

RESEARCH

Open Access

Acoustic source localization in mixed field using spherical microphone arrays

Qinghua Huang^{1*} and Tong Wang²

Abstract

Spherical microphone arrays have been used for source localization in three-dimensional space recently. In this paper, a two-stage algorithm is developed to localize mixed far-field and near-field acoustic sources in free-field environment. In the first stage, an array signal model is constructed in the spherical harmonics domain. The recurrent relation of spherical harmonics is independent of far-field and near-field mode strengths. Therefore, it is used to develop spherical estimating signal parameter via rotational invariance technique (ESPRIT)-like approach to estimate directions of arrival (DOAs) for both far-field and near-field sources. In the second stage, based on the estimated DOAs, simple one-dimensional MUSIC spectrum is exploited to distinguish far-field and near-field sources and estimate the ranges of near-field sources. The proposed algorithm can avoid multidimensional search and parameter pairing. Simulation results demonstrate the good performance for localizing far-field sources, or near-field ones, or mixed field sources.

Keywords: Far-field; Near-field; Source localization; Spherical array; Spherical harmonics

1 Introduction

Acoustic source localization using microphone arrays has many applications, such as video conferences, intelligent systems, and robotics. It has received great attention since almost four decades [1,2]. In most of array signal processing applications, the wavefront is assumed to be planar, that is, all the sources are located in the far-field (FF) of an array. In this case, the parameter that characterizes a source location is its direction of arrival (DOA) [2]. In the near-field (NF) of an array, the range information should be integrated into the array signal model for accurately characterizing sources [3]. Although plane wave assumption can simplify the modeling and processing, it cannot hold in near-field applications and results in analysis errors. Moreover, in some practical applications, the signals collected by microphone arrays are often the mixture of far-field and near-field sources. Each source may be located in the near-field or far-field of an array [4-9]. The localization methods for the mixed field sources should discriminate far-field and near-field sources. Then for far-

field sources, they only estimate DOA information. For near-field sources, range information is also estimated.

If an acoustic source locates in three-dimensional (3D) space, its position information is jointly described by range and bearing (azimuth and elevation). The geometric structure of a microphone array is very important for the localization performance. Currently, most localization techniques used a uniform linear array (ULA) or a uniform circular array (UCA) [1,2,5-7,9-11] to estimate source positions. Planar arrays, such as cross array and uniform rectangular array (URA) are the straightforward extensions of the ULA and can estimate both azimuth and elevation [8,12,13]. ULAs cause a 180° ambiguity in the azimuth estimation. UCAs can provide 360° azimuthal coverage due to its circular symmetry in the azimuth plane. The main drawback of planar arrays including UCAs is that they provide a smaller aperture in the elevation plane compared to the azimuth plane, resulting in poor estimation of elevation angles [10]. Some arbitrary array configurations were investigated to localize sources [14-17]. They were different from array uniformity that traditional localization approaches require. The array structure was selected according to some specific practical applications. Spherical microphone arrays have 3D symmetrical geometry configuration and can capture higher order sound field information. The 3D structure

* Correspondence: qinghua@shu.edu.cn

¹School of Communication and Information Engineering, Shanghai University, Shanghai 200072, People's Republic of China
Full list of author information is available at the end of the article

advantage facilitates more accurate sound source localization. Moreover, they can be analyzed within the mathematical framework of the spherical Fourier transform (SFT) which greatly simplifies processing in the space domain. Therefore, they have received considerable attention and have a wide variety of applications in the fields of source localization, beamforming, and acoustic analysis [18-20]. In this paper, we aim to develop a novel algorithm able to accurately estimate the locations of mixed field sources using spherical arrays.

