MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

Experimental validation of vector-based EB-ESPRIT technique for the localization of early reflections in a reverberant room

Jo, Byeongho[1]
Korea Advanced Institute of Science and Technology (KAIST)
School of Electrical Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon, South Korea

Choi, Jung-Woo[2]
Korea Advanced Institute of Science and Technology (KAIST)
School of Electrical Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon, South Korea

ABSTRACT

The eigenbeam (EB)-ESPRIT is one of the popular parametric direction-of-arrival (DOA) estimation techniques for a spherical microphone array. The EB-ESPRIT directly estimates the directional parameters by using a single recurrence relation of spherical harmonics, and has also been applied for the localization of early reflections in a reverberant environment. However, there are couple of limitations that hinder the use of the EB-ESPRIT for a real room data. The EB-ESPRIT exhibits a matrix ill-conditioning problem for sources positioned near the equator of the spherical coordinates, which can frequently occur by echoes impinging from those angles. Furthermore, the number of simultaneously detectable sources is also limited. To overcome these limitations, the vector-based EB-ESPRIT technique was introduced by utilizing three recurrence relations of spherical harmonics corresponding to x-, y-, and z- components of a vector in the Cartesian coordinates. In this work, we conduct experiments in a real reverberant room to validate the performance of vector-based EB-ESPRIT, and demonstrate that the vector-based EB-ESPRIT can successfully localize the first-order echoes from the data recorded in the reverberant room.

Keywords: EB-ESPRIT, recurrence relations, spherical microphone array
I-INCE Classification of Subject Number: 74

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

1. INTRODUCTION

The directions-of-arrival (DOAs) estimation of sound sources using a spherical microphone array (SMA) has been investigated as an essential part of room acoustics [1] [2]. The sound field captured by a SMA can be transformed into the spherical harmonic domain (SHD) data or eigen-beams (EB) and processed by various eigenbeam techniques. Using the eigenbeam-based beamformer approach, Sun *et al.* estimated distinct room reflections [3] in a reverberant room. Using the room impulse responses measured by SMA, Tervo *et al.* presented the eigenbeam maximum likelihood (ML) method for DOA estimation of coherent acoustic reflections [4]. Mabande *et al.* studied the room geometry inference by using the eigenbeam-based beamformers and the time differences of arrivals (TDOAs) [5].

Unlike the beamformer-based techniques, the eigenbeam estimation of signal parameters via the rotational invariance technique (EB-ESPRIT) has an advantage that it does not need to an exhaustive grid search [6]. The EB-ESPRIT uses a single recurrence relation of spherical harmonics to estimate the DOAs in a parametric way. However, as the EB-ESPRIT has a matrix ill-conditioning problem and can detect a limited number of sources [7], the DOA estimation of acoustic reflections in a reverberant room is impeded, in practice.

There have been several works to overcome these problems. For instance, to avoid the ill-conditioning problem, Sun *et al.* proposed the rotation of the reference coordinates [7], and Huang *et al.* proposed two-step spherical harmonics ESPRIT-type algorithms [8]. A nonsingular EB-ESPRIT using the two sine-based recurrence relations of spherical harmonics was also presented by authors [9]. However, all these methods are accompanied by some problems. For the coordinates rotation, for instances, one should determine the proper rotation angle for each axis, and even after rotation the sound sources can still be on the equator. The two-step methods require additional elevation and azimuth pairing algorithms. Furthermore, due to the properties of the elevation parameters, both two-step methods and nonsingular EB-ESPRIT have fundamental degradation near the equator and poles, respectively.

In this work, we propose the vector-based EB-ESPRIT that uses three independent recurrence relations of spherical harmonics [10]. From the three recurrence relations, the directional vectors of early reflections in Cartesian coordinates are uniquely extracted, and hence, the proposed method is not induced any singularity problem. Furthermore, by utilizing a common transformation matrix, we directly estimate DOAs in a single stage without any pairing algorithms. For the performance validation in a real environment, we carried out experiments in a reverberant room with varying reverberation times (T_{60}). The experimental results in a real room show that the proposed EB-ESPRIT can estimate DOAs of early reflections successfully within reasonably small errors.

