PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Robust localization of early reflections in a room using semi real-valued EB-ESPRIT with three recurrence relations and Laplacian constraint

Byeongho JO[1]; Jung-Woo CHOI[2]

[1, 2] Korea Advanced Institute of Science and Technology, Republic of Korea

ABSTRACT

To localize early reflections in a reverberant room, the direction-of-arrival (DOA) estimation technique has been evolved in various ways. Among many techniques, the subspace-based localization techniques such as eigenbeam estimation of signal parameters via rotational invariance techniques (EB-ESPRIT) have been extensively researched for their ability to directly estimate DOA parameters from spherical microphone recordings. The EB-ESPRIT estimates DOAs by computing the directional parameters from a single recurrence relation of spherical harmonics. For estimating DOAs of early reflections, however, the EB-ESPRIT is inappropriate due to the insufficient number of detectable echoes and its vulnerability to noise. In this study, we attempt to overcome these issues by using semi real-valued processing of spherical harmonics and extra constraints. The proposed technique can compute directional parameters more robustly by deriving the real-valued forms of EB-ESPRIT equations and by imposing additional Laplacian constraint. We show that the real-valued formula solved by the joint eigenvalue decomposition (JEVD) algorithm can enhance the accuracy of DOA estimation as compared to all existing EB-ESPRIT based techniques. In simulations, we validate the robustness of the proposed technique in terms of the estimation error under different signal-to-noise ratios (SNRs).

Keywords: Three Recurrence relations, Spherical microphone array, Early reflections

1. INTRODUCTION

There have been many studies to estimate the directions-of-arrival (DOAs) or localize the early reflections of a room using a spherical microphone array (SMA), which is applicable to the room geometry inference (1), robust adaptive beamforming (2), de-reverberation (3), and sound field reproduction with compensation of the reverberant components (4). Mabande *et al*. proposed the room geometry inference method using the eigenbeam beamforming techniques with time differences of arrival (TDOAs) estimated by the cross-correlation analysis (1, 5). Morgenstern *et al*. proposed to estimate DOAs of reflections using the de-correlation technique with a loudspeaker array and a microphone array (6, 7). Tervo *et al*. introduced the maximum-likelihood (ML) method based on the stochastic model to estimate the coherent early reflections (8).

These studies have limitations that the exhaustive grid-search of peaks should be accompanied from the computed beamforming map. In contrast, the eigenbeam estimation of signal parameters via rotational invariant techniques (EB-ESPRIT) has the advantage that it directly estimates DOAs in a parametric way (9). However, the limited number of detectable sound sources and the possibility of matrix ill-conditioning problems due to reflections impinging from the equator direction hinder the use of EB-ESPRIT in highly reverberant conditions (10, 11). For these reasons, many alternative types of EB-ESPRIT techniques have been proposed to increase the number of detectable sources by using the different recurrence relations or to resolve the matrix ill-conditioning problem (12-15).

In this work, we introduce the novel EB-ESPRIT technique that combines three recurrence relations and the semi real-valued processing. The semi real-valued processing is known to improve

[1] byeongho@kaist.ac.kr
[2] jwoo@kaist.ac.kr

the estimation performance albeit with the reduced computational complexity (16-19). In addition, we estimate the directional parameter matrices by using iterative joint eigenvalue decomposition (JEVD) method (20-22) with the additional constraint based on the Helmholtz equation. By using the additional constraint, the diagonalization of JVED is restricted by a physical constraint, which reduces the effect of the noises. The estimation performance of the proposed method is validated by the simulated spherical microphone array data installed in a box-shaped room under various noise conditions.

2. SPHERICAL HARMONIC DOMAIN MODEL

2.1 Complex-valued data model

The 2D fields $f(\Omega)$ defined on a unit sphere ($\mathcal{S}^2$) can be converted into the spherical harmonic domain (SHD) through the spherical Fourier transform (SFT) as (23)

$$f_{nm} \coloneqq \int_{\mathcal{S}^2} f(\Omega) Y_n^m(\Omega)^* d\Omega, \tag{1}$$

where $\Omega \in \{(\theta,\phi) \mid \theta \in [0,\pi], \phi \in [0,2\pi]\}$ denotes 2D angles in the spherical coordinates and $(\cdot)^*$ denotes the complex conjugate. Spherical harmonics is defined as $Y_n^m(\Omega) = \mathcal{N}_{nm} \mathcal{P}_n^m(\cos\theta) e^{im\phi}$, where $\mathcal{N}_{nm} = \{(2n+1)(n-m)!/4\pi(n+m)!\}^{1/2}$ and $\mathcal{P}_n^m(\cdot)$ is the associated Legendre function. For a unit amplitude plane-wave with DOA Ω_{PW}, the SHD signal is transformed from the spherical microphone array (SMA) recording data as

$$f_{PW,nm} = b_n(kr) Y_n^m(\Omega_{PW})^*, \tag{2}$$

where k denotes the wavenumber, r is the radius of the spherical microphone array and $b_n(kr)$ is a mode-strength that depends on the radius and the boundary condition of the array surface. To get rid of this dependency, the SHD signal is divided by the $b_n(kr)$ of the respective orders ($n = 1, \cdots, N$). After compensation, the mode-compensated SHD (MC-SHD) signal of a unit amplitude plane-wave can be written as $a_{PW,nm} = Y_n^m(\Omega_{PW})^*$ (24).

