and L. Deng, *Automatic Speech Recognition*. Berlin, Germany: Springer, 2016.
- [3] A. Politis, S. Tervo, and V. Pulkki, “COMPASS: Coding and multidirectional parameterization of Ambisonic sound scenes,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2018, pp. 6802–6806.
- [4] E. Mabande, K. Kowalczyk, H. Sun, and W. Kellermann, “Room geometry inference based on spherical microphone array eigenbeam processing,” *J. Acoust. Soc. Amer.*, vol. 134, no. 4, pp. 2773–2789, Oct. 2013.
- [5] S. Tervo and A. Politis, “Direction of arrival estimation of reflections from room impulse responses using a spherical microphone array,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 10, pp. 1539–1551, Jun. 2015.
- [6] L. Remaggi, P. J. Jackson, P. Coleman, and W. Wang, “Acoustic reflector localization: Novel image source reversion and direct localization methods,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 2, pp. 296–309, Dec. 2016.
- [7] V. Krishnaveni, T. Kesavamurthy, and B. Aparna, “Beamforming for direction-of-arrival (DOA) estimation—a survey,” *Int. J. Comput. Appl.*, vol. 61, no. 11, pp. 4–11, Jan. 2013.
- [8] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Hoboken, NJ, USA: Wiley, 2004.
- [9] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.
- [10] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Trans. Antennas. Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
- [11] B. D. Rao and K. S. Hari, “Performance analysis of root-MUSIC,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 37, no. 12, pp. 1939–1949, Dec. 1989.
- [12] R. Roy and T. Kailath, “ESPRIT-estimation of signal parameters via rotational invariance techniques,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
- [13] D. P. Jarrett, E. A. Habets, and P. A. Naylor, “3D source localization in the spherical harmonic domain using a pseudointensity vector,” in *Proc. Eur. Signal Process. Conf.*, Aalborg, Denmark, Aug. 2010, pp. 442–446.
- [14] A. H. Moore, C. Evers, and P. A. Naylor, “Direction of arrival estimation in the spherical harmonic domain using subspace pseudointensity vectors,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 1, pp. 178–192, Sep. 2016.
- [15] O. Nadiri and B. Rafaely, “Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 10, pp. 1494–1505, Jul. 2014.
- [16] L. Madmoni and B. Rafaely, “Direction of arrival estimation for reverberant speech based on enhanced decomposition of the direct sound,” *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 1, pp. 131–142, Dec. 2018.
- [17] S. Hafezi, A. H. Moore, and P. A. Naylor, “3D acoustic source localization in the spherical harmonic domain based on optimized grid search,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, China, Mar. 2016, pp. 415–419.
- [18] S. Hafezi, A. H. Moore, and P. A. Naylor, “Multiple source localization in the spherical harmonic domain using augmented intensity vectors based on grid search,” in *Proc. Eur. Signal Process. Conf.*, Budapest, Hungary, Aug. 2016, pp. 602–606.
- [19] S. Hafezi, A. H. Moore, and P. A. Naylor, “Augmented intensity vectors for direction of arrival estimation in the spherical harmonic domain,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 10, pp. 1956–1968, Aug. 2017.
- [20] M. Berzborn and M. Vorländer, “Investigations on the directional energy decay curves in reverberation rooms,” in *Proc. Euronoise*, Heraklion, Greece, May 2018, pp. 2005–2010.
- [21] B. Alary, P. Massé, V. Välimäki, and M. Noisternig, “Assessing the anisotropic features of spatial impulse responses,” in *Proc. EAA Spatial Audio Signal Process. Symp.*, Paris, France, Sep. 2019, pp. 43–48.
- [22] A. Politis and V. Pulkki, *Higher-Order Directional Audio Coding*. Hoboken, NJ, USA: Wiley, 2017, pp. 141–159.
- [23] L. McCormack, A. Politis, O. Scheuregger, and V. Pulkki, “Higher-order processing of spatial impulse responses,” in *Proc. 23rd Int. Congr. Acoust.*, Aachen, Germany, Sep. 2019, pp. 4909–4916.
- [24] H. Teutsch and W. Kellermann, “Eigen-beam processing for direction-of-arrival estimation using spherical apertures,” in *Proc. Joint Workshop Hands-Free Speech Commun. Microphone Arrays*, Mar. 2005, pp. c13–c14.
- [25] H. Teutsch and W. Kellermann, “Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2008, pp. 5276–5279.
- [26] X. Li, S. Yan, X. Ma, and C. Hou, “Spherical harmonics MUSIC versus conventional MUSIC,” *Appl. Acoust.*, vol. 72, no. 9, pp. 646–652, Sep. 2011.
- [27] N. Epain and C. T. Jin, “Spherical harmonic signal covariance and sound field diffuseness,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 10, pp. 1796–1807, Jun. 2016.
- [28] B. Jo and J.-W. Choi, “Direction of arrival estimation using nonsingular spherical ESPRIT,” *J. Acoust. Soc. Amer.*, vol. 143, no. 3, pp. EL181–EL187, Mar. 2018.