2. SPHERICAL HARMONIC DOMAIN PROCESSING

First, the sound field is measured by a SMA in order to estimate DOAs of early reflections. Suppose that the radius of SMA is r, and capturing position is given by (r, θ, ϕ) in the spherical coordinates. Then, the sound field of frequency ω captured by

SMA at (r, θ, ϕ) can be described as a weighted summation of the spherical harmonics [1]

$$p(k, r, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} p_{nm}(k, r) Y_n^m(\theta, \phi), \tag{1}$$

where $k = \omega/c$ (c: sound speed) is the wavenumber, and $Y_n^m(\cdot)$ denotes the spherical harmonics with order n and degree m,

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

Here, $P_n^m(\cdot)$ is the associated Legendre function.

The captured data transformed into the SHD, called spherical harmonic coefficients $p_{nm}(k, r)$, can be computed from the measured data $p(k, r, \theta, \phi)$ by using the spherical Fourier transform (SFT) as

$$p_{mm}(k, r) = \int_0^{2\pi} \int_0^{\pi} p(k, r, \theta, \phi) [Y_n^m(\theta, \phi)]^* \sin\theta d\theta d\phi, \tag{3}$$

where $(\cdot)^*$ denotes the complex conjugate.

When D multiple plane waves are simultaneously incident on the SMA, the spherical harmonic coefficients can be derived as

$$p_{nm}(k, r) = \sum_{d=1}^{D} b_n(kr) [Y_n^m(\Omega_s^d)]^* s_d(k), \tag{4}$$

where $s_d(k)$ and $\Omega_s^d = (\theta_s^d, \phi_s^d)$ denote the complex amplitudes and DOAs of sources, respectively. The mode strength is represented as $b_n(kr)$, which depends on the radius and the boundary condition of the SMA. For this work, we use the SMA with the rigid boundary condition ($b_n(kr) = 4\pi i^{n+1}/[(kr)^2 h_n^{(1)\prime}(kr)]$), where $h_n^{(1)\prime}(\cdot)$ is the derivative of a spherical Henkel function of the first kind [1]. In order to do the EB processing with frequency-independent eigenbeams, the mode strength compensation should be accompanied as follows [2],

$$a_{nm}(k) = p_{nm}(k, r)/b_n(kr) = \sum_{d=1}^{D} [Y_n^m(\Omega_s^d)]^* s_d(k). \tag{5}$$

The mode strength compensated data named the directional harmonic coefficients $a_{nm}(k)$ can be extracted up to a limited order of spherical harmonics (N) because the sound field is discretely sampled with a limited number of microphones in practice. In this manner, $a_{nm}(k)$ can be re-written in a vector form up to the harmonics order N as

$$\mathbf{a}(k) = \mathbf{Y}^H \mathbf{s}(k), \tag{6}$$

where $\mathbf{a}(k) = [a_{0,0}(k),\ a_{1,-1}(k),\ \cdots,\ a_{N,N}(k)]^T \in \mathbb{C}^{(N+1)^2 \times 1}$, $\mathbf{s}(k) = [s_1(k),\ \cdots,\ s_D(k)]^T \in \mathbb{C}^{D \times 1}$, and $\mathbf{Y}^H \in \mathbb{C}^{(N+1)^2 \times D}$ is the array manifold matrix in the SHD whose d-th column is given by

$$\mathbf{y}\left(\Omega_s^{(d)}\right) = [\underbrace{Y_0^0(\Omega_s^d),}_{n=0}\ \underbrace{Y_1^{-1}(\Omega_s^d), Y_1^0(\Omega_s^d), Y_1^1(\Omega_s^d),}_{n=1}\ \cdots,\ \underbrace{Y_N^{-N}(\Omega_s^d),\ \cdots,\ Y_N^N(\Omega_s^d)}_{n=N}]^H, \tag{7}$$

where $\{\cdot\}^H$ denotes the conjugate transpose.

For the EB processing, a covariance matrix should be constructed first. From multiple snapshots or observations, the covariance matrix can be computed as follows,

$$\mathbf{R}(k) = E\{\mathbf{a}(k)\mathbf{a}(k)^H\} = \mathbf{Y}^H E\{\mathbf{s}(k)\mathbf{s}(k)^H\}\mathbf{Y}, \tag{8}$$

where $E\{\cdot\}$ denotes the expectation operator.