Many techniques were proposed to estimate DOAs of multiple acoustic sources. Multiple signal classification (MUSIC) and estimating signal parameter via rotational invariance techniques (ESPRIT) are two subspace techniques [21,22]. The latter avoids multidimensional search in the parameter space. Goossens and Rogier proposed a unitary spherical ESPRIT algorithm based on the spherical phase-mode excitation that yielded accurate estimates with low computational complexity [23]. The eigenbeam (EB)-ESPRIT algorithm for spherical arrays was presented in [24] with its performance analysis for robust localization in reverberant environments. It only exploited the relation between a fixed order of spherical harmonics. Many approaches were based on beamforming [22]. Argentieri and Danes proposed an online beamspace MUSIC method with a beamforming scheme to localize sound sources in robotics [25]. Sun et al. proposed several steered beamformer-based and subspace-based localization techniques in the spherical EB domain [26]. They localized early reflections in room acoustic environments. Wu et al. used sparse recovery to localize sources and formulate super-resolution beamforming in the spherical harmonic domain [27]. However, when a source is close to the spherical array, the array signal model based on the far-field assumption is no longer valid. Independent component analysis (ICA) was used to estimate source locations. It employed higher order statistics and directly identified basis vectors containing the source location information. It was applied in near-field or far-field [28]. The ICA-based method was used to estimate DOAs for spherical microphone arrays [29,30]. It fails to localize sources which are not statistically independent. Source localization can be considered as an overcomplete basis representation problem using a grid of spatial locations. Many sparse recovery methods were used to estimate source DOAs [31,32]. If the source is in 3D space, the number of basis is large and the computational complexity is high. Some approaches assumed that one source was dominant over the others in some time-frequency zones [11,33]. They extended the single-source DOA algorithm over these zones to estimate multiple source locations. They were based on the sparse representation of the observation signals in the time-frequency domain.

In many practical applications, the observations collected by an array may be either mixed far-field and

near-field signals or multiple far-field signals or multiple near-field sources. Most of the above techniques localize sources in far-field or near-field. In recent years, source localization in mixed near-field and far-field has been developed using MUSIC algorithm [5-7], ESPRIT-like technique [8], or sparse signal reconstruction method [4] based on a linear array. Jiang et al. proposed a 3D source localization algorithm with a cross array [9]. First, the elevation angles are obtained based on the generalized ESPRIT method. Similar to the root MUSIC method, the range parameters are estimated with the elevation estimates. Finally, a MUSIC pseudo-spectrum function is used to get the azimuth angles with the elevation and range estimates. Due to the 3D symmetrical structure, spherical arrays have been widely used in far-field source localization [23,24]. A spherical microphone array was used in the near-field, and a new close-talking microphone array was proposed in [34]. It can adaptively compensate for the distance and orientation of a near-field source. Fisher and Rafaely presented a near-field spherical microphone array and defined the near-field criterion in terms of the array order and radius [35]. They analyzed spherical microphone array capabilities in the near-field and designed a radial filter discriminating the distances between the sources incident from the same direction [36]. Although the aforementioned work considers the near-field processing of the spherical microphone array, near-field or mixed field source localization via a spherical microphone array has not yet been studied. Based on the recurrent relation of the spherical harmonics, only DOAs were estimated for mixed field sources simultaneously [37]. However, how to distinguish near-field and far-field sources and how to estimate the ranges of near-field sources were not considered.

The aim of our work is to develop a new method that is able to localize mixed far-field and near-field sources simultaneously using spherical arrays in free-field environment. It can avoid the parameter pairing problem and complex multidimensional search. Three-dimensional MUSIC method scans the azimuth, elevation, and range parameter space and brings very high computational complexity. Therefore, it is not practical for direct source localization in 3D space. First, we construct the mixed near-field and far-field array signal model in the spherical harmonics domain. The mixed steering matrix in the spherical harmonics domain only contains the source DOA and range information. Moreover, the DOA and range information is decoupled. Exploiting the recurrent relation between spherical harmonics, we extend the spherical ESPRIT method to simultaneously discriminate directions of multiple far-field and near-field sources. This avoids two-dimensional parameter space search and the azimuth and elevation pairing. Based on the estimated DOAs, the ranges of near-field sources

can be easily obtained using the one-dimensional MUSIC algorithm with high resolution.

The remainder of this paper is organized as follows: Section 2 introduces mixed field array signal model in the spherical harmonics domain. A two-stage source localization with spherical arrays is developed in Section 3. Simulation results in Section 4 are presented to demonstrate the performance of the proposed algorithm. Conclusions are given in Section 5.

2 Array signal model for mixed field sources

To clarify the notations, scalars are denoted as italic letters (a, b, A, B, \dots), column vectors as lowercase boldface letters ($\mathbf{a}, \mathbf{b}, \dots$), and matrices as boldface capitals ($\mathbf{A}, \mathbf{B}, \dots$). The superscripts $T, *$, and H denote transpose, complex conjugation, and conjugate transpose, respectively. $\text{diag}(\cdot)$ defines a diagonal matrix and $\arg(\cdot)$ calculates the phase.