For D plane waves, the MC-SHD signal up to the truncated harmonics order N can be written as the vector form:

$$\mathbf{a} = \mathbf{Y}^H \mathbf{s}, \tag{3}$$

where $\mathbf{a} = [a_{0,0}, a_{1,-1}, a_{1,0}, a_{1,1}, \cdots, a_{N,N}]^T$, $\mathbf{s} = [s_1, \cdots, s_d, \cdots, s_D]^T$, s_d is the complex amplitude of d-th source, and $\mathbf{Y}^H$ is the steering matrix in SHD whose d-th column is given by $\mathbf{y}(\Omega_d) = [Y_0^0(\Omega_d), Y_1^{-1}(\Omega_d), Y_1^0(\Omega_d), Y_1^1(\Omega_d), \cdots, Y_N^N(\Omega_d)]^H$. $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and conjugate transpose, respectively. We assume that all MC-SHD signals are made by the short-time Fourier transform (STFT), SFT, and mode-compensation in order. To simplify the notations, time frame and frequency indices of MC-SHD signals are omitted. For EB-ESPRIT type algorithms, the covariance matrix should be constructed for extracting the signal subspaces as follows,

$$\mathbf{R}_c = E\{\mathbf{a}\mathbf{a}^H\} = \mathbf{Y}^H E\{\mathbf{s}\mathbf{s}^H\}\mathbf{Y} = \mathbf{U}_s \mathbf{\Sigma}_s \mathbf{U}_s{}^H, \tag{4}$$

where $\mathbf{U}_s$ and $\mathbf{\Sigma}_s$ are the signal subspace and the corresponding eigenvalue matrix, respectively. $E\{\cdot\}$ denotes the expectation operator. The subscript c of $\mathbf{R}_c$ represents the complex-valued matrix, which is to distinguish the complex-valued covariance matrix from the real-valued covariance matrix introduced in the following section.

2.2 Real-valued data model

For the unitary or semi real-valued processing (12, 18), the real-valued covariance matrix should be constructed. First, the real spherical harmonics is defined as follows:

$$\hat{Y}_n^m(\Omega) = \begin{cases} \sqrt{2}\mathcal{N}_{n|m|}\mathcal{P}_n^{|m|}(\cos\theta)\sin(|m|\phi), & m < 0 \\ \mathcal{N}_{nm}\mathcal{P}_n^m(\cos\theta), & m = 0 \\ \sqrt{2}\mathcal{N}_{nm}\mathcal{P}_n^m(\cos\theta)\cos(m\phi), & m > 0 \end{cases}. \tag{5}$$

Using the complex symmetry property of the spherical harmonics and Equation (5), the complex spherical harmonic matrix can be transformed by a following $(2n+1)\times(2n+1)$ unitary matrix for order $n \geq 1$:

$$\mathbf{Q}_n = \frac{1}{\sqrt{2}} \begin{bmatrix} i\mathbf{I}_n & \mathbf{0}_n & (-1)^2 i\mathbf{J}_n \\ \mathbf{0}_n^T & \sqrt{2} & \mathbf{0}_n^T \\ \mathbf{G}_n & \mathbf{0}_n & \mathbf{J}_n\mathbf{G}_n \end{bmatrix}, \tag{6}$$

where $\mathbf{Q}_0 = 1$, and $\mathbf{G}_n$ is a $n \times n$ anti-diagonal matrix with ones in anti-diagonal positions and zeros elsewhere, $\mathbf{0}_n$ represents a column vector containing of n zeros, and $\mathbf{J}_n$ is a $n \times n$ anti-diagonal matrix whose $(i, n+1-i)$-th element is set to $(-1)^i$. With $\mathbf{Q}_n$ matrices ($n = 0, \cdots, N$), one can construct a unitary matrix $\mathbf{Q} = \text{blkdiag}\{\mathbf{Q}_0, \cdots, \mathbf{Q}_N\}$. As shown in (12), a real-valued steering matrix can be obtained as $\mathbf{Q}\mathbf{Y}^H = \hat{\mathbf{Y}}^T = [\hat{\mathbf{y}}_1, \cdots, \hat{\mathbf{y}}_D]$, where $\hat{\mathbf{y}} = [\hat{Y}_0^0(\theta,\phi), \hat{Y}_1^{-1}(\theta,\phi), \hat{Y}_1^0(\theta,\phi), \cdots, \hat{Y}_N^N(\theta,\phi)]^T$.

Based on the proof in (12), the real-valued covariance matrix can be computed from the complex-valued covariance matrix as

$$\mathbf{R}_r = \text{Re}\{\mathbf{Q}\mathbf{R}_c\mathbf{Q}^H\} = \hat{\mathbf{Y}}^T E\{\hat{\mathbf{s}}\hat{\mathbf{s}}^H\}\hat{\mathbf{Y}} = \hat{\mathbf{U}}_s\hat{\mathbf{\Sigma}}_s\hat{\mathbf{U}}_s^H \tag{7}$$

where $\hat{\mathbf{s}} = [\text{Re}\{\mathbf{s}\}, \text{Im}\{\mathbf{s}\}]$, $\hat{\mathbf{U}}_s$ is the real-valued signal subspace, $\hat{\mathbf{\Sigma}}_s$ is the real-valued eigenvalue matrix, and $\text{Re}\{\cdot\}$ extracts the real part of the argument. In the following section, the real-valued signal subspace in Equation (7) is utilized for the recurrence relations of complex spherical harmonics.

3. SEMI REAL-VALUED EB-ESPRIT

The proposed semi real-valued (SR)-EB-ESPRIT utilizes three recurrence relations of spherical harmonics whose directional parameters correspond to x, y, and z directions in the Cartesian coordinates (25, 26):

$$\begin{aligned} \sin\theta_d\cos\phi_d Y_n^m(\Omega_d)^* &= (w_{n+1}^{m-1}Y_{n+1}^{m-1}(\Omega_d)^* - w_n^{-m}Y_{n-1}^{m-1}(\Omega_d)^* - w_{n+1}^{-(m+1)}Y_{n+1}^{m+1}(\Omega_d)^* + w_n^m Y_{n-1}^{m+1}(\Omega_d)^*)/2 \\ \sin\theta_d\sin\phi_d Y_n^m(\Omega_d)^* &= (w_{n+1}^{m-1}Y_{n+1}^{m-1}(\Omega_d)^* - w_n^{-m}Y_{n-1}^{m-1}(\Omega_d)^* + w_{n+1}^{-(m+1)}Y_{n+1}^{m+1}(\Omega_d)^* - w_n^m Y_{n-1}^{m+1}(\Omega_d)^*)/2i, \\ \cos\theta_d Y_n^m(\Omega_d)^* &= v_n^m Y_{n-1}^m(\Omega_d)^* + v_{n+1}^m Y_{n+1}^m(\Omega_d)^* \end{aligned} \tag{8}$$

where $w_n^m = \sqrt{(n-m-1)(n-m)/(2n-1)(2n+1)}$ and $v_n^m = \sqrt{(n-m)(n+m)/(2n-1)(2n+1)}$.