For de-correlating the coherent early reflections due to distinct walls, smoothing techniques should be applied. The spherical harmonics ESPRIT-type algorithms ([6], [8], [9]) can estimate DOAs of early echoes in a reverberant environment by using the smoothing techniques such as frequency smoothing [11], temporal smoothing [12], modal smoothing [13], and spherical harmonic smoothing [14]. In this work, the smoothed covariance matrix $\bar{\mathbf{R}}$ is constructed by averaging the time-frequency (TF) data as

$$\bar{\mathbf{R}} = \frac{1}{J_t J_f} \sum_{t=1}^{J_t} \sum_{f=1}^{J_f} \mathbf{a}_t(f)\mathbf{a}_t^H(f), \tag{9}$$

where J_t and J_f denote the total number of time frames and frequency bins.

3. VECTOR-BASED EB-ESPRIT

Unlike the conventional EB-ESPRIT [6] and other spherical harmonics ESPRIT-type algorithms [8] [9], the vector-based EB-ESPRIT utilizes three recurrence relations of spherical harmonics as follows [15],

$$\begin{aligned}
\sin\theta_s e^{i\phi_s}[Y_n^m(\Omega_s)]^* &= w_{n+1}^{m-1}[Y_{n+1}^{m-1}(\Omega_s)]^* - w_n^{-m}[Y_{n-1}^{m-1}(\Omega_s)]^* \\
\sin\theta_s e^{-i\phi_s}[Y_n^m(\Omega_s)]^* &= -w_{n+1}^{-(m+1)}[Y_{n+1}^{m+1}(\Omega_s)]^* + w_n^{m}[Y_{n-1}^{m+1}(\Omega_s)]^* \\
\cos\theta_s[Y_n^m(\Omega_s)]^* &= v_n^m[Y_{n-1}^{m}(\Omega_s)]^* + v_{n+1}^m[Y_{n+1}^{m}(\Omega_s)]^*,
\end{aligned} \tag{10}$$

where

$$\begin{aligned}
w_n^m &= \sqrt{(n-m-1)(n-m)/(2n-1)(2n+1)} \\
v_n^m &= \sqrt{(n-m)(n+m)/(2n-1)(2n+1)}.
\end{aligned} \tag{11}$$

In order to jointly utilize three recurrence relations with different orders, three relations are re-written in a matrix form with the shifted and order-reduced harmonic matrix $\mathbf{Y}^{(\mu,\nu)}$

$$\begin{aligned}
\mathbf{Y}^{(0,0)H}\boldsymbol{\Phi}_{xy+} &= \mathbf{W}_{n+1}^{m-1}\mathbf{Y}^{(+1,-1)H} - \mathbf{W}_n^{-m}\mathbf{Y}^{(-1,-1)H} \\
\mathbf{Y}^{(0,0)H}\boldsymbol{\Phi}_{xy-} &= -\mathbf{W}_{n+1}^{-(m+1)}\mathbf{Y}^{(+1,+1)H} + \mathbf{W}_n^{m}\mathbf{Y}^{(-1,+1)H} \\
\mathbf{Y}^{(0,0)H}\boldsymbol{\Phi}_{z} &= \mathbf{V}_n^{m}\mathbf{Y}^{(-1,0)H} + \mathbf{V}_{n+1}^{m}\mathbf{Y}^{(+1,0)H},
\end{aligned} \tag{12}$$

where $\boldsymbol{\Phi}_{xy+}$, $\boldsymbol{\Phi}_{xy-}$, and $\boldsymbol{\Phi}_z$ are $D \times D$ complex-valued diagonal matrices containing directional parameters ($\sin\theta_s^d e^{i\phi_s^d}$, $\sin\theta_s^d e^{-i\phi_s^d}$, and $\cos\theta_s^d$). $\mathbf{W}_{n+\mu}^{m+\nu}$ 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 $\mathbf{Y}^{(\mu,\nu)H} \in \mathbb{C}^{N^2 \times D}$ is given by

$$\begin{aligned}\mathbf{y}_{\mu}^{\nu}\left(\Omega_s^{(d)}\right) =& [\underbrace{Y_{0+\mu}^{0+\nu}(\Omega_s^{(d)}),}_{n=0}\ \underbrace{Y_{1+\mu}^{-1+\nu}(\Omega_s^{(d)}),\ Y_{1+\mu}^{0+\nu}(\Omega_s^{(d)}),\ Y_{1+\mu}^{1+\nu}(\Omega_s^{(d)}),}_{n=1}\ \cdots,\\ & \underbrace{Y_{(N-1)+\mu}^{-(N-1)+\nu}(\Omega_s^{(d)}),\ \cdots,\ Y_{(N-1)+\mu}^{(N-1)+\nu}(\Omega_s^{(d)})}_{n=N-1}]^{H}.\end{aligned} \tag{13}$$

