Applied Acoustics 136 (2018) 41–47

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Fast direction-of-arrival estimation algorithm for multiple wideband acoustic sources using multiple open spherical arrays

Xi Pan, Huayang Wang*, Zhiyi Lou, Yongchao Su

School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China

ARTICLE INFO

Keywords:
Direction of arrival(DOA)
Multiple wideband acoustic sources
Orthonormal propagator method (OPM)
Spherical harmonic (SH)
Spherical microphone arrays (SMAs)

ABSTRACT

The orthonormal propagator method (OPM) is extended to the spherical harmonic (SH) domain using multiple open spherical microphone arrays (SMAs). Compared with the multiple signal classification method (MUSIC), the computational loads of the OPM can be lower (by one or two orders), because the OPM does not require eigendecomposition of the cross spectral matrix (CSM) of the received sensor signals. Moreover, multiple open SMAs have wider frequency ranges of operation because their large radii reduce their lower frequency bounds and their small radii increase their higher frequency bounds. In this work, the performances of the OPM using both the spherical Fourier transform (SFT) components in the SH domain (SH-SFT-OPM) and the cross spectral matrix of the SFT components in the SH domain (SH-CSM-OPM) are analyzed theoretically and then evaluated in terms of their root-mean-square errors under various signal-to-noise ratio (SNR) conditions and direction-of-arrival estimation snapshots. The performance results are then compared with those of the SH-MUSIC. The simulation results confirm that the SH-CSM-OPM performs similarly to the SH-MUSIC, even for low SNRs, and that the computational complexity of the SH-SFT-OPM is significantly lower than that of the SH-MUSIC.

1. Introduction

Spherical microphone arrays (SMAs) have been used in various applications, including sound-field reproduction [1–4], beamforming [5–7], and direction-of-arrival (DOA) estimation of sound sources [8–12]. Subspace-based methods in the spherical harmonic (SH) domain, a DOA estimation method using SMAs, and the division of observation space into a noise subspace and a source subspace have all demonstrated high resolution capabilities in a three-dimensional space [8,10,11]. Subspace-based methods in the SH domain all use the eigendecomposition of the cross-spectral matrix (CSM) of the SH coefficients, which is obtained using the spherical Fourier transform (SFT). One of the most famous of these high-resolution methods—SH multiple signal classification (SH-MUSIC)—has shown significant improvements over MUSIC alone [13]. However, obtaining the eigenvectors is computationally expensive when the number of microphones is large. The orthonormal propagator method (OPM) is a subspace-based method that does not require the eigendecomposition of the CSM of the received sensor signals [14], and the propagator is a linear operator that depends only on the steering vectors. Thus, the OPM can significantly reduce the computational complexity.

There are two main methods for the DOA estimation of wideband signals. One is the incoherent signal subspace method (ISSM), and the other is the coherent signal subspace method (CSSM). The key idea of the ISSM is to decompose wideband signals into narrowband sectors and apply narrowband techniques to each sector independently. After then, the results from all of the frequency sectors are averaged [15]. This method, though simple, suffers from performance degradation when the signal-to-noise ratio (SNR) varies at each frequency, and it requires a DOA estimation for each frequency sector, which results in a high computational cost. The CSSM transforms the CSMs of many frequency sectors into one general CSM at a single focusing frequency using a focusing matrix that depends on the frequency sector [16]. The drawbacks of CSSM are that these focusing matrices require initial DOA values and that the estimation performance is very sensitive to these initial values. However, the SFT using SMAs has the advantage that the frequency-dependent components are decoupled from the angular-dependent components [10]. As a result, the sound source DOA is no longer dominated by the frequency of the source. By removing the frequency-dependent components, the DOA estimation methods can be applied to narrowband signals without requiring either focusing matrices or knowledge of the initial DOA values. Thus, by using the SFT together with the frequency removal techniques, both a better DOA estimation performance and a lower computational complexity can be achieved.

For open spheres, the modal strength contains poor numerical

* Corresponding author at: Beijing Institute of Technology, 3rd Building 5, South Zhongguancun Street Haidian District, Beijing 100081, China.
E-mail address: wanghuayang@bit.edu.cn (H. Wang).

https://doi.org/10.1016/j.apacoust.2018.01.014
Received 16 July 2017; Received in revised form 8 November 2017; Accepted 14 January 2018
Available online 22 February 2018
0003-682X/

conditioning at the fixed frequencies [5], which causes excessive noise at these frequencies. To overcome this problem, a method using multiple open SMAs (MOSMAs) is proposed. By choosing an appropriate radius ratio for a MOSMA, not only can numerically robust open SMAs be obtained, but the frequency range of operation can be increased as well, which means that a smaller radius increases the upper frequency range bound and a larger radius increases the lower frequency range bound.

To perform a fast DOA estimation for multiple wideband acoustic sources, the proposed method incorporates a wider frequency band of operation with the faster abilities of the OPM method in the SH domain by using MOSMAs. The remainder of this paper is organized as follows. In Section 2, the MOSMA model is presented. Section 3, the proposed DOA algorithm, using both the SH-SFT-OPM and the SH-CSM-OPM for multiple wideband acoustic sources, is derived, and its computational complexity is analyzed. In Section 4.1, the performance of the MOSMA is discussed, and the frequency ranges of operation for various dual open sphere radii are presented for the given number of microphones. In Section Section 4.2, the performances of the SH-SFT-OPM, the SH-CSM-OPM, and the SH-MUSIC are analyzed theoretically and simulated in terms of their RMSEs for both various SNRs and various DOA estimation snapshots. Finally, Section Section 5 concludes the paper.

Throughout this paper, vectors are denoted by lower-case bold letters, and matrices are represented by upper-case bold letters. The superscripts $_T$, $_*$, and $_H$ denote transpose, complex conjugate, and conjugate-transpose, respectively.

2. Multiple open SMA model

To simplify both the description and the physical construction of the algorithm, we first restrict the problem to that of a dual open SMA, as shown in Fig. 1, composed of two open spheres: one with a radius of r_1 and another with a radius of $r_2 = \rho r_1$, where ρ is the ratio of the radii of the two spheres. Distributed evenly over each of the surfaces of these two spheres are L omnidirectional pressure microphones with positions of $\Omega_l = (\theta_l,\phi_l)$ $(l = 1, 2 \dots L)$. The position vector of the lth microphone of the dth sphere, in Cartesian coordinates, is $\boldsymbol{r}_{dl} = (r_{dl}\sin\theta_l\cos\phi_l, r_{dl}\sin\theta_l\sin\phi_l, r_{dl}\cos\theta_l)^T$ $(d = 1, 2)$, where r_{dl} is the distance from the origin to that microphone. The azimuth angle ϕ is measured counterclockwise from the x-axis, and the elevation angle θ is measured down from the z-axis.