The above three recurrence relations can be re-written as the matrix form with the shifted and order-reduced ($N-1$) harmonics matrix $\mathbf{Y}^{(\mu,\nu)}$ as

$$\begin{aligned} \mathbf{Y}^{(0,0)H}\mathbf{\Phi}^{(1)} &= (\mathbf{W}_{n+1}^{m-1}\mathbf{Y}^{(+1,-1)H} - \mathbf{W}_n^{-m}\mathbf{Y}^{(-1,-1)H} - \mathbf{W}_{n+1}^{-(m+1)}\mathbf{Y}^{(+1,+1)H} + \mathbf{W}_n^m\mathbf{Y}^{(-1,+1)H})/2 \\ \mathbf{Y}^{(0,0)H}\mathbf{\Phi}^{(2)} &= (\mathbf{W}_{n+1}^{m-1}\mathbf{Y}^{(+1,-1)H} - \mathbf{W}_n^{-m}\mathbf{Y}^{(-1,-1)H} + \mathbf{W}_{n+1}^{-(m+1)}\mathbf{Y}^{(+1,+1)H} - \mathbf{W}_n^m\mathbf{Y}^{(-1,+1)H})/2i, \\ \mathbf{Y}^{(0,0)H}\mathbf{\Phi}^{(3)} &= \mathbf{V}_n^m\mathbf{Y}^{(-1,0)H} + \mathbf{V}_{n+1}^m\mathbf{Y}^{(+1,0)H} \end{aligned} \tag{9}$$

where $\mathbf{\Phi}^{(1)}$, $\mathbf{\Phi}^{(2)}$, and $\mathbf{\Phi}^{(3)}$ are $D \times D$ real-valued diagonal matrices containing directional parameters $\sin\theta_d\cos\phi_d$, $\sin\theta_d\sin\phi_d$, and $\cos\theta_d$, respectively. $\mathbf{W}_{n+\mu}^{m+\nu}$ and $\mathbf{V}_{n+\mu}^{m+\nu}$ are $N^2 \times N^2$ real-valued diagonal matrices containing $w_{n+\mu}^{m+\nu}$ and $v_{n+\mu}^{m+\nu}$, respectively. The d-th column of $\mathbf{Y}^{(\mu,\nu)H}$ is defined as $\mathbf{y}_\mu^\nu(\Omega_d) = [Y_{0+\mu}^{0+\nu}(\Omega_d), Y_{1+\mu}^{-1+\nu}(\Omega_d), Y_{1+\mu}^{0+\nu}(\Omega_d), Y_{1+\mu}^{1+\nu}(\Omega_d), \cdots, Y_{(N-1)+\mu}^{(N-1)+\nu}(\Omega_d)]^H$.

As the directional parameter matrices ($\mathbf{\Phi}^{(1)}$, $\mathbf{\Phi}^{(2)}$, and $\mathbf{\Phi}^{(3)}$) contain DOAs (Ω_d), one has to solve the Equation (9) for the matrices ($\mathbf{\Phi}^{(1)}$, $\mathbf{\Phi}^{(2)}$, and $\mathbf{\Phi}^{(3)}$). However, $\mathbf{Y}^{(\mu,\nu)}$ matrices cannot be computed directly from the MC-SHD signals. We utilize a relationship between the real-valued signal subspace ($\hat{\mathbf{U}}_s$) and the complex-valued steering matrix ($\mathbf{Y}^H$) in SHD (12), given by

$$\underline{\mathbf{U}}_s = \mathbf{Q}^H\hat{\mathbf{U}}_s = \mathbf{Y}^H\hat{\mathbf{T}}, \tag{10}$$

where $\hat{\mathbf{T}}$ is a $D \times D$ real-valued nonsingular matrix. Using Equation (10), we can transform a real-valued signal subspace ($\hat{\mathbf{U}}_s$) obtained from a real-valued covariance matrix ($\mathbf{R}_r$) into a complex-valued signal subspace ($\underline{\mathbf{U}}_s$) that can be applied to recurrence relations of complex spherical harmonics (Equation (8)), and obtain a real-valued common transformation matrix $\hat{\mathbf{T}}$ as well. From Equations (7) and (10), and by multiplying $\hat{\mathbf{T}}$ matrix, Equation (9) can be re-written as

$$\begin{aligned}
\underline{\mathbf{U}}_s^{(0,0)}\boldsymbol{\Psi}^{(1)} &= (\mathbf{W}_{n+1}^{m-1}\underline{\mathbf{U}}_s^{(+1,-1)} - \mathbf{W}_n^{-m}\underline{\mathbf{U}}_s^{(-1,-1)} - \mathbf{W}_{n+1}^{-(m+1)}\underline{\mathbf{U}}_s^{(+1,+1)} + \mathbf{W}_n^{m}\underline{\mathbf{U}}_s^{(-1,+1)})/2 \\
\underline{\mathbf{U}}_s^{(0,0)}\boldsymbol{\Psi}^{(2)} &= (\mathbf{W}_{n+1}^{m-1}\underline{\mathbf{U}}_s^{(+1,-1)} - \mathbf{W}_n^{-m}\underline{\mathbf{U}}_s^{(-1,-1)} + \mathbf{W}_{n+1}^{-(m+1)}\underline{\mathbf{U}}_s^{(+1,+1)} - \mathbf{W}_n^{m}\underline{\mathbf{U}}_s^{(-1,+1)})/2i \\
\underline{\mathbf{U}}_s^{(0,0)}\boldsymbol{\Psi}^{(3)} &= \mathbf{V}_n^{m}\underline{\mathbf{U}}_s^{(-1,0)} + \mathbf{V}_{n+1}^{m}\underline{\mathbf{U}}_s^{(+1,0)}
\end{aligned}, \tag{11}$$

where $\boldsymbol{\Psi}^{(I)} = \hat{\mathbf{T}}^{-1}\boldsymbol{\Phi}^{(I)}\hat{\mathbf{T}}$ $(\forall I \in \{1,2,3\})$. The least-squares solution of Equation (11) yield $\boldsymbol{\Psi}^{(I)}$ matrices. The least-squares solution can be unique when the pseudo-inverse matrix $\underline{\mathbf{U}}_s^{(0,0)}$ is the overdetermined system, which corresponds $N^2 \geq D$. From this inequality condition, the maximum number of detectable sources is determined to N^2.