The goal of this algorithm is estimating the directional parameter matrices $\mathbf{\Phi}_{xy+}$, $\mathbf{\Phi}_{xy-}$, and $\mathbf{\Phi}_{z}$. However, the shifted harmonic matrices $\mathbf{Y}^{(\mu,\nu)H}$ in the Equation 12 cannot be obtained from the measurement data. Instead, we utilize a relation between the signal eigenvector matrix $\mathbf{U}_s$ and array manifold matrix $\mathbf{Y}^H$ in SHD as follows [6],

$$\mathbf{U}_s = \mathbf{Y}^H\mathbf{T}, \tag{14}$$

where $\mathbf{U}_s$ is the signal eigenvector matrix that can be computed by extracting eigenvectors corresponding D largest eigenvalues from the covariance matrix, and the transformation matrix $\mathbf{T}$ is the invertible $D \times D$ matrix. Using the Equation 14, the Equation 12 is reformulated with respect to $\mathbf{U}_s$ as

$$\begin{aligned}\mathbf{U}_s^{(0,0)}\mathbf{\Psi}_{xy+} &= \underbrace{\mathbf{W}_{n+1}^{m-1}\mathbf{U}_s^{(+1,-1)} - \mathbf{W}_{n}^{-m}\mathbf{U}_s^{(-1,-1)}}_{\mathbf{\Lambda}_{xy+}}\\ \mathbf{U}_s^{(0,0)}\mathbf{\Psi}_{xy-} &= \underbrace{-\mathbf{W}_{n+1}^{-(m+1)}\mathbf{U}_s^{(+1,+1)} + \mathbf{W}_{n}^{m}\mathbf{U}_s^{(-1,+1)}}_{\mathbf{\Lambda}_{xy-}}\\ \mathbf{U}_s^{(0,0)}\mathbf{\Psi}_{z} &= \underbrace{\mathbf{V}_{n}^{m}\mathbf{U}_s^{(-1,0)} + \mathbf{V}_{n+1}^{m}\mathbf{U}_s^{(+1,0)}}_{\mathbf{\Lambda}_{z}},\end{aligned} \tag{15}$$

where $\mathbf{\Psi}_{\Delta} = \mathbf{T}^{-1}\mathbf{\Phi}_{\Delta}\mathbf{T}$ ($\Delta = \{xy+,\ xy-,\ z\}$) are transformed directional parameter matrices.

The Equation 15 can be solved with respect to $\mathbf{\Psi}_{\Delta}$ by the least-squares solutions under $N^2 \geq D$ as follows,

$$\hat{\mathbf{\Psi}}_{\Delta} = \{\mathbf{U}_s^{(0,0)}\}^{\dagger}\mathbf{\Lambda}_{\Delta}, \tag{16}$$

where $(\cdot)^{\dagger}$ denotes the Moore-Penrose inverse operator. From the condition ($N^2 \geq D$), the number of detectable sources (D_{max}) is determined to N^2, which is larger than conventional EB-ESPRIT ($D_{max} = (N-1)^2$) and two-step EB-ESPRIT($D_{max} = \lfloor N^2/2 \rfloor$), and smaller than nonsingular EB-ESPRIT ($D_{max} = \lfloor N^2 + N/2 \rfloor$).

Instead of extracting the eigenvalues of $\hat{\mathbf{\Psi}}_{\mathbf{\Delta}}$ matrices separately ([6], [8], [9]), we utilize the left eigenvector matrix ($\mathbf{U}_{xy+}$) of $\mathbf{\Psi}_{xy+}$ ($= \mathbf{U}_{xy+}\mathbf{\Phi}_{xy+}\mathbf{U}_{xy+}^{-1}$) matrix as a common transformation matrix to estimate the directional parameter matrices $\mathbf{\Phi}_{xy+}$, $\mathbf{\Phi}_{xy-}$, and $\mathbf{\Phi}_{z}$ as follows,

$$\begin{aligned}\hat{\mathbf{\Phi}}_{xy+} &= \mathbf{U}_{xy+}^{-1}\mathbf{\Psi}_{xy+}\mathbf{U}_{xy+}\\ \hat{\mathbf{\Phi}}_{xy-} &= \mathbf{U}_{xy+}^{-1}\mathbf{\Psi}_{xy-}\mathbf{U}_{xy+}\\ \hat{\mathbf{\Phi}}_{z} &= \mathbf{U}_{xy+}^{-1}\mathbf{\Psi}_{z}\mathbf{U}_{xy+}.\end{aligned} \tag{17}$$