Consider S plane waves with the wave number k impinging on the array and with $k = 2\pi f/c$, where f is the frequency, c is the speed of sound, and the DOA of the sth plane wave is $\Omega_s = (\theta_s,\phi_s)$ $(s = 1, 2 \dots S)$. Then, the wavenumber vector of the sth propagating plane wave is denoted by $\boldsymbol{k}_s = -(k_s\sin\theta_s\cos\phi_s, k_s\sin\theta_s\sin\phi_s, k_s\cos\theta_s)^T$. In the spatial

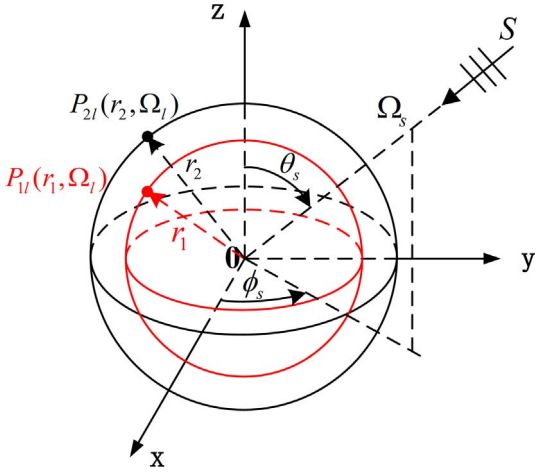

Fig. 1. Illustration of the dual spherical array geometry.

domain, the sound pressure at each of the L microphones on the two SMAs can be written as

$$\boldsymbol{x}_1(k) = \boldsymbol{A}_1(k)\boldsymbol{s}(k) + \boldsymbol{n}_1(k), \tag{1}$$

$$\boldsymbol{x}_2(k) = \boldsymbol{A}_2(k)\boldsymbol{s}(k) + \boldsymbol{n}_2(k), \tag{2}$$

where $\boldsymbol{x}_d(k) = [x_{d1}(k), x_{d2}(k), \cdots x_{dL}(k)]^T \in C^{L\times 1}$ represents the observation vectors of the two SMAs, respectively; $\boldsymbol{s}(k) = [s_1(k), s_2(k), \cdots s_S(k)]^T \in C^{S\times 1}$ represents the signal waveform vectors, and $\boldsymbol{n}_d(k) = [n_{d1}(k), n_{d2}(k), \cdots, n_{dL}(k)]^T \in C^{L\times 1}$ represents the respective additive noise vectors that are both uncorrelated with the signal vector $\boldsymbol{s}(k)$, $\boldsymbol{A}_d(k) \in C^{L\times S}$ is the steering matrix, and $\boldsymbol{A}_d(k)$ can be written as

$$\begin{aligned}\boldsymbol{A}_d(k) &= \begin{bmatrix} e^{-i\boldsymbol{k}_1^T\boldsymbol{r}_{d1}} & e^{-i\boldsymbol{k}_2^T\boldsymbol{r}_{d1}} & \dots & e^{-i\boldsymbol{k}_S^T\boldsymbol{r}_{d1}} \\ e^{-i\boldsymbol{k}_1^T\boldsymbol{r}_{d2}} & e^{-i\boldsymbol{k}_2^T\boldsymbol{r}_{d2}} & \dots & e^{-i\boldsymbol{k}_S^T\boldsymbol{r}_{d2}} \\ \vdots & \vdots & & \vdots \\ e^{-i\boldsymbol{k}_1^T\boldsymbol{r}_{dL}} & e^{-i\boldsymbol{k}_2^T\boldsymbol{r}_{dL}} & \dots & e^{-i\boldsymbol{k}_S^T\boldsymbol{r}_{dL}} \end{bmatrix} \\ &= [\boldsymbol{a}_{d1}(k) \quad \boldsymbol{a}_{d2}(k) \quad \cdots \quad \boldsymbol{a}_{dS}(k)]\end{aligned} \tag{3}$$

where $\boldsymbol{a}_{ds}(k) \in C^{L\times 1}$ is the steering vector, with

$$\boldsymbol{a}_{ds}(k) = \left[e^{-i\boldsymbol{k}_s^T\boldsymbol{r}_{d1}}, e^{-i\boldsymbol{k}_s^T\boldsymbol{r}_{d2}}, \cdots, e^{-i\boldsymbol{k}_s^T\boldsymbol{r}_{dL}}\right]^T, \tag{4}$$

Here, represents the pressure of the sth unit plane wave that is incident from the direction Ω_s on the lth microphone of the dth SMA. In the SH domain, this can be expressed as [17]

$$e^{-i\boldsymbol{k}_s^T\boldsymbol{r}_{dl}} \approx \sum_{n=0}^{N}\sum_{m=-n}^{n} b_n(kr_d)Y_{nm}^*(\Omega_s)Y_{nm}(\Omega_l), \tag{5}$$

where $i = \sqrt{-1}$; n is the order, ranging from 0 to N; N is the highest SH order, which is determined by the sampling set Ω_l of the microphones on the sphere [1]; m is the degree, ranging from $-n$ to n; and $b_n(kr_d) = 4\pi i^n \times j_n(kr_d)$ is the far-field modal strength for open SMAs, where $j_n(kr_d)$ is the spherical Bessel function. Fig. 2(a) shows the magnitude of b_n as a function of f with $r_d = 0.1$ m for several orders n. Y_{nm} is the SH of both the order n and degree m, defined as

$$Y_{nm}(\Omega) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}P_{nm}(\cos\theta)e^{im\phi}, \tag{6}$$

where $P_{nm}(\cos\theta)$ is the associated Legendre function. According to (5), the steering matrix $\boldsymbol{A}_d(k)$ can be expressed as

$$\boldsymbol{A}_d(k) = \boldsymbol{Y}(\Omega_L)\boldsymbol{B}(kr_d)\boldsymbol{Y}^H(\Omega_S), \tag{7}$$

where $\boldsymbol{Y}(\Omega_L) \in C^{L\times(N+1)^2}$ is an SH matrix given by

$$\boldsymbol{Y}(\Omega_L) = \begin{bmatrix} Y_{00}(\Omega_1) & Y_{1-1}(\Omega_1) & Y_{10}(\Omega_1) & Y_{11}(\Omega_1) & \cdots & Y_{NN}(\Omega_1) \\ Y_{00}(\Omega_2) & Y_{1-1}(\Omega_2) & Y_{10}(\Omega_2) & Y_{11}(\Omega_2) & \cdots & Y_{NN}(\Omega_2) \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ Y_{00}(\Omega_L) & Y_{1-1}(\Omega_L) & Y_{10}(\Omega_L) & Y_{11}(\Omega_L) & \cdots & Y_{NN}(\Omega_L) \end{bmatrix}, \tag{8}$$