To estimate simultaneously $\boldsymbol{\Phi}^{(I)}$ matrices from the estimated $\boldsymbol{\Psi}^{(I)}$ matrices, one should compute joint-diagonalizing matrix. In the following section, we show how to compute the joint-diagonalizing matrix $\mathbf{A}$ by using a block coordinate procedure and additional constraints. From the estimated $\mathbf{A}$ matrix that corresponds to the inverse of the real-valued common transformation matrix ($\hat{\mathbf{T}}^{-1}$), the $\hat{\boldsymbol{\Phi}}^{(I)}$ matrices can be computed from $\hat{\boldsymbol{\Phi}}^{(I)} = \mathbf{A}\boldsymbol{\Psi}^{(I)}\mathbf{A}^{-1}$. Using the elements of the $\hat{\boldsymbol{\Phi}}^{(I)}$ matrices, DOA of d-th source (Ω_d) can be estimated with a four-quadrant inverse tangent ($\mathrm{atan}\{\cdot\}$) as

$$\begin{aligned}
\hat{\theta}_d &= \mathrm{atan2}\{\sqrt{\{\mathrm{Re}\{[\hat{\boldsymbol{\Phi}}^{(1)}]_{d,d}\}\}^2 + \{\mathrm{Re}\{[\hat{\boldsymbol{\Phi}}^{(2)}]_{d,d}\}\}^2}, \mathrm{Re}\{[\hat{\boldsymbol{\Phi}}^{(3)}]_{d,d}\}\} \\
\hat{\phi}_d &= \mathrm{atan2}\{\mathrm{Re}\{[\hat{\boldsymbol{\Phi}}^{(2)}]_{d,d}\}, \mathrm{Re}\{[\hat{\boldsymbol{\Phi}}^{(1)}]_{d,d}\}\}
\end{aligned}, \tag{12}$$

4. JOINT DIAGONALIZATION WITH CONSTRAINT

4.1 A block coordinate procedure

As mentioned in (15), $\boldsymbol{\Psi}^{(I)}$ matrices are jointly diagonalized by using the joint eigenvalue decomposition algorithms, which can be achieved by minimizing

$$\mathcal{J}(\mathbf{A}) = \sum_{I=1}^{3}\left\|\mathrm{ZDiag}\left\{\mathbf{A}\boldsymbol{\Psi}^{(I)}\mathbf{A}^{-1}\right\}\right\|_F^2, \tag{13}$$

where $\mathrm{ZDiag}\{\cdot\}$ sets to zero the diagonal of the matrix element, and $\|\cdot\|_F$ denotes the Frobenius norm of the matrix. In this work, we utilize a block coordinate procedure of the JAPAM algorithm (Joint eigenvalue decomposition algorithms using a Parameterized Matrix) (22). The block coordinate procedure divides the problem of estimating $\mathbf{A}$ by minimizing $\mathcal{J}(\mathbf{A})$ into $D(D-1)/2$ successive simpler optimization problems by using the following matrices,

$$\mathbf{A} = \prod_{i=1}^{D-1}\prod_{j=i+1}^{D}\mathbf{A}^{(i,j)}, \tag{14}$$

where matrices $\mathbf{A}^{(i,j)}$ are equal to the identity matrix that depends on four unknown parameters: $A_1^{(i,j)}$, $A_2^{(i,j)}$, $A_3^{(i,j)}$, $A_4^{(i,j)}$:

$$\begin{cases}
\forall (i',j') \notin \{(i,i),(i,j),(j,i),(j,j)\}, [\mathbf{A}^{(i,j)}]_{i',j'} = \delta_{i',j'} \\
[\mathbf{A}^{(i,j)}]_{i,i} = A_1^{(i,j)}, [\mathbf{A}^{(i,j)}]_{i,j} = A_2^{(i,j)} \\
[\mathbf{A}^{(i,j)}]_{j,i} = A_3^{(i,j)}, [\mathbf{A}^{(i,j)}]_{j,j} = A_4^{(i,j)}
\end{cases}, \tag{15}$$

where δ is the Kronecker delta function. For (i,j) step, the following matrices are updated as

$$\begin{cases}
\mathbf{A} \leftarrow \mathbf{A}^{(i,j)}\mathbf{A} \\
\boldsymbol{\Psi}^{(I)} \leftarrow \mathbf{A}^{(i,j)}\boldsymbol{\Psi}^{(I)}(\mathbf{A}^{(i,j)})^{-1}, \ \forall I \in \{1,2,3\}
\end{cases}. \tag{16}$$

In (i,j)-update, we define the following matrices

$$\boldsymbol{\Theta}^{(I)} = \mathbf{A}^{(i,j)}\boldsymbol{\Psi}^{(I)}(\mathbf{A}^{(i,j)})^{-1}, \ \forall I \in \{1,2,3\}. \tag{17}$$

Then, the cost function (Equation (13)) can be re-written as

$$C(\mathbf{A}^{(i,j)}) = \sum_{I=1}^{3}(|[\boldsymbol{\Theta}^{(I)}]_{i,j}|^2 + |[\boldsymbol{\Theta}^{(I)}]_{j,i}|^2). \tag{18}$$