Let the diagonal components of the estimated directional parameter matrices ($\hat{\mathbf{\Phi}}_\Delta$) as $[\hat{\mathbf{\Phi}}_{xy+}]_{d,d} = \hat{\Phi}^d_{xy+}$, $[\hat{\mathbf{\Phi}}_{xy-}]_{d,d} = \hat{\Phi}^d_{xy-}$, and $[\hat{\mathbf{\Phi}}_{z}]_{d,d} = \hat{\Phi}^d_{z}$. Then, the DOA of the d-th source can be computed from the directional vector in the Cartesian coordinate ($[x^d_s, y^d_s, z^d_s] = \mathrm{Re}\{[(\hat{\Phi}^d_{xy+} + \hat{\Phi}^d_{xy-})/2, (\hat{\Phi}^d_{xy+} - \hat{\Phi}^d_{xy-})/2i, \hat{\Phi}^d_z]\}$) as

$$\begin{aligned} \hat{\theta}^d_s &= \mathrm{atan2}\left\{ \sqrt{(x^d_s)^2 + (y^d_s)^2}, z^d_s \right\} \\ \hat{\phi}^d_s &= \mathrm{atan2}\{y^d_s, x^d_s\}, \end{aligned} \tag{18}$$

for all sources ($d = 1, \cdots, D$). $\mathrm{Re}\{\cdot\}$ extracts the real part of the argument and $\mathrm{atan2}\{\cdot,\cdot\}$ is a four-quadrant inverse tangent with two variables.

4. PERFORMANCE EVALUATION

4.1. Setup

To validate the echo localization performance, we conducted a series of experiments in the real room environment. For comparison with the ideal situation, we carried out simulations with the same room setup and reverberation times. The impulse responses of the ideal simulation environment were generated with a spherical microphone array impulse response generator [16] and convolved with the white Gaussian noise to generate the simulated microphone data. The baselines to compare this work were the semi-RTS-SHESPRIT (real-valued two-step spherical harmonics ESPRIT) [8] and sine-based EB-ESPRIT [9]. Note that the conventional EB-ESPRIT [6] is not included because it cannot estimate enough number of early reflections. The EB-MUSIC spectrum was used for resolving the ambiguity problem of the sine-based EB-ESPRIT. For the semi-RTS-SHESPRIT, the pairing of elevation and azimuth angles was done by choosing the best matches with true DOAs.

A stationary white Gaussian noise was consistently played as a direct sound. The SMA signals were measured for 5 seconds and sampled at 48 kHz. The directional harmonic coefficients were generated by SFT, and short time Fourier transform (STFT)

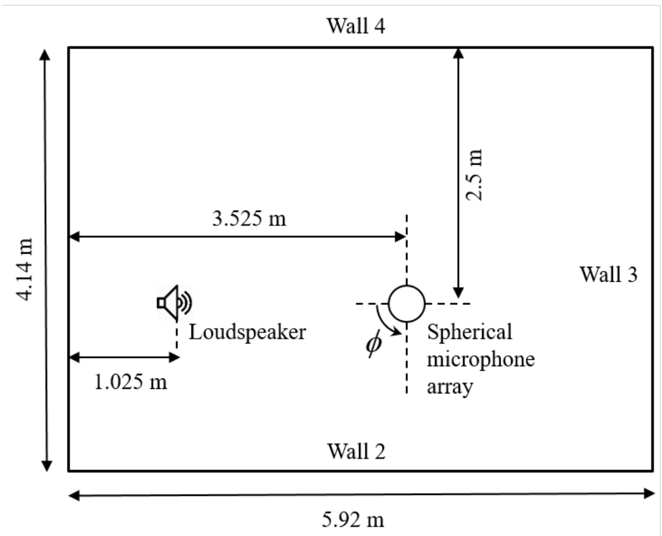

Figure 1: The measurement setup

Table 1: Ground truth DOAs of the direct sound and early reflections in a room

	Direct sound	Ceiling	Floor	Wall 2	Wall 3	Wall 4
θ	90°	40.9°	130.8°	90°	90°	90°
ϕ	0°	0°	0°	52.7°	180°	296.6°

with Hanning window of 480 samples (10 ms) in length, a hop size of 240 samples (5 ms), and DFT size of 512.