$\boldsymbol{B}(kr_d) \in C^{(N+1)^2\times(N+1)^2}$ is a far-field modal strength diagonal matrix written as

$$\boldsymbol{B}(kr_d) = diag(b_0(kr_d), b_1(kr_d), b_1(kr_d), \cdots, b_N(kr_d)), \tag{9}$$

and $\boldsymbol{Y}(\Omega_S) \in C^{S\times(N+1)^2}$ is an SH matrix containing the source signal directions, which can be expressed as

$$\boldsymbol{Y}(\Omega_S) = \begin{bmatrix} Y_{00}(\Omega_1) & Y_{1-1}(\Omega_1) & Y_{10}(\Omega_1) & Y_{11}(\Omega_1) & \cdots & Y_{NN}(\Omega_1) \\ Y_{00}(\Omega_2) & Y_{1-1}(\Omega_2) & Y_{10}(\Omega_2) & Y_{11}(\Omega_2) & \cdots & Y_{NN}(\Omega_2) \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ Y_{00}(\Omega_S) & Y_{1-1}(\Omega_S) & Y_{10}(\Omega_S) & Y_{11}(\Omega_S) & \cdots & Y_{NN}(\Omega_S) \end{bmatrix}, \tag{10}$$

Substituting (7) into both (1) and (2) and multiplying each of the results from the left by $\boldsymbol{Y}^H(\Omega_L) \times \boldsymbol{V}$ provides the SFT coefficient model of the observation signal, which is given as follows:

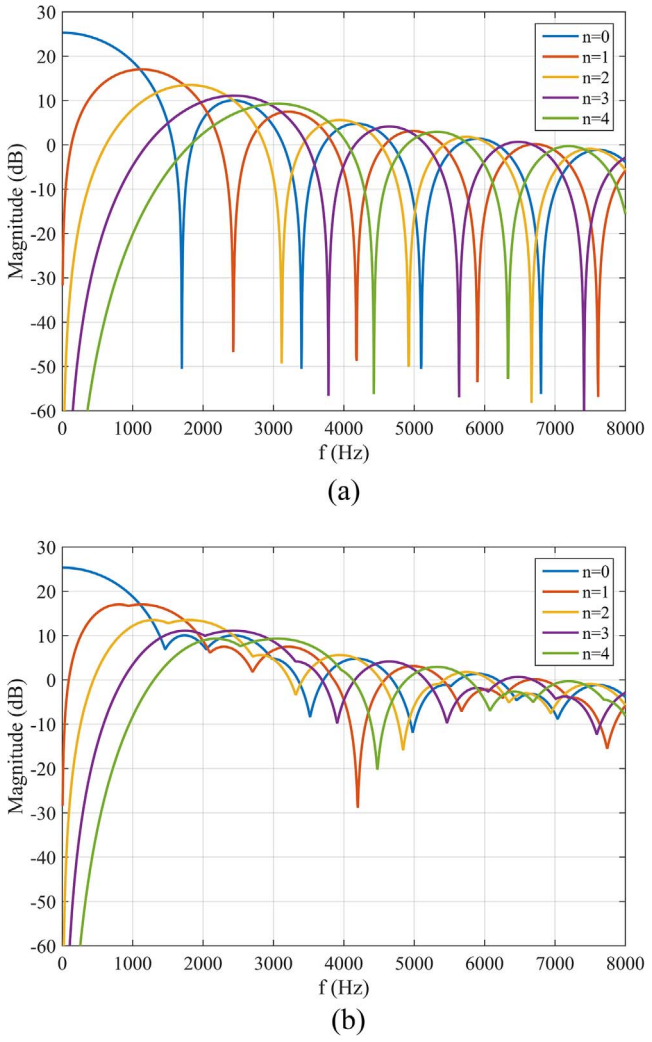

Fig. 2. Magnitude of b_n for $n = 0...4$ as a function of f. (a) Single open spherical arrays with $r = 0.1$ m and (b) dual open spherical arrays with $r = 0.1$ m& 0.14 m.

$$\boldsymbol{Y}^H(\Omega_L) \times \boldsymbol{V} \times \boldsymbol{x}_d(k) = \boldsymbol{Y}^H(\Omega_L) \times \boldsymbol{V} \times \boldsymbol{Y}(\Omega_L)\boldsymbol{B}(kr_d)\boldsymbol{Y}^H(\Omega_s)\boldsymbol{s}(k) + \boldsymbol{n}_{dnm}(k), \tag{11}$$

where $\boldsymbol{Y}^H(\Omega_L) \times \boldsymbol{V} \times \boldsymbol{Y}(\Omega_L) = \boldsymbol{I}_{(N+1)^2}$ [1], with $\boldsymbol{I}_{(N+1)^2} \in C^{(N+1)^2\times(N+1)^2}$ denoting the identity matrix and $\boldsymbol{V} \in C^{L\times L}$. $\boldsymbol{V} = diag(\alpha_1,\alpha_2,\cdots,\alpha_L)$ represents the sampling weights. Thus, the SFT coefficient model of the observation signal can be rewritten as

$$\boldsymbol{x}_{dnm}(k) = \boldsymbol{B}(kr_d)\boldsymbol{Y}^H(\Omega_s)\boldsymbol{s}(k) + \boldsymbol{n}_{dnm}(k), \tag{12}$$

where $\boldsymbol{x}_{dnm}(k) \in C^{(N+1)^2\times 1}$ represents the SFT coefficient vectors of the two SMAs, respectively, $\boldsymbol{n}_{dnm}(k) \in C^{(N+1)^2\times 1}$ represents the respective additive noise SFT coefficient vectors, and $\boldsymbol{n}_{dnm}(k) = \boldsymbol{Y}^H(\Omega_L) \times \boldsymbol{V} \times \boldsymbol{n}_d(k)$.

For multiple open spheres, to avoid using a $b_n(kr)$ with small values, the data set with the largest value of $b_n(kr)$ at each wave number k and each order n, $b_n(kr_d)$ in (5), is rewritten as

$$b_n(kr) = 4\pi i^n \times \max[|j_n(kr_1)|,|j_n(kr_2)|,\cdots,|j_n(kr_D)|], \tag{13}$$

where max[•] is used to obtain the maximum elements of an array, and r representing any value from r_1 to r_D, is given as

$$r = r_d,\ |j_n(kr_d)| = \max[|j_n(kr_1)|,|j_n(kr_2)|,\cdots,|j_n(kr_D)|]. \tag{14}$$