As we focus on $\mathbf{\Theta}_{i,j}^{(I)}$ and $\mathbf{\Theta}_{j,i}^{(I)}$ matrices, Equation (18) can be transformed to the simpler cost function form with 2×2 matrices as

$$C(\tilde{\mathbf{A}}^{(i,j)}) = \sum_{I=1}^{3}(|[\tilde{\mathbf{\Theta}}^{(I)}]_{1,2}|^2 + |[\tilde{\mathbf{\Theta}}^{(I)}]_{2,1}|^2), \tag{19}$$

where

$$\begin{aligned}&\tilde{\mathbf{\Theta}}^{(I),(i,j)} = \tilde{\mathbf{A}}^{(i,j)}\tilde{\mathbf{\Psi}}^{(I),(i,j)}(\tilde{\mathbf{A}}^{(i,j)})^{-1} = \\ &\begin{bmatrix}[\mathbf{\Theta}^{(I)}]_{i,i} & [\mathbf{\Theta}^{(I)}]_{i,j}\\ [\mathbf{\Theta}^{(I)}]_{j,i} & [\mathbf{\Theta}^{(I)}]_{j,j}\end{bmatrix} = \begin{bmatrix}A_1^{(i,j)} & A_2^{(i,j)}\\ A_3^{(i,j)} & A_4^{(i,j)}\end{bmatrix}\begin{bmatrix}[\mathbf{\Psi}^{(I)}]_{i,i} & [\mathbf{\Psi}^{(I)}]_{i,j}\\ [\mathbf{\Psi}^{(I)}]_{j,i} & [\mathbf{\Psi}^{(I)}]_{j,j}\end{bmatrix}\begin{bmatrix}A_1^{(i,j)} & A_2^{(i,j)}\\ A_3^{(i,j)} & A_4^{(i,j)}\end{bmatrix}^{-1}, \forall I \in \{1,2,3\}\end{aligned}. \tag{20}$$

In this work, without assumptions in (22), two elements $[\tilde{\mathbf{\Theta}}^{(I)}]_{1,2}$ and $[\tilde{\mathbf{\Theta}}^{(I)}]_{2,1}$ are computed as

$$\begin{aligned}[\tilde{\mathbf{\Theta}}^{(I)}]_{1,2} &= A_1^{(i,j)}A_2^{(i,j)}\left(\Psi_{j,j}^{(I)} - \Psi_{i,i}^{(I)}\right) + (A_1^{(i,j)})^2\Psi_{i,j}^{(I)} - (A_2^{(i,j)})^2\Psi_{j,i}^{(I)}\\ [\tilde{\mathbf{\Theta}}^{(I)}]_{1,2} &= A_3^{(i,j)}A_4^{(i,j)}\left(\Psi_{i,i}^{(I)} - \Psi_{j,j}^{(I)}\right) + (A_4^{(i,j)})^2\Psi_{j,i}^{(I)} - (A_3^{(i,j)})^2\Psi_{i,j}^{(I)}\end{aligned}. \tag{21}$$

4.2 JEVD with Laplacian constraint

The sound field satisfies the Helmholtz equation (27). Therefore, the following equation always holds for the wavenumber components (k_x, k_y, k_z) in the Cartesian coordinates:

$$k_x^2 + k_y^2 + k_z^2 = k^2. \tag{22}$$

In an ideal case without noises, $\mathbf{\Phi}^{(1)}$, $\mathbf{\Phi}^{(2)}$, and $\mathbf{\Phi}^{(3)}$ matrices obtained from the acquired sound field are the real-valued diagonal matrices with ($\cos\phi_d \sin\theta_d$, $\sin\phi_d \sin\theta_d$, and $\cos\theta_d$) as their (d,d) -th element, respectively. Therefore, the following expression should be satisfied:

$$\begin{aligned}&\mathbf{\Phi}^{(1)}\mathbf{\Phi}^{(1)} + \mathbf{\Phi}^{(2)}\mathbf{\Phi}^{(2)} + \mathbf{\Phi}^{(3)}\mathbf{\Phi}^{(3)}\\ &= \hat{\mathbf{T}}^{-1}(\mathbf{\Psi}^{(1)}\mathbf{\Psi}^{(1)} + \mathbf{\Psi}^{(2)}\mathbf{\Psi}^{(2)} + \mathbf{\Psi}^{(3)}\mathbf{\Psi}^{(3)})\hat{\mathbf{T}} = \mathbf{I}_{D\times D}\end{aligned}. \tag{23}$$

However, in practical situations, it is impossible to satisfy Equation (23) because of the noises. In this work, we propose to use a robust JEVD algorithm by adding the iterative JEVD algorithm (22) into the above Laplacian constraint (L-constraint). The cost function of the proposed JEVD algorithm is changed from the cost function of a block coordinate procedure (Equation (19)) as

$$C_{\text{const}}(\tilde{\mathbf{A}}^{(i,j)}) = (1-\lambda)\sum_{I=1}^{3}(\left|[\tilde{\mathbf{\Theta}}^{(I)}]_{1,2}\right|^2 + \left|[\tilde{\mathbf{\Theta}}^{(I)}]_{2,1}\right|^2) + \lambda(\left|[\tilde{\mathbf{\Xi}}]_{1,2}\right|^2 + \left|[\tilde{\mathbf{\Xi}}]_{2,1}\right|^2 + \left|[\tilde{\mathbf{\Xi}}]_{1,1}-1\right|^2 + \left|[\tilde{\mathbf{\Xi}}]_{2,2}-1\right|^2), \tag{24}$$

where $\tilde{\mathbf{\Xi}}^{(i,j)} = \tilde{\mathbf{A}}^{(i,j)}(\tilde{\mathbf{\Psi}}^{(1)(i,j)}\tilde{\mathbf{\Psi}}^{(1)(i,j)} + \tilde{\mathbf{\Psi}}^{(2)(i,j)}\tilde{\mathbf{\Psi}}^{(2)(i,j)} + \tilde{\mathbf{\Psi}}^{(3)(i,j)}\tilde{\mathbf{\Psi}}^{(3)(i,j)})(\tilde{\mathbf{A}}^{(i,j)})^{-1} \in \mathbb{R}^{2\times 2}$, and elements of $\tilde{\mathbf{\Xi}}^{(i,j)}$ are computed as follows,