- [29] B. Jo and J.-W. Choi, "Sine-based EB-ESPRIT for source localization," in *Proc. IEEE Sensor Array Multichan. Signal Process. Workshop*, Sheffield, U.K., Jul. 2018, pp. 326–330.
- [30] B. Jo and J.-W. Choi, "Parametric direction-of-arrival estimation with three recurrence relations of spherical harmonics," *J. Acoust. Soc. Amer.*, vol. 145, no. 1, pp. 480–488, Jan. 2019.
- [31] A. Herzog and E. Habets, "Eigenbeam-ESPRIT for DOA-vector estimation," *IEEE Signal Process. Lett.*, vol. 26, no. 4, pp. 572–576, Feb. 2019.
- [32] B. Jo and J.-W. Choi, "Robust localization of early reflections in a room using semi real-valued EB-ESPRIT with three recurrence relations and laplacian constraint," in *Proc. 23rd Int. Congr. Acoust.*, Aachen, Germany, Sep. 2019, pp. 4949–4956.
- [33] H. Sun, H. Teutsch, E. Mabande, and W. Kellermann, "Robust localization of multiple sources in reverberant environments using EB-ESPRIT with spherical microphone arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Prague, Czechia, May 2011, pp. 117–120.
- [34] Q. Huang, L. Zhang, and F. Yong, "Two-step spherical harmonics ESPRIT-type algorithms and performance analysis," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 9, pp. 1684–1697, Sep. 2018.
- [35] A. Herzog and E. Habets, "Online DOA estimation using real eigenbeam ESPRIT with propagation vector matching," in *Proc. EAA Spatial Audio Signal Process. Symp.*, Paris, France, Sep. 2019, pp. 19–24.
- [36] A. Herzog and E. Habets, "On the relation between DOA-Vector Eigenbeam ESPRIT and subspace pseudointensity-vector," in *Proc. Eur. Signal Process. Conf.*, Corua, Spain, Sep. 2019, pp. 1–5.
- [37] N. Barrett and S. Berge, "A new method for B-format to binaural transcoding," in *Proc. Audio Eng. Soc. Conf.*, Oct. 2010, Paper 6-5.
- [38] D. Jarrett, E. Habets, and P. Naylor, *Theory and Applications of Spherical Microphone Array Processing*. New York, NY, USA: Springer, 2017.
- [39] B. Rafaely, *Fundamentals of Spherical Array Processing*. New York, NY, USA: Springer, 2015.
- [40] B. Jo and J.-W. Choi, "Spherical harmonic smoothing for localizing coherent sound sources," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 10, pp. 1969–1984, Aug. 2017.
- [41] N. Gumerov and R. Duraiswami, *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*. College Park, MD, USA: Elsevier, 2004.
- [42] C. Choi, J. Ivanic, M. Gordon, and K. Ruedenberg, "Rapid and stable determination of rotation matrices between spherical harmonics by direct recursion," *J. Chem. Phys.*, vol. 111, no. 19, pp. 8825–8831, Nov. 1999.
- [43] T. Fu and X. Gao, "Simultaneous diagonalization with similarity transformation for non-defective matrices," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Toulouse, France, May 2006, vol. 4, pp. IV-1137–IV-1140.
- [44] X. Luciani and L. Albera, "Joint eigenvalue decomposition using polar matrix factorization," in *Proc. Int. Conf. Latent Variable Anal. Signal Separation*, Sep. 2010, pp. 555–562.
- [45] R. André, X. Luciani, and E. Moreau, "A coupled joint eigenvalue decomposition algorithm for canonical polyadic decomposition of tensors," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop*, Rio de Janeiro, Brazil, Jul. 2016, pp. 1–5.
- [46] X. Luciani and L. Albera, "Joint eigenvalue decomposition of non-defective matrices based on the LU factorization with application to ICA," *IEEE Signal Process.*, vol. 63, no. 17, pp. 4594–4608, Jun. 2015.
- [47] G. Strang, *Introduction to Linear Algebra*. Wellesley, MA, USA: Wellesley Cambridge Press, 1993.
- [48] O. Thiergart and E. A. P. Habets, "Robust direction-of-arrival estimation of two simultaneous plane waves from a B-format signal," in *Proc. IEEE Conv. Elect. Electron. Eng. Isr.*, Eilat, Israel, Nov. 2012, pp. 1–5.
- [49] D. A. Linebarger, R. D. DeGroat, and E. M. Dowling, "Efficient direction-finding methods employing forward/backward averaging," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2136–2145, Aug. 1994.
- [50] H. Gene and F. V. L. Golub, Charles, *Matrix Computations*, vol. 3. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996, vol. 3.
- [51] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 95–107, Jan. 1995.
- [52] S. Boyd and L. Vandenberghe, *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*. Cambridge, U.K.: Cambridge Univ. Press, 2018.
- [53] R. Hardin and N. Sloane, "Mclarens improved snub cube and other new spherical designs in three dimensions," *Discrete Comput. Geometry*, vol. 15, no. 4, pp. 429–441, Apr. 1996.
- [54] N. Huleihel and B. Rafaely, "Spherical array processing for acoustic analysis using room impulse responses and time-domain smoothing," *J. Acoust. Soc. Amer.*, vol. 133, no. 6, pp. 3995–4007, Jun. 2013.
- [55] L. Kumar and R. Hegde, "Stochastic Cramer-Rao bound analysis for estimation in spherical harmonics domain," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1030–1034, Dec. 2014.
- [56] D. Jarrett, E. Habets, M. Thomas, and P. Naylor, "Rigid sphere room impulse response simulation: Algorithm and applications," *J. Acoust. Soc. Amer.*, vol. 132, no. 3, pp. 1462–1472, Sep. 2012.
- [57] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.
- [58] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2015, pp. 5206–5210.
- [59] C. T. Jin, N. Epain, and A. Parthy, "Design, optimization and evaluation of a dual-radius spherical microphone array," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 1, pp. 193–204, Oct. 2013.
- [60] V. Pulkki, S. Delikaris-Manias, and A. Politis, *Parametric Time-Frequency Domain Spatial Audio*. Hoboken, NJ, USA: Wiley, 2018.
- [61] V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503–516, Jun. 2007.