Fig. 2(b) shows the magnitude of b_n for the dual open spherical arrays with $r = 0.1$ m and 0.14 m. This figure illustrates that zeroes are avoided in b_n. Using (13), the combination of the SFT coefficient model from the two open spheres is given by

$$\boldsymbol{x}_{nm}(k) = \boldsymbol{B}(kr)\boldsymbol{Y}^H(\Omega_s)\boldsymbol{s}(k) + \boldsymbol{n}_{nm}(k), \tag{15}$$

where is the combined SFT coefficient vector of the two SMAs. The elements of $\boldsymbol{B}(kr)$ consist of $b_n(kr)$ from (13) and $\boldsymbol{n}_{nm}(k) = \boldsymbol{Y}^H(\Omega_L) \times \boldsymbol{V} \times \boldsymbol{n}(k)$, where $\boldsymbol{n}(k)$ is the additive noise vector. The steering matrix of the SFT coefficients $\boldsymbol{B}(kr)\boldsymbol{Y}^H(\Omega_s)$ contains both frequency and angular information. Therefore, to decouple the frequency-dependent components from the angular-dependent components, we simply multiply (15) from the left by $\boldsymbol{B}^{-1}(kr)$ to obtain the SFT components $\boldsymbol{P}_{nm} \in C^{(N+1)^2\times 1}$, which are expressed as [10]

$$\boldsymbol{P}_{nm}(k) = \boldsymbol{Y}^H(\Omega_s)\boldsymbol{s}(k) + \boldsymbol{N}(k), \tag{16}$$

where $\boldsymbol{N}(k) \in C^{(N+1)^2\times 1}$ and

$$\boldsymbol{N}(k) = \boldsymbol{B}^{-1}(kr) \times \boldsymbol{Y}^H(\Omega_L) \times \boldsymbol{V} \times \boldsymbol{n}(k), \tag{17}$$

If we define the matrix $\boldsymbol{F}(k) \in C^{(N+1)^2\times L}$ as

$$\boldsymbol{F}(k) = \boldsymbol{B}^{-1}(kr) \times \boldsymbol{Y}^H(\Omega_L) \times \boldsymbol{V}, \tag{18}$$

$\boldsymbol{N}(k)$ can be rewritten as

$$\boldsymbol{N}(k) = \boldsymbol{F}(k) \times \boldsymbol{n}(k). \tag{19}$$

3. Proposed DOA estimation method

The SFT components in (16) can be used to perform the DOA estimation using the OPM through the following processes. There are methods for obtaining the propagator $\boldsymbol{P}$: one using the SFT components and one using the CSM of these components. In this section, both the OPM using the SFT components in the SH domain (SH-SFT-OPM) and the OPM using the CSM of the SFT components in SH domain (SH-CSM-OPM) are deduced.

First, in practical applications of the wideband case, the frequency value f represents the whole frequency band of the sound source. According to the equation $k = 2\pi f/c$, which represents the ranges of k based on the source bandwidth, we assume that there are X frequency sectors, so that the SFT components for the wideband $\boldsymbol{P}_{nm} \in C^{(N+1)^2\times X}$ can be obtained from $\boldsymbol{P}_{nm}(k)$ in (16), which is given as

$$\boldsymbol{P}_{nm} = [\boldsymbol{P}_{nm}(k_1),\boldsymbol{P}_{nm}(k_2),...\boldsymbol{P}_{nm}(k_X)], \tag{20}$$

Thus, the CSM $\boldsymbol{R}_{nm} \in C^{(N+1)^2\times(N+1)^2}$ of the $\boldsymbol{P}_{nm}$ for the X frequency sectors can be estimated as

$$\begin{aligned}\overline{\boldsymbol{R}}_{nm} &= \tfrac{1}{X}\boldsymbol{P}_{nm}\boldsymbol{P}_{nm}^H \\ &= \boldsymbol{Y}^H(\Omega_s)\overline{\boldsymbol{R}}_s\boldsymbol{Y}(\Omega_s) + \sigma^2\overline{\boldsymbol{R}}_N\end{aligned} \tag{21}$$

where σ^2 is the variance of the noise, and $\overline{\boldsymbol{R}}_s \in C^{S\times S}$ and $\overline{\boldsymbol{R}}_N \in C^{(N+1)^2\times(N+1)^2}$ are the signal CSM and the noise CSM, respectively, which are defined as

$$\overline{\boldsymbol{R}}_s = \frac{1}{X}\sum_{x=1}^{X} \boldsymbol{s}(k_x)\boldsymbol{s}^H(k_x), \tag{22}$$

$$\overline{\boldsymbol{R}}_N = \frac{1}{X}\sum_{x=1}^{X} \boldsymbol{F}(k_x)\boldsymbol{F}^H(k_x), \tag{23}$$

The noise CSM $\overline{\boldsymbol{R}}_N$ is not spatially white. To use the propagator method, whitening must be performed. The whitening CSM $\overline{\boldsymbol{R}}_{nm}^w \in C^{(N+1)^2\times(N+1)^2}$ can be obtained as follows:

$$\begin{aligned}\overline{\boldsymbol{R}}_{nm}^w &= \overline{\boldsymbol{R}}_N^{-1/2}\overline{\boldsymbol{R}}_{nm}\overline{\boldsymbol{R}}_N^{-1/2} \\ &= \overline{\boldsymbol{R}}_N^{-1/2}\boldsymbol{Y}^H(\Omega_s)\overline{\boldsymbol{R}}_s\boldsymbol{Y}(\Omega_s)\overline{\boldsymbol{R}}_N^{-1/2} + \sigma^2\boldsymbol{I}\end{aligned} \tag{24}$$

Then, the definition of the propagator is based on both the partition of the SFT components $\boldsymbol{P}_{nm}$ and the CSM $\overline{\boldsymbol{R}}_{nm}^w$ of $\boldsymbol{P}_{nm}$, according to

$$\boldsymbol{P}_{nm} = [\boldsymbol{P}_A;\boldsymbol{P}_B], \tag{25}$$

$$\overline{\boldsymbol{R}}_{nm}^{w} = [\boldsymbol{G},\boldsymbol{H}], \tag{26}$$

where $\boldsymbol{P}_A \in C^{S\times X}$ denotes the sub-matrix consisting of the first S rows of the matrix $\boldsymbol{P}_{nm}$, and $\boldsymbol{P}_B \in C^{U\times X}$, with $U = (N+1)^2 - S$ representing the rest of the rows of matrix $\boldsymbol{P}_{nm}$. To guarantee $U > 0$, it is necessary that $S < (N+1)^2$. $\boldsymbol{G} \in C^{(N+1)^2\times S}$ denotes the sub-matrix consisting of the first S columns of the matrix $\overline{\boldsymbol{R}}_{nm}^{w}$, and $\boldsymbol{H} \in C^{(N+1)^2\times U}$ represents the rest of the columns of the matrix $\overline{\boldsymbol{R}}_{nm}^{w}$. Assuming that no two signals are coherent, the full-rank properties of both matrices $\boldsymbol{P}_{nm}$ and $\overline{\boldsymbol{R}}_{nm}^{w}$ indicate that the rows of $\boldsymbol{P}_A$ and the columns of $\boldsymbol{G}$ are linearly independent. In the no-noise case, there exists a matrix $\boldsymbol{P}$ —the propagator—that satisfies the following equations:

$$\boldsymbol{P}_B = \boldsymbol{P}^H \boldsymbol{P}_A, \tag{27}$$

$$\boldsymbol{H} = \boldsymbol{G}\boldsymbol{P}, \tag{28}$$

In the presence of noise, although the partitions of $\boldsymbol{P}_{nm}$ and $\overline{\boldsymbol{R}}_{nm}^{w}$ still hold, relations (27) and (28) no longer do. The propagators $\hat{\boldsymbol{P}}_{SFT}$ and $\hat{\boldsymbol{P}}_{csm}$ can be obtained via the least-squares method as follows:

$$\hat{\boldsymbol{P}}_{SFT} = (\boldsymbol{P}_A \boldsymbol{P}_A^H)^{-1} \boldsymbol{P}_A \boldsymbol{P}_B^H, \tag{29}$$

$$\hat{\boldsymbol{P}}_{csm} = (\boldsymbol{G}^H \boldsymbol{G})^{-1} \boldsymbol{G}^H \boldsymbol{H}, \tag{30}$$