$$\begin{aligned}[\tilde{\mathbf{\Xi}}]_{1,2} &= A_1^{(i,j)}A_2^{(i,j)}(\bar{\Psi}_{(4)}^{(i,j)} - \bar{\Psi}_{(1)}^{(i,j)}) + (A_1^{(i,j)})^2\bar{\Psi}_{(2)}^{(i,j)} - (A_2^{(i,j)})^2\bar{\Psi}_{(3)}^{(i,j)}\\ [\tilde{\mathbf{\Xi}}]_{2,1} &= A_3^{(i,j)}A_4^{(i,j)}(\bar{\Psi}_{(1)}^{(i,j)} - \bar{\Psi}_{(4)}^{(i,j)}) + (A_4^{(i,j)})^2\bar{\Psi}_{(3)}^{(i,j)} - (A_3^{(i,j)})^2\bar{\Psi}_{(2)}^{(i,j)}\\ [\tilde{\mathbf{\Xi}}]_{1,1} &= A_1^{(i,j)}A_4^{(i,j)}\bar{\Psi}_{(1)}^{(i,j)} - A_1^{(i,j)}A_3^{(i,j)}\bar{\Psi}_{(2)}^{(i,j)} + A_2^{(i,j)}A_4^{(i,j)}\bar{\Psi}_{(3)}^{(i,j)} - A_2^{(i,j)}A_3^{(i,j)}\bar{\Psi}_{(4)}^{(i,j)}\\ [\tilde{\mathbf{\Xi}}]_{1,2} &= A_1^{(i,j)}A_4^{(i,j)}\bar{\Psi}_{(4)}^{(i,j)} + A_1^{(i,j)}A_3^{(i,j)}\bar{\Psi}_{(2)}^{(i,j)} - A_2^{(i,j)}A_4^{(i,j)}\bar{\Psi}_{(3)}^{(i,j)} - A_2^{(i,j)}A_3^{(i,j)}\bar{\Psi}_{(1)}^{(i,j)}\end{aligned}, \tag{25}$$

where

$$\begin{aligned}\bar{\Psi}_{(1)}^{(i,j)} &= \sum_{I=1}^{3}([\mathbf{\Psi}^{(I)}]_{i,i}[\mathbf{\Psi}^{(I)}]_{i,i} + [\mathbf{\Psi}^{(I)}]_{i,j}[\mathbf{\Psi}^{(I)}]_{j,i})\\ \bar{\Psi}_{(2)}^{(i,j)} &= \sum_{I=1}^{3}([\mathbf{\Psi}^{(I)}]_{i,i}[\mathbf{\Psi}^{(I)}]_{i,j} + [\mathbf{\Psi}^{(I)}]_{i,j}[\mathbf{\Psi}^{(I)}]_{j,j})\\ \bar{\Psi}_{(3)}^{(i,j)} &= \sum_{I=1}^{3}([\mathbf{\Psi}^{(I)}]_{j,i}[\mathbf{\Psi}^{(I)}]_{i,i} + [\mathbf{\Psi}^{(I)}]_{j,j}[\mathbf{\Psi}^{(I)}]_{j,i})\\ \bar{\Psi}_{(4)}^{(i,j)} &= \sum_{I=1}^{3}([\mathbf{\Psi}^{(I)}]_{j,i}[\mathbf{\Psi}^{(I)}]_{i,j} + [\mathbf{\Psi}^{(I)}]_{j,j}[\mathbf{\Psi}^{(I)}]_{j,j})\end{aligned}. \tag{26}$$

In Equation (24), λ is the parameter that determines the weight between the original cost function (Equation (19)) and the L-constraint, ranging from 0 to 1. In this work, for each update of (i,j), we

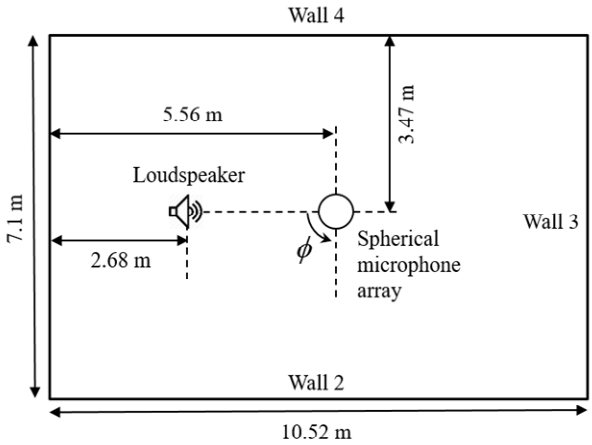
solve the unconstraint optimization problem with the cost function in Equation (24) by using the Quasi-Newton method (28, 29).

5. EVALUATION

5.1 Setup

To investigate the estimation performance for the early reflections, we conducted simulations with synthesized room impulse responses. The impulse responses of the room were generated with a spherical microphone array impulse response generator (30) and convolved with the white Gaussian noise to generate MC-SHD data. The rigid body SMA with radius 0.042 m that composed of 32 uniformly sampled microphones was used. The SMA signals were sampled at 16 kHz. The STFT was done by Rectangular window of 480 samples (30 ms) in length, a hop size of 120 samples (7.5 ms), and DFT size of 512.