The spherical coordinate system is used to describe the positions of sensors and source signals in 3D space shown in Figure 1. A total of L identical and isotropic sensors mount on the rigid spherical surface with radius R . Each sensor element is unambiguously defined by its elevation θ_l and azimuth ϕ_l ($l = 1, 2, \dots, L$), measured from the positive z -axis and x -axis, respectively. Thus, $\mathbf{R}_l = (R, \theta_l, \phi_l)$ describes the sensor position. Consider a point source located at $\mathbf{r}_d = (r_d, \vartheta_d, \varphi_d)$, where r_d is the distance measured from the center of the spherical array, and ϑ_d and φ_d represent the elevation and azimuth of the source, respectively. For the spherical microphone array with an order N , the near-field extent is suggested [35,36]

$$r_N \approx \frac{N}{k}, \quad (1)$$

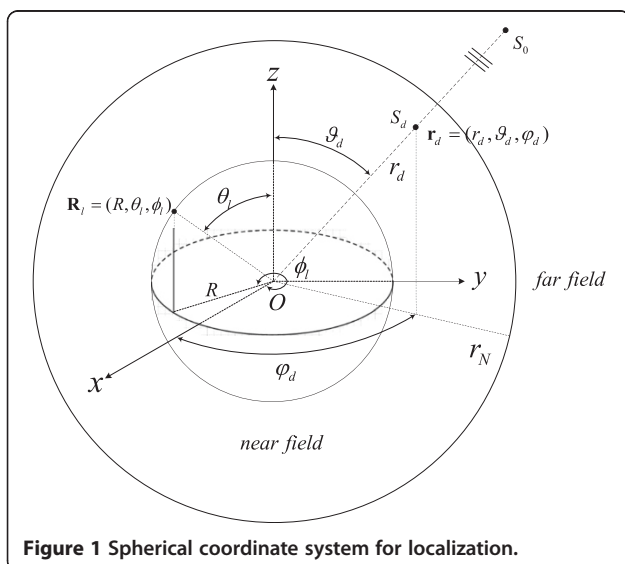


Figure 1 Spherical coordinate system for localization.

where k is the wavenumber. The maximal wavenumber is

$$k_{\max} = \frac{N}{R}. \quad (2)$$

By combining (1) and (2), the criterion for r_d to be in the near-field can be written as

$$r_d < r_N \approx \frac{k_{\max}}{k} R. \quad (3)$$

Assuming there are D source signals impinging on the spherical array. The first D_1 sources are assumed to be far-field signals, while the remaining $D_2 = D - D_1$ sources locate within the near-field extent of the array. In the presence of additive noises, the model in the space-frequency domain is represented as

$$\mathbf{x}(k) = \mathbf{A}_F \mathbf{s}_F(k) + \mathbf{A}_N \mathbf{s}_N(k) + \mathbf{v}(k), \quad (4)$$

where $\mathbf{x}(k) = [x_1(k) \ x_2(k) \ \dots \ x_L(k)]^T$ denotes an observation vector composed of the pressure samples at each sensor at a frequency corresponding to the wavenumber k . $\mathbf{v}(k) = [v_1(k) \ v_2(k) \ \dots \ v_L(k)]^T$ is a noise vector, and $\mathbf{s}_F(k) = [s_1(k) \ s_2(k) \ \dots \ s_{D_1}(k)]^T$ and $\mathbf{s}_N(k) = [s_{D_1+1}(k) \ s_{D_1+2}(k) \ \dots \ s_D(k)]^T$ are far-field and near-field source signals, respectively. They are assumed to be statistically independent and well separated. $\mathbf{A}_F \in \mathbb{C}^{L \times D_1}$ and $\mathbf{A}_N \in \mathbb{C}^{L \times D_2}$ are the corresponding physical steering matrices assuming the array is in free-field. The D sources are incident from DOAs $\Phi_d = (\vartheta_d, \varphi_d)$, where $d = 1, 2, \dots, D$. Range information r_{d_2} is important only for near-field sources, where $d_2 = D_1 + 1, D_1 + 2, \dots, D$. The objective is to estimate the azimuths and elevations for far-field sources and joint azimuth-elevation-ranges for near-field sources.