Then, either $\hat{\boldsymbol{P}}_{SFT}$ or $\hat{\boldsymbol{P}}_{csm}$ is used to construct $\boldsymbol{Q} = [\hat{\boldsymbol{P}}, -\boldsymbol{I}_{(N+1)^2-S}]^T$ and set either $\boldsymbol{Q}^H \boldsymbol{P}_{nm} = 0$ or $\boldsymbol{Q}^H \overline{\boldsymbol{R}}_{nm}^{w} = 0$. Thus, the vector space $\boldsymbol{Q}$ is orthogonal to the direction vectors $\boldsymbol{a}(\Omega_s)$, where $\boldsymbol{a}(\Omega_s)$ is the steering vector that is a column of $\boldsymbol{Y}^H(\Omega_s)$ or $\overline{\boldsymbol{R}}_N^{-1/2} \boldsymbol{Y}^H(\Omega_s)$, and

$$\boldsymbol{Q}^H \boldsymbol{a}(\Omega_s) = 0, \tag{31}$$

Therefore, we can construct the following DOA estimator:

$$\boldsymbol{F}_{PM}(\Omega_s) = \boldsymbol{a}^H(\Omega_s) \boldsymbol{Q}\boldsymbol{Q}^H \boldsymbol{a}(\Omega_s), \tag{32}$$

To introduce a projection operator into the noise subspace, we can replace the matrix Q with its orthonormalized version:

$$\boldsymbol{Q}_0 = \boldsymbol{Q}(\boldsymbol{Q}^H \boldsymbol{Q})^{-1/2}, \tag{33}$$

Then, we can obtain the following pseudo-spectrum:

$$\boldsymbol{F}_{OPM}(\Omega_s) = \frac{1}{\boldsymbol{a}^H(\Omega_s) \boldsymbol{Q}_0 \boldsymbol{Q}_0^H \boldsymbol{a}(\Omega_s)}. \tag{34}$$

By scanning Ω_s, the largest peaks of the pseudo-spectrum indicate the directions of the source.

3.1. Computational cost

A major advantage of the propagator method is that it does not require eigendecomposition of the CSM. The computational gain of the OPM over the MUSIC is on the order of $O(S/L)$ [14]. Thus, the computational gain of the SH-CSM-OPM over the SH-MUSIC is on the order of $O(S/(N+1)^2)$. However, the SH-SFT-OPM does not require the CSM of the SFT components, which results in a lower computational complexity.

4. Simulation

In this section, the simulation is divided into two parts: one for the SMA performance and the other for the DOA estimation performance. The effects of both the spherical array structure parameters and the frequency range of the acoustic source signal on the SMA performance are mainly discussed in accordance with a stochastic process. Subsequently, the time cost of the MATLAB-based DOA estimation as well as the DOA estimation performance under various SNRs and various snapshots are investigated through calculations and simulations. All of the simulations are performed using MATLAB 2015b running on an Intel i5, 2.50 GHz processor with 8 GB of memory under the Win10 operating system.

4.1. SMA performance

The SMAs are employed in the estimation of the DOA. Therefore, the performance of the DOA estimation depends primarily on how accurately the SFT components $\boldsymbol{P}_{nm}(k)$ can be obtained from the received microphone signals as functions of the frequency. In this section, we evaluate the performance of the SMA using the SNR of the SFT components (SNR-SFT) and analyze the factors that limit it.

Taking into account errors from various sources, we evaluated the performance of an SMA via a stochastic process. For a given microphone array design, L sets of microphone positions were created randomly with a given average positioning error. The pressure received by the microphones of these L sets of SMAs was then calculated for S plane-wave directions, and random noise was added to the pressure values. The SNR-SFT using the SMA can be expressed as

$$SNR\text{-}SFT_{(dB)} = 10\log_{10}\left(\frac{\|\boldsymbol{P}_{nm}(k)\|_F^2}{\|\boldsymbol{P}_{nm}(k) - \hat{\boldsymbol{P}}_{nm}(k)\|_F^2}\right), \tag{35}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, and $\boldsymbol{P}_{nm}(k)$ and $\hat{\boldsymbol{P}}_{nm}(k)$ are the theoretical estimated values of the SFT components, respectively, and can be written as

$$\boldsymbol{P}_{nm}(k) = \boldsymbol{Y}^H(\Omega_s) \boldsymbol{s}(k), \tag{36}$$

$$\hat{\boldsymbol{P}}_{nm}(k) = \boldsymbol{F}(k) \boldsymbol{A}(k) \boldsymbol{s}(k) + \boldsymbol{F}(k) \boldsymbol{n}(k). \tag{37}$$

The SNR-SFT is determined by four main factors. The first factor is the SNR of the received microphone signals, which determines the values of $\boldsymbol{n}(k)$. The function $\boldsymbol{F}(k)$ is determined by both the modal strength b_n and the highest SH order N, and these two factors are each determined by the configuration of the SMAs. The fourth factor is the rate of change of frequency (RCF) of the sound source signals, which controls the values of $\boldsymbol{s}(k)$. That is, a higher RCF yields a wider frequency band of the sound source signals for both the given number of snapshots and the given sampling frequency, which can influence the magnitude of the SNR-SFT.