The room setup is described in the left side of Figure 1. The height of the room is 3.03 m, and the height of the center of the SMA and loudspeaker unit are 1.41 m. The sound field was decomposed up to the fourth-order of spherical harmonics ($N = 4$). To de-correlate the highly correlated early reflections, the real-valued covariance matrix ($\mathbf{R}_r$) was averaged over 500 time frames and frequency range within 2.275–3.9 kHz ($kr = 1.75 \sim 3$) (31, 32). The signal-to-noise ratio (SNR) is defined to the total power of the sound sources to the total power of the self-microphone noises. The self-microphone noises were generated with the uncorrelated white Gaussian noises. The ground truth DOAs of direct sound and early reflections are shown in the right side of Figure 1. The estimation errors between the true DOAs and estimated DOAs were computed as $\Delta\Omega_d = \cos^{-1}(\cos\theta_d \cos\hat{\theta}_d + \cos(\phi_d - \hat{\phi}_d)\sin\theta_d \sin\hat{\theta}_d)$. The root-mean-square error (RMSE) is defined as $\mathrm{RMSE}(\Omega) = \{(1/KD)\sum_{\kappa=1}^{K}\sum_{d=1}^{D}|\Delta\Omega_d|^2\}^{1/2}$ with $K = 300$ independent trials. The reverberation time of the simulated room was set to 300 ms.



	θ	ϕ
Direct sound	90°	0°
Ceiling	41°	0°
Floor	135°	0°
Wall 2	90°	69°
Wall 3	90°	180°
Wall 4	90°	293°

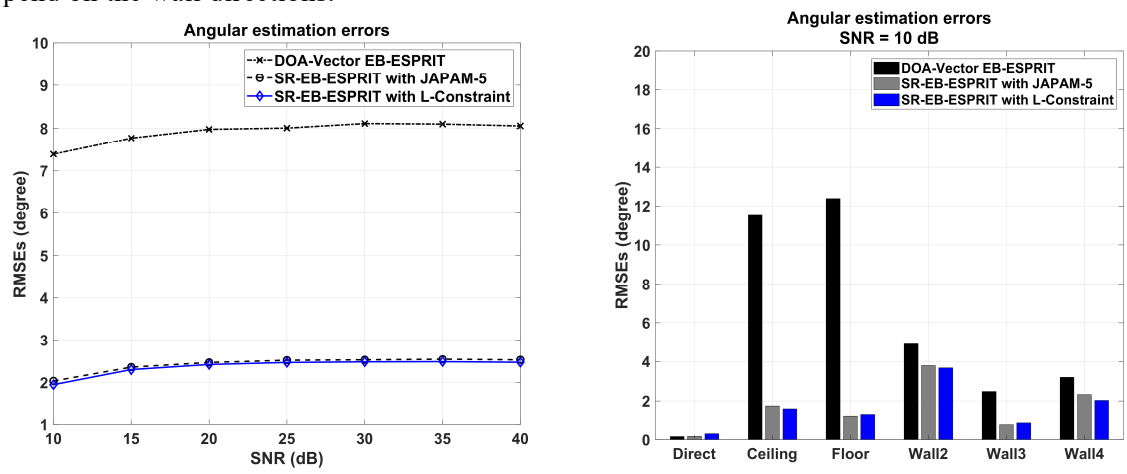
Figure 1 – The room setup and ground truth DOAs of the direct sound and early reflections in a room.

The baselines of this work were DOA-vector EB-ESPRIT (15) and the SR-EB-ESPRIT using JAPAM-5 algorithm without L-constraint (22). The diagonalizing matrix $\mathbf{A}$ in the first iteration was initialized as the inverse of eigenvector matrix of $\mathbf{\Psi}^{(1)}$ for JAPAM-5 algorithm and the proposed optimization algorithm. The maximum iteration number was set to 150, and the stopping criterion for JAPAM-5 was set as same in (22). In the proposed method, λ was set to 0.05. The optimization with the proposed cost function (Equation (24)) was solved by using the unconstraint multivariable and nonlinear optimization solver (`fminunc`) of the MATLAB®. The optimization algorithm was selected as 'Quasi-Newton' algorithm (29), and the initial points of $A_1^{(i,j)}$, $A_2^{(i,j)}$, $A_3^{(i,j)}$, and $A_4^{(i,j)}$ were selected from $\{0.3, 0.65, 1\}$ that minimize the cost function output ($C_{\mathrm{const}}(\tilde{\mathbf{A}}^{(i,j)})$) for each (i, j) step. The total iteration number of the proposed algorithm was set to 1.

5.2 Results

In Figure 2 (a), RMSEs evaluated in various SNR conditions (SNR = 10 dB ~ 40 dB) are shown. RMSEs for all algorithms change slightly as SNR decreases. This is because the artifacts of the late

reverberant components have dominant influences on the performance of algorithms. The RMSEs of SR-EB-ESPRIT with JAPAM-5 and with L-constraint are only one-fourth compared to the case of DOA-vector EB-ESPRIT in all SNR regions. The proposed SR-EB-ESPRIT with L-constraint estimates slightly better than the SR-EB-ESPRIT with JAPAM-5, which is caused by adding the physical constraints based on the Helmholtz equation. To validate the performance for each reflection, the RMSE for each wall was examined in SNR = 10 dB condition (Figure 2 (b)). Without direct sound, RMSEs of DOA-vector EB-ESPRIT are quite high. Unlike the RMSEs of DOA-vector EB-ESPRIT, both RMSEs of SR-EB-ESPRIT with JAPAM-5 and SR-EB-ESPRIT with L-constraint are reasonably small (under 4 degrees), but their performance gap is ignorable. To summarize, both SR-EB-ESPRIT techniques proposed in this work show outstanding performance compared to the state-of-the art technique. The L-constraint brings a slight performance gain in overall RMSEs, but the differences depend on the wall directions.



(a) (b)

Figure 2 – (a) RMSEs for various SNRs (10 ~ 40 dB)

(b) RMSEs of the direct sound and the early reflections with SNR = 10 dB.

6. CONCLUSIONS

We proposed the DOA estimation method for direct and early reflection sounds in a room by using the semi real-valued processing and joint eigenvalue decomposition with constraint. The proposed method can reduce the computational complexity by using the real-valued covariance matrix and real-valued common transformation matrix. In addition, by using the L-constraint based on the Helmholtz equation, the robustness of the proposed method is slightly improved. Through the simulation results in various SNRs, we showed that the estimation performance of the proposed method outperforms other baseline algorithms in a room.