The measurement setup is described in Figure 1. The height of the room was 2.52 m, and the height of the center of the SMA and loudspeaker unit were set to 1.08 m. The SMA used to capture the sound field in the room was the Eigenmike®, which has 32 calibrated microphones placed on the rigid spherical body with a radius of $r = 0.042\,\text{m}$. The sound field was decomposed up to the fourth-order of spherical harmonics ($N = 4$) to secure enough number of detectable sources for localizing early reflections of all comparing algorithms (Semi-RTS SHESPRIT: $D_{\max} = 9$, sine-based EB-ESPRIT: $D_{\max} = 18$, proposed algorithm: $D_{\max} = 16$). For a sound source, a 1-inch full-range loudspeaker unit with custom enclosure of 40 cc volume was used. For de-correlating early reflections, the covariance matrix was averaged over all time-frames and frequency range within 0.65–3.9 kHz range($kr \in [0.5,\ 3]$).

The signal-to-noise ratio (SNR) with respect to the background noise and microphone self-noise was maintained around 48 dB within the frequency of interest. For the error analysis, the ground truth DOAs of direct sound and early reflections are computed from the known wall locations and their first-order image source positions (Table 1). The angular estimation errors between the true and estimated DOAs are computed as $\Delta\Omega_s^d = \cos^{-1}\left(\cos\theta_s^d \cos\hat{\theta}_s^d + \cos(\phi_s^d - \hat{\phi}_s^d)\sin\theta_s^d \sin\hat{\theta}_s^d\right)$. To evaluate the localization performance of the proposed method in practical acoustic environments, three reverberation times were considered (T_{60} = 0.22, 0.35, and 0.46 s) by changing the number of absorption boards (Figure 2). The noise reduction coefficient (NRC) of the absorption board is around 0.75.

The directional harmonic coefficients $\mathbf{a}(k)$ were computed by compensating with the soft-limiting regularization filter [17] of 25 dB maximum allowed amplification, in order to prevent the excessive amplification of high order harmonics in the low sensitivity regions. The soft-limiting filter in the frequency domain $d_n(kr)$ is constructed as follows

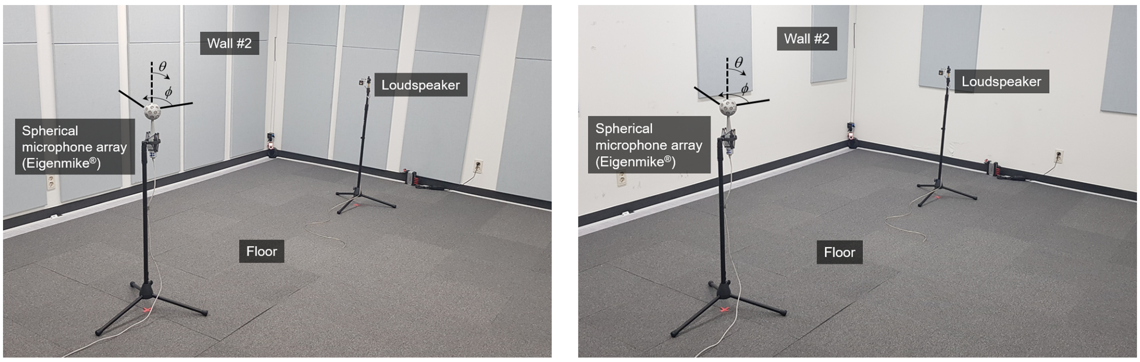

Figure 2: Experimental configurations: spherical array and loudspeaker arrangement in a room. Left: T_{60} = 0.22 s. Right: T_{60} = 0.46 s.

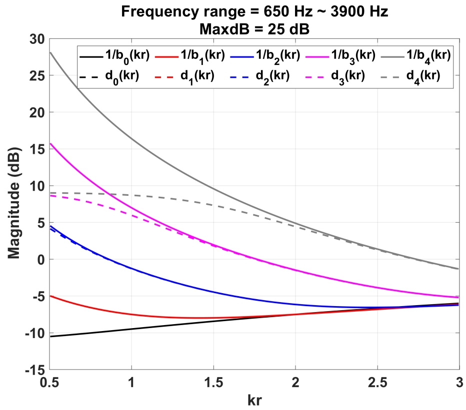

Figure 3: Magnitude of $1/b_n(kr)$ *and* $d_n(kr)$ *with orders of* $n = 0$ *to* 4