Fig. 3 shows the simulated SNR-SFT corresponding to three different SMA configurations, as shown in Table 1. Array 1 used a single open sphere with $r = 0.1m$, and Array 2 used a single open sphere with $r = 0.14m$. Array 3 used a dual open sphere with $r = 0.1m$ and $0.14m$. On the surface of each sphere were 32 evenly distributed omnidirectional pressure microphones with positions of $\Omega_l = (\theta_l, \phi_l)$ $(l = 1,2...32)$, which were the same as the positions of the microphones in the em32 Eigenmike® microphone array [18]. The average position error of the sensor was set to 3% of the sphere radius, which approximately

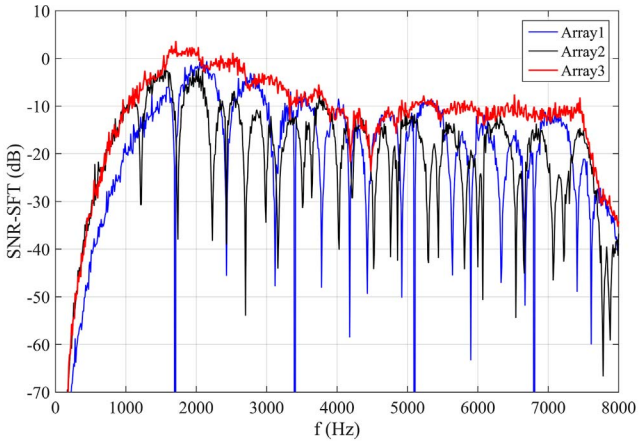

Fig. 3. Magnitude of SNR-SFT for $N = 4$ as a function of f. The different arrays are indicated by the line color. Blue line: Array 1 using a single open sphere with $r = 0.1$ m; Black line: Array 2 using a single open sphere with $r = 0.14$ m; Red line: Array 3 using a dual open sphere with $r = 0.1$ m& 0.14 m. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Details of the three arrays for the simulation.

Array name	Configuration	Radius	Order	D × L
Array 1	Single open sphere	0.1 m	4	1 × 32
Array 2	Single open sphere	0.14 m	4	1 × 32
Array 3	Dual open sphere	0.1 m & 0.14 m	4	2 × 32

corresponds to 0.3 mm when r is 0.1 m. The highest SH order was $N = 4$ and was calculated with $N = \lfloor\sqrt{L}-1\rfloor$ [1], where $\lfloor\cdot\rfloor$ denotes the floor function.

For these three simulations, three incoherent far-field sine signals swept linearly in the frequency range of 0.1–9.1 kHz at 10 kHz/s RCF were used, and their directions were (120°,90°), (90°,180°), and (60°,270°), respectively. The SNRs of the received microphone signals were set as 0 dB. The number of snapshots was 2,048, and the sampling frequency was set as 20.48 kHz. The snapshots were composed of 900 frequency sectors and had a frequency resolution of 10 Hz.

As shown in Fig. 3, the performances of the three SMAs differed greatly. First, at low frequency values, the modal strength b_n was small, which resulted in significant errors in the SFT components $\boldsymbol{P}_{nm}$. That is, at low frequencies, some SFT components deviated from $\boldsymbol{P}_{nm}$. This phenomenon governs the lower bound of the frequency range of operation of the SMA. By comparing the values of Arrays 1 and 2, the SFT-SNR can be obtained. When the number of microphones is given, a larger SMA radius results in a lower bound.

Second, at high frequency values, the performance of the SMA is limited by spatial aliasing [19]. Spatial aliasing occurs mainly above the so-called spatial aliasing frequency, at which the distance between a microphone and its nearest neighbors is equal to a half-wavelength. In this frequency region, the higher-order SFT components are aliased into lower-order components. The spatial aliasing frequency can be considered the higher bound of the frequency range of operation of the SMA. This bound depends directly on the size of the array, and when the number of microphones is given, a smaller SMA radius results in a higher maximum frequency.

Finally, for Arrays 1 and 2 with single open spheres and frequencies f at which $b_n(kr)$ is equal to zero, the SNR-SFTs are ill-conditioned and small, and the array robustness is poor owing to, for example, errors in the microphone positioning or sensor self-noise. For Array 3, with dual open spherical arrays, the zeroes of $b_n(kr)$ can be avoided by selecting an optimal ratio of the two radii ρ that is close to the optimal value given by $\rho = 1 + \pi/2N$ [5]. Furthermore, the large radius reduces the lower bound, and the small radius increases the higher bound, which results in a wider SMA frequency range of operation.

According to the analysis of the DOA estimation performance under different SNRs in Section Section 4.2, we define the frequency range of operation of an array as that for which the SNR-SFT is greater than −5 dB. Given the configuration of the dual open sphere, which is similar to that of Array 3, the ratio ρ of the two radii is equal to 1.4. Considering the actual situation, the range of the radius r_1 extends from 0.05 to 0.5 m. The frequency ranges of operation for different radii are shown in Fig. 4, which were found using the same conditions as those in the simulation above. Fig. 4 shows that the small radius corresponds to the higher bound and the large radius corresponds to the lower bound, which is the same conclusion as the one drawn above. These results can be used to guide the size design of the SMAs for the sound source target of the given frequency range.

4.2. DOA estimation performance

To evaluate the DOA estimation performance, three incoherent far-field sine signals swept linearly in the frequency range of 1.2–3.2 kHz at 8 kHz/s RCF with directions of (120°,90°), (90°,180°), and (60°,270°) were used. The SNRs of the received microphone signals were set as

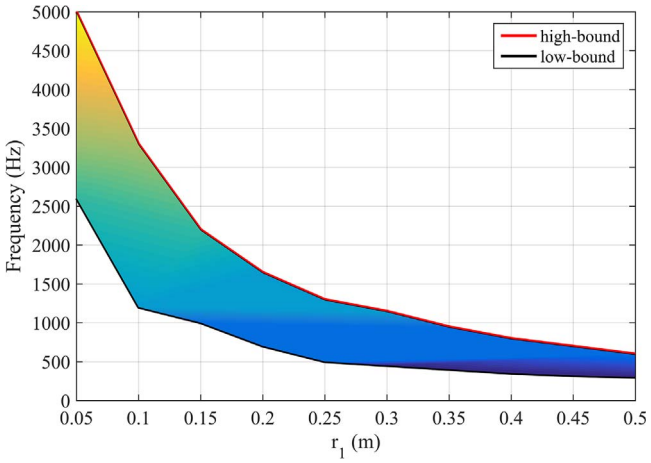

Fig. 4. Frequency range of operation with different radii r_1, using the dual open spherical array with 32 evenly distributed microphones on each sphere, and the ratio of the two radii $\rho = 1.4$.

0 dB, the number of snapshots was 2,048, and the sampling frequency was set to 20.48 kHz. Meanwhile, the frequency band was divided into 200 frequency sectors, resulting in a frequency resolution of 10 Hz.

First, we tested the time cost of the DOA estimation based on the SH-SFT-OPM, SH-CSM-OPM, and SH-MUSIC using Array 3. The grid intervals of elevation and azimuth were (5°,10°), (2°,4°), and (1°,2°). The average times for the ten trials are shown in Table 2.