ACKNOWLEDGEMENTS

This work was supported by the Korea Research Foundation (KRF) under Contract No. 2019R1A2C1007393 and by the BK21 Plus program through the National Research Foundation (NRF) funded by the Ministry of Education of Korea.

REFERENCES

1. Mabande E, Kowalczyk K, Sun H, Kellermann W. Room geometry inference based on spherical microphone array eigenbeam processing. J Acoust Soc Am. 2013; 134(4) 2773–2789.
2. Sun H, Yan S, Svensson U. P. Space domain optimal beamforming for spherical microphone arrays. Proc IEEE Int Conf Acoust Speech Signal Process; 2010. p. 117–120.
3. Peled Y, Rafaely B. Method for dereverberation and noise reduction using spherical microphone arrays. Proc IEEE Int Conf Acoust Speech Signal Process; 2010. p. 113–116.
4. Betlehem T, Poletti M. A. Two dimensional sound field reproduction using higher order sources to exploit room reflections. J Acoust Soc Am. 2014; 135(4) 1820–1833.

5. Sun H, Mabande E, Kowalczyk K, Kellermann W. Joint DOA and TDOA estimation for 3D localization of reflective surfaces using eigenbeam MVDR and spherical microphone arrays. Proc IEEE Int Conf Acoust Speech Signal Process. 2011; p. 113–116.
6. Morgenstern H, Zotter F, Rafaely B. Joint spherical beam forming for directional analysis of reflections in rooms. J. Acoust. Soc. Am. 2012; 131(4) 3207.
7. Morgenstern H, Rafaely B. Modal smoothing for analysis of room reflections measured with spherical microphone and loudspeaker arrays. J Acoust Soc Am. 2018; 143(2) 1008–1018.
8. Tervo S, Politis A. Direction of arrival estimation of reflections from room impulse responses using a spherical microphone array. IEEE/ACM Trans Audio Speech Lang Process. 2015; 23(10) 1539–1551.
9. Teutsch H, Kellermann W. Eigen-beam processing for direction-of-arrival estimation using spherical apertures. Proc. Joint Workshop Hands-Free Commun and Microphone Arrays; 2005.
10. Sun H, Teutsch H, Mabande E, Kellermann W. Robust localization of multiple sources in reverberant environments using EB-ESPRIT with spherical microphone arrays. Proc IEEE Int Conf Acoust Speech Signal Process; 2011. p. 117–120.
11. Sun, H, Mabande E, Kowalczyk K, Kellermann W. Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing. J Acoust Soc Am. 2012; 131(4) 2828–2840.
12. Huang Q, Zhang L, Yong F. Two Step Spherical Harmonics ESPRIT-Type Algorithms and Performance Analysis. IEEE/ACM Trans Audio Speech Lang Process. 2018; 26(9) 1684–1697.
13. Jo B, Choi J. W. Direction of arrival estimation using nonsingular spherical ESPRIT. J Acoust Soc Am. 2018; 143(3) EL181–EL187.
14. Jo B, Choi J. W. Parametric direction-of-arrival estimation with three recurrence relations of spherical harmonics. J Acoust Soc Am. 2019; 145(1) 480–488.
15. Herzog A, Habets E. A. Eigenbeam-ESPRIT for DOA-Vector Estimation. IEEE Signal Process. Lett. 2019; 26(4) 572–576.
16. Huarng K. C, Yeh C. C. A unitary transformation method for angle-of-arrival estimation. IEEE Trans Signal Process. 1991; 39(4) 975–977.
17. Haardt M, Nossek J. A. Unitary ESPRIT: How to obtain increased estimation accuracy with a reduced computational burden. IEEE Trans signal process. 1995; 43(5) 1232–1242.
18. Goossens R, Rogier H. Unitary spherical ESPRIT: 2D-angle estimation with spherical arrays for scalar fields. IET signal process. 2009; 3(3) 221–231.
19. Krishnaveni V, Kesavamurthy T, Aparna B. Beamforming for direction-of-arrival (DOA) estimation-a survey. Int J Comput Appl. 2013; 61(11).
20. Fu T, Gao X. Simultaneous diagonalization with similarity transformation for non-defective matrices. Proc IEEE Int Conf Acoust Speech Signal Process; 2006. p. IV.
21. Balda E. R, Cheema S. A, Weiss A, Yeredor A, Haardt M. Perturbation analysis of joint eigenvalue decomposition algorithms. Proc IEEE Int Conf Acoust Speech Signal Process; 2017. p. 3101–3105.
22. André R, Luciani X, Moreau E. A new class of block coordinate algorithms for the joint eigenvalue decomposition of complex matrices. Signal Process. 2018; 145 78–90.
23. Rafaely B. Fundamentals of spherical array processing; 2015.
24. Jarrett D, Habets E, Naylor P. Theory and applications of spherical microphone array processing; 2017.
25. Choi C, Ivanic J, Gordon M. S, Ruedenberg K. Rapid and stable determination of rotation matrices between spherical harmonics by direct recursion. J Chem Phys. 1999; 111(19) 8825–8831.
26. Gumerov N. A, Duraiswami R. Fast multipole methods for the Helmholtz equation in three dimensions; 2005.
27. Kinsler L. E, Frey A. R, Coppens A. B, Sanders J. V. Fundamentals of Acoustics. 4th Edition; 1999.
28. Chong E. K, Zak S. H. An introduction to optimization; 2013.
29. Shanno D. F. Conditioning of quasi-Newton methods for function minimization. Math of comput. 1970; 24(111) 647–656.
30. Jarrett D, Habets E. A. P, Thomas M. R. P, Naylor P. A. Rigid sphere room impulse response simulation: Algorithm and applications. J Acoust Soc Am. 2012; 132(3) 1462–1472.
31. Khaykin D, Rafaely B. Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach. Proc IEEE Workshop on Appl of Signal Process Audio and Acoustics; 2009. p. 221–224.
32. Huleihel N, Rafaely B. Spherical array processing for acoustic analysis using room impulse responses and time-domain smoothing. J Acoust Soc Am. 2013; 133(6) 3995–4007.