(Figure 3),

$$d_n(kr) = \frac{2 \cdot 10^{5/4}}{\pi} \frac{|b_n(kr)|}{b_n(kr)} \tan^{-1}\left(\frac{\pi}{2 \cdot 10^{5/4} \cdot |b_n(kr)|}\right). \tag{19}$$

4.2. Results

In Figure 4, the root mean squares of the angular estimation errors are shown with three different T_{60}. In the simulation results, the errors of two baselines (Sine-based EB-ESPRIT and Semi-RTS-SHESPRIT) increase as the reverberation time increases. By contrast, the proposed algorithm does not have such tendency and outperforms other algorithms with quite low errors of under 10° in all T_{60} conditions. In the experimental results, the relation between the accuracy and reverberation time is rather weak, and errors of all algorithms increase. Despite the fact that the error of the proposed algorithm in $T_{60} = 0.22$ s is rather high (about 15°) compared to the simulation, the vector-based EB-ESPRIT still outperforms other algorithms in all T_{60} conditions.

To investigate errors of different early reflections separately, the estimation errors of the direct sound and the different walls were examined for the $T_{60} = 0.46$ s condition

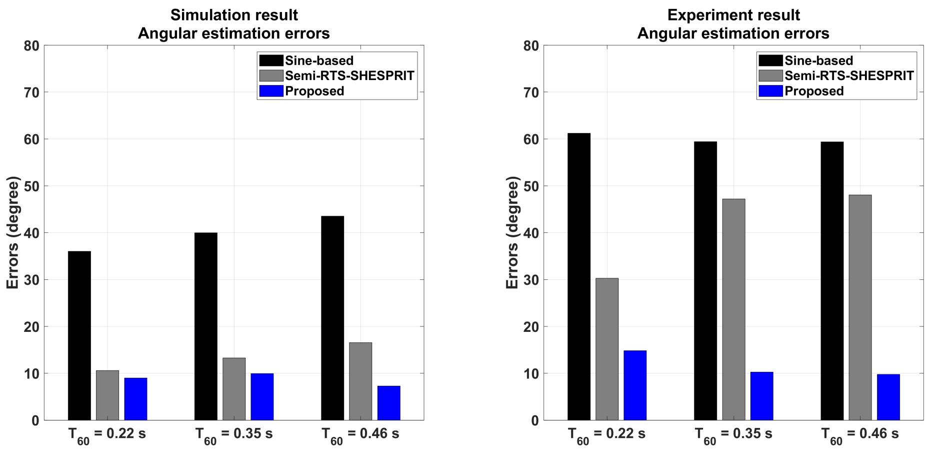

Figure 4: Angular estimation errors for different T_{60}*. Left: Simulation results. Right: Experimental results.*

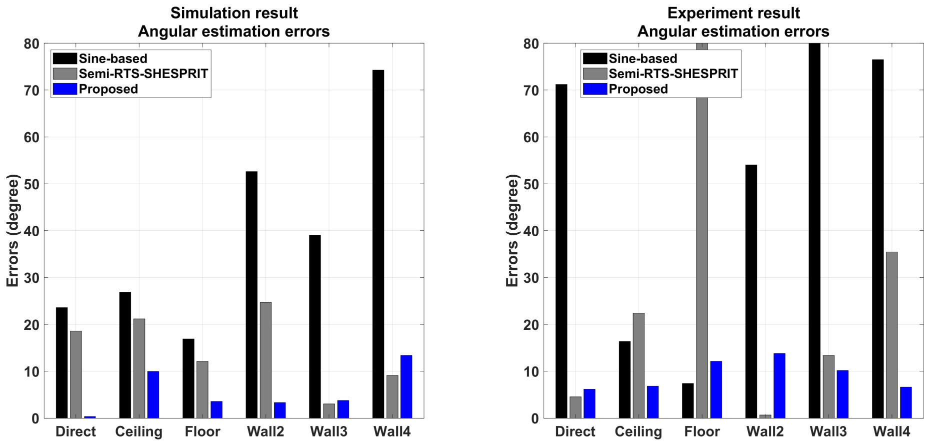

Figure 5: Estimation errors of direct sound and early reflections with $T_{60} = 0.46$ s. Left: Simulation results. Right: Experimental results.