Table 2 shows that the computational time increased quickly as the grid interval became smaller. The SH-SFT-OPM method required the least amount of time because it did not need to compute either the covariance matrix or perform whitening, while the SH-CSM-OPM and SH-MUSIC methods required almost the same amount of time. The spatial spectra of the DOA estimation corresponding to these three methods with elevation and azimuth grid intervals of (1°,2°) are shown in Fig. 5(a–c). By comparing the estimation results of various algorithms, the following conclusions can be drawn. The estimation accuracy of the SH-SFT-OPM was the worst, as shown in Fig. 5(a). Fig. 5(b) and (c) show that the estimation accuracies of the SH-CSM-OPM and the SH-MUSIC were similar and that their spectrum peaks were sharper than those of the SH-SFT-OPM.

A comparison of the DOA estimation performances of the three estimation algorithms was performed in terms of the RMSE, which can be written as

$$RMSE = \frac{1}{S}\sum_{s=1}^{S}\sqrt{\frac{1}{W}\sum_{w=1}^{W}(\hat{\theta}_{s,w}-\theta_s)^2 + (\hat{\phi}_{s,w}-\phi_s)^2}, \tag{38}$$

where S and W are the numbers of plane waves and independent Monte Carlo trials, respectively. $(\hat{\theta}_{s,w},\hat{\phi}_{s,w})$ represents the estimated values of the sth plane wave in the wth Monte Carlo trial with the impinging direction (θ_s,ϕ_s). In our simulation, the elevation and azimuth grid intervals were (1°,1°). The SNRs of the received microphone signals varied from −10 to 20 dB. The incident sound source signal parameters remained the same as those mentioned above. The RMSEs of 500 independent Monte Carlo trials for each of these three estimation

Table 2
Comparison of computation time.

Grid intervals of elevation and azimuth (degree)	Time (s)		
	SH-SFT-OPM	SH-CSM-OPM	SH-MUSIC
(5°,10°)	0.074	0.085	0.087
(2°,4°)	0.135	0.164	0.170
(1°,2°)	0.428	0.510	0.521

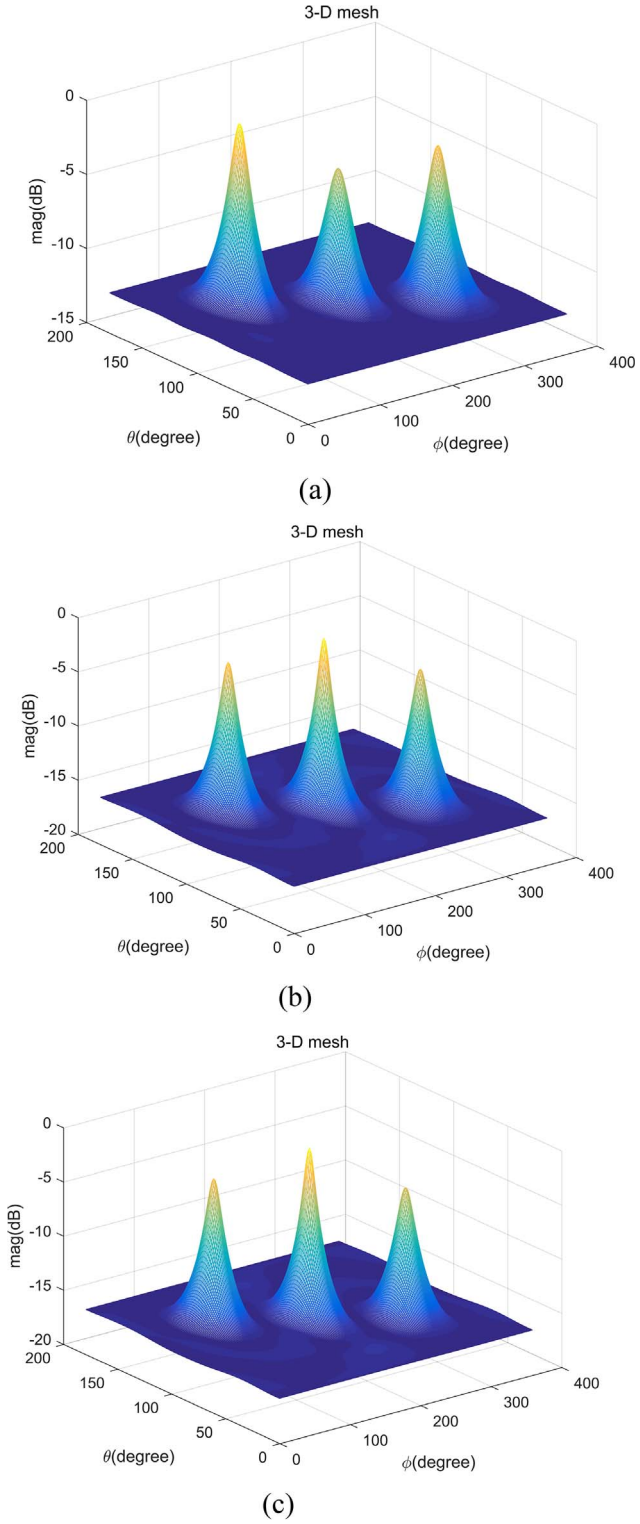

Fig. 5. Spatial spectrum of the DOA estimation of three sources at (120°,90°), (90°,180°), and (60°,270°) for Array 3: using the (a) SH-SFT-OPM, (b) SH-CSM-OPM, and (c) SH-MUSIC.

algorithms are shown in Fig. 6.

Fig. 6 shows that the RMSE of the SH-SFT-OPM was the worst. This was caused by the colored noise $\boldsymbol{N}(k)$. The other two algorithms yielded similar results, with SNRs between -5 and 20 dB and RMSEs of less than one degree. In summary, the SH-CSM-OPM, using a dual open spherical array, provided good estimation performances for wideband acoustic sources and did not require eigenvalue decomposition, which

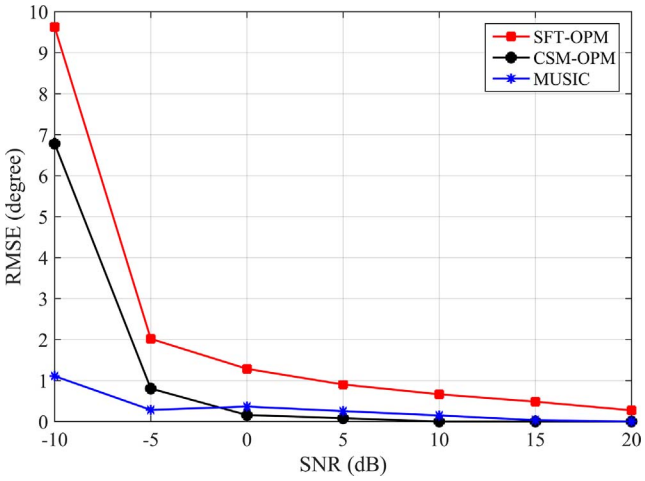

Fig. 6. DOA estimation performance of three sources at (120°,90°), (90°,180°), and (60°,270°) for Array 3. The figure shows the RMSE vs. the SNR based on the SH-SFT-OPM, SH-CSM-OPM, and SH-MUSIC.

greatly reduced the computational complexity.