Byeongho Jo (Student Member, IEEE) received the B.S. degree in electronic and electrical engineering from Sungkyunkwan University, Suwon, South Korea, in 2015. He received the M.S. degree in electrical engineering (EE) from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2018. He is currently working toward the Ph.D. degree in EE at KAIST, Daejeon, South Korea. Since 2015, he has been in Smart Sound Systems Laboratory, developing spherical microphone array systems for estimating the direction of arrival. His

research interests include spherical microphone array signal processing, acoustic signal processing, and beamforming.



Franz Zotter received the M.Sc. degree in electrical and audio engineering from the Graz University of Technology in 2004, and in 2009 he received the Ph.D. degree in natural science from the University of Music and Performing Arts in Graz, Austria. He joined the Institute of Electronic Music an Acoustics (IEM) in 2004 as Research Assistant, became Senior Scientist in 2008, and Assistant Professor (tenure track) in 2019. He authored the book *Ambisonics* (Springer, 2019) with M. Frank, and his interests include spherical array signal processing and applications in music and virtual reality. Ass.Prof. Zotter is a member of the German Acoustical Society (DEGA), is its current TC Virtual Acoustics Chair and was awarded DEGA's Lothar Cremer medal in 2012. He is also member of the German Tonmeister Society (VDT) and the Audio Engineering Society (AES).



Jung-Woo Choi (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in mechanical engineering from Korea Institute of Science and Technology (KAIST), South Korea, in 1999, 2001, 2005, respectively. He was a Postdoctoral Research Associate with the Center for Noise and Vibration Control (NOVIC) at KAIST from 2005 to 2006. From 2006 to 2007, he was a Postdoctoral Researcher at the Institute of Sound and Vibration Research (ISVR), University of Southampton, U.K. From 2007 to 2011, he was with Samsung Electronics at the Samsung Advanced

Institute of Technology (SAIT) in Korea. He was a Research Associate Professor in the Department of Mechanical Engineering, KAIST, until 2014. In 2015, he joined the Department of Electrical Engineering, KAIST, as Assistant Professor and became Associate Professor in 2018. He is the coauthor of the book *Sound Visualization and Manipulation* (Wiley, 2013). His current research interests include sound field reproduction, sound focusing, array signal processing, and their applications. Prof. Choi is a member of Acoustical Society of America, Institute of Noise Control Engineers-USA, Acoustical Society of Korea, Korean Society of Noise and Vibration Engineering.