(Figure 5). As described in the previous articles [8] [9], the performance degradation near the equator ($\theta = 90°$: Direct, Wall2, Wall3, Wall4) of the sine-based EB-ESPRIT can be observed due to the slow rate of change of the sine function near the equator in the simulation results. Likewise, the errors of the early reflections near the poles ($\theta = 0°, 180°$: Ceiling and Floor) of the semi-RTS-SHESPRIT are higher than those of other early reflections except for the wall 2. It is because of the slow rate of change of the cosine function near the poles. The exceptional result of the wall 2 may be caused by the high contribution of the 2nd or higher reflections. Those tendencies can also be observed in the experimental results except for the wall 4. Although overall errors of three algorithms increase in the experiments, the performances of the proposed algorithm are still better on average than others in both simulation and experiment results.

5. CONCLUSION

The vector-based EB-ESPRIT utilizes three different recurrence relations of spherical harmonics corresponding to the axis-directed vectors in the Cartesian coordinates. By utilizing all measured data in SHD and uses the atan2 function, the vector-based EB-ESPRIT can resolve major problems of the conventional EB-ESPRIT. In this work, we showed that the vector-based EB-ESPRIT not only can estimate enough number of early echoes but also outperforms other algorithms in terms of estimation accuracy. Through the analysis of simulation and experimental data acquired from different reverberation conditions, we demonstrated that the estimation performance of the vector-based EB-ESPRIT outperforms other baseline algorithms in the reverberant environment.

6. ACKNOWLEDGEMENTS

This work was supported by the BK21 Plus program through the National Research Foundation (NRF) funded by the Ministry of Education of Korea.

7. REFERENCES

[1] B. Rafaely. *Fundamentals of spherical array processing*. Springer, 2015.

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

[3] H. Sun, E. Mabande, K. Kowalczyk, and W. Kellermann. Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing. *The Journal of the Acoustical Society of America*, 131(4):2828–2840, 2012.

[4] S. Tervo and A. Politis. Direction of arrival estimation of reflections from room impulse responses using a spherical microphone array. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 23(10):1539–1551, 2015.

[5] E. Mabande, K. Kowalczyk, H. Sun, and W. Kellermann. Room geometry inference based on spherical microphone array eigenbeam processing. *The Journal of the Acoustical Society of America*, 134(4):2773–2789, 2013.

[6] H. Teutsch and W. Kellermann. Eigen-beam processing for direction-of-arrival estimation using spherical apertures. In *Proc. Joint Workshop on Hands-Free Communication and Microphone Arrays*, 2005.

[7] 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 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 117–120. IEEE, 2011.

[8] Q. Huang, L. Zhang, and F. Yong. Two-step spherical harmonics esprit-type algorithms and performance analysis. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2018.

[9] B. Jo and J.-W. Choi. Direction of arrival estimation using nonsingular spherical esprit. *The Journal of the Acoustical Society of America*, 143(3):EL181–EL187, 2018.

[10] B. Jo and J.-W. Choi. Parametric direction-of-arrival estimation with three recurrence relations of spherical harmonics. *The Journal of the Acoustical Society of America*, 145(1):480–488, 2019.

[11] D. Khaykin and B. Rafaely. Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 221–224, 2009.

[12] N. Huleihel and B. Rafaely. Spherical array processing for acoustic analysis using room impulse responses and time-domain smoothing. *The Journal of the Acoustical Society of America*, 133(6):3995–4007, 2013.

[13] H. Morgenstern and B. Rafaely. Modal smoothing for analysis of room reflections measured with spherical microphone and loudspeaker arrays. *The Journal of the Acoustical Society of America*, 143(2):1008–1018, 2018.

[14] B. Jo and J.-W. Choi. Spherical harmonic smoothing for localizing coherent sound sources. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(10):1969–1984, 2017.

[15] C.H. Choi, J. Ivanic, M.S. Gordon, and K. Ruedenberg. Rapid and stable determination of rotation matrices between spherical harmonics by direct recursion. *The Journal of Chemical Physics*, 111(19):8825–8831, 1999.

[16] D.P. Jarrett, E.A.P. Habets, M.R.P. Thomas, and P.A. Naylor. Rigid sphere room impulse response simulation: Algorithm and applications. *The Journal of the Acoustical Society of America*, 132(3):1462–1472, 2012.

[17] B. Bernschütz, C. Pörschmann, S. Spors, S. Weinzierl, and B. der Verstärkung. Soft-limiting der modalen amplitudenverstärkung bei sphärischen mikrofonarrays im plane wave decomposition verfahren. In *Deutsche Jahrestagung für Akustik (DAGA)*, pages 661–662, 2011.