In the case of a constant sampling rate, the length of the time window is determined by the number of snapshots, and a narrower time window results in fewer frequency sectors X and a lower frequency resolution. When the frequency range of the sound signals is not included in the X frequency sectors, the sound signals cannot be recognized. This is a limitation of the Fourier transform and is a factor in determining the performance of the proposed DOA algorithm.

To analyze the DOA estimation performance for various numbers of snapshots, corresponding simulation experiments were performed for seven different numbers of snapshots (64, 128, 256, 512, 1024, 2048, and 4096) with the rest of the parameters remaining the same as the values described in the first paragraph of this section. Each snapshot was subjected to 500 independent Monte Carlo experiments, and the RMSEs of various snapshots were calculated based on SH-SFT-OPM, SH-CSM-OPM, and SH-MUSIC, as shown in Fig. 7.

The results show that the RMSE increased rapidly when the number of snapshots was lower than 512, which is consistent with the conclusion that a lower number of snapshots yields worse DOA estimation performance. However, when the number of snapshots was greater than 1024, only the SH-CSM-OPM and SH-MUSIC algorithms followed the principle that a greater number of snapshots leads to a better estimation performance, while the performance of the SH-SFT-OPM algorithm

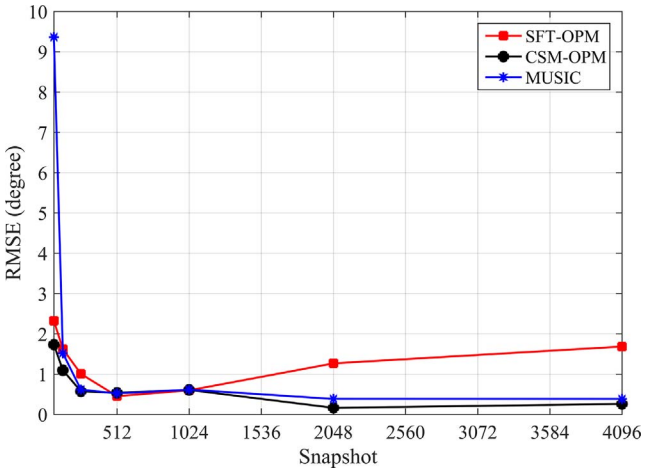

Fig. 7. DOA estimation performance of three sources at (120°,90°), (90°,180°), and (60°,270°) for Array 3. The figure shows the RMSE vs. the snapshot number based on SH-SFT-OPM, SH-CSM-OPM, and SH-MUSIC.

deteriorated when the number of snapshots increased. The reason for this was that the SH-SFT-OPM algorithm did not whiten the colored noise $\boldsymbol{N}(k)$ that was introduced during both the SFT and the frequency-component removal. Therefore, with an increase in the number of snapshots, colored noise accumulated, and the SFT-OPM algorithm achieved a poorer DOA estimation performance.

5. Conclusions

Multiple open spherical arrays were developed for the DOA estimation of multiple wideband sources based on the SH OPM algorithm. Compared with single spherical arrays, the multiple open spherical array structure is more exact and more suitable for wideband acoustic source DOA estimation. The simulation results demonstrated both that fewer computations were required for the SH-SFT-OPM and SH-CSM-OPM algorithms than for the SH-MUSIC algorithm and that these two algorithms could also obtain accurate DOA estimation results for multiple wideband sources by using multiple open spherical arrays, even in cases of lower received signal SNRs and fewer snapshots. Possible future work will extend the OPM algorithm to EB-ESPRIT to make real-time implementation possible.

Acknowledgements

This work was supported by the NSFC – China [grant number 61401027].

References

[1] Rafaely B. Analysis and design of spherical microphone arrays. IEEE Trans Speech Audio Process 2005;13(1):135–43.
[2] Jin CT, Epain N, Parthy A. Design, optimization and evaluation of a dual-radius spherical microphone array. IEEE Trans Audio Speech Lang Process 2014;22(1):193–204.
[3] Abhayapala TD, Ward DB. Theory and design of high order sound field microphones using spherical microphone array. In: 2002 IEEE international conference on acoustics, speech, and signal processing (ICASSP); 2002. p. II-1949–52.
[4] Prasanga Samarasinghe, Abhayapala T, Poletti M. Wavefield analysis over large areas using distributed higher order microphones. IEEE/ACM Trans Audio Speech Lang Process 2014;22(3):647–58.
[5] Balmages I, Rafaely B. Open-sphere designs for spherical microphone arrays. IEEE Trans Audio Speech Lang Process 2007;15(2):727–32.
[6] Meyer J, Elko G. A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield. In: 2002 IEEE international conference on acoustics, speech, and signal processing (ICASSP); 2002. p. II-1781–4.
[7] Li ZY, Duraiswami R. Flexible and optimal design of spherical microphone arrays for beamforming. IEEE Trans Audio Speech Lang Process 2007;15(2):702–14.
[8] Sun H, Teutsch H, Mabande E, Kellermann W. Robust localization of multiple sources in reverberant environments using EB-ESPRIT with spherical microphone arrays. In: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP); 2011. p. 117–20.
[9] Nadiri O, Rafaely B. Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test. IEEE Trans Audio Speech Lang Process 2014;22(10):1494–505.
[10] Khaykin D, Rafaely B. Coherent signals direction-of-arrival estimation using a spherical microphone array: frequency smoothing approach. In: 2009 IEEE workshop on applications of signal processing to audio and acoustics; 2009. p. 221–4.
[11] Huang Q, Zhang G, Xiang L, Fang Y. Unitary transformations for spherical harmonics MUSIC. Signal Process 2017;131:441–6.
[12] Abhayapala TD, Bhatta H. Coherent broadband source localization by modal space processing. 10th international conference on telecommunications, 2003. ICT 2003, vol. 2. IEEE; 2003. p. 1617–23.
[13] Kumar L, Hegde RM. Near-field acoustic source localization and beamforming in spherical harmonics domain. IEEE Trans Signal Process 2016;64(13):3351–61.
[14] Marcos S, Marsal A, Benidir M. The propagator method for source bearing estimation. Signal Process 1995;42(2):121–38.
[15] Claudio EDD, Parisi R. WAVES: weighted average of signal subspaces for robust wideband direction finding. IEEE Trans Signal Process 2002;49(10):2179–91.
[16] Wang H, Kaveh M. Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources. IEEE Trans Acoust Speech Signal Process 2003;33(4):823–31.
[17] Williams EG. Fourier acoustics: sound radiation and nearfield acoustical holography. New York: Academic; 1999.
[18] em32 Eigenmike® microphone array. < https://www.mhacoustics.com/home > .
[19] Rafaely B, Weiss B, Bachmat E. Spatial aliasing in spherical microphone arrays. IEEE Trans Signal Process 2007;55(3):1003–10.