The localization algorithm is developed under the following assumptions:

1. The number of all sources is known.
2. The incident source signals are statistically independent.
3. The noise is zero-mean, complex circular Gaussian, and spatially uniform white, and is statistically independent of all the signals [6].

2.1 Spherical harmonic representation for spherical array processing

One advantage of spherical arrays is that it can be analyzed within the mathematical framework of the spherical Fourier transform which greatly simplifies processing in the space domain. Any square integrable function on a

sphere $g(\theta, \phi)$ can be denoted by g_{nm} using the following SFT:

$$g_{nm} = \int_0^{2\pi} \int_0^\pi g(\theta, \phi) [Y_n^m(\theta, \phi)]^* \sin\theta d\theta d\phi, \quad (5)$$

where the integral covers the entire surface of the unit sphere S^2 , and $Y_n^m(\theta, \phi)$ is the spherical harmonic of order n and degree m defined as

$$Y_n^m(\theta, \phi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos\theta) e^{im\phi}, \quad (6)$$

where $P_n^m(\cos\theta)$ is the associated Legendre polynomial [38]. The corresponding inverse Fourier transform is

$$g(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n g_{nm} Y_n^m(\theta, \phi). \quad (7)$$

The spherical harmonics are orthonormal, i.e., [19],

$$\int_0^{2\pi} \int_0^\pi Y_n^m(\theta, \phi) [Y_{n'}^{m'}(\theta, \phi)]^* \sin\theta d\theta d\phi = \delta_{nn'} \delta_{mm'}, \quad (8)$$

where $\delta_{nn'} = 1$ for $n = n'$, and $\delta_{nn'} = 0$ otherwise. Equation 8 is the orthonormality of continuous spherical harmonics. However, spherical microphone arrays perform spatial sampling of continuous functions defined on a sphere. Spatial sampling, similar to time-domain sampling, requires limited spatial bandwidth (limited harmonic order) to avoid aliasing [39].

Consider the highest order of the spherical microphone array is up to N . $\Omega_l = (\theta_l, \phi_l)$ denotes the elevation and the azimuth of the l th sensor. $\mathbf{Y}(\Omega)$ is defined as an $L \times (N+1)^2$ spherical harmonic matrix as follows:

$$\mathbf{Y}(\Omega) = \begin{bmatrix} Y_0^0(\Omega_1) & Y_1^{-1}(\Omega_1) & Y_1^0(\Omega_1) & Y_1^1(\Omega_1) & \cdots & Y_N^N(\Omega_1) \\ Y_0^0(\Omega_2) & Y_1^{-1}(\Omega_2) & Y_1^0(\Omega_2) & Y_1^1(\Omega_2) & \cdots & Y_N^N(\Omega_2) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_0^0(\Omega_L) & Y_1^{-1}(\Omega_L) & Y_1^0(\Omega_L) & Y_1^1(\Omega_L) & \cdots & Y_N^N(\Omega_L) \end{bmatrix}. \quad (9)$$

According to the inverse transform truncated up to order N in (7), the observations of the spherical microphone array can be expressed in the following form:

$$\mathbf{x}(k) = \mathbf{Y}(\Omega) \mathbf{x}_{nm}(k), \quad (10)$$

where $\mathbf{x}_{nm}(k)$ is a $(N+1)^2 \times 1$ transform coefficient vector in the spherical harmonics domain. In the same way, the noise can be expressed as

$$\mathbf{v}(k) = \mathbf{Y}(\Omega) \mathbf{v}_{nm}(k), \quad (11)$$

where $\mathbf{v}_{nm}(k)$ is a $(N+1)^2 \times 1$ transform coefficient vector in the spherical harmonics domain.

According to the spherical harmonic representation of the sound field [40,41], when sources are in the far-field

of an array, the element $a_F(l, d_1, k)$ of the far-field steering matrix \mathbf{A}_F is independent of the source distance and can be expressed using spherical harmonics as

$$a_F(l, d_1, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n b_n(kR) [Y_n^m(\vartheta_{d_1}, \varphi_{d_1})]^* Y_n^m(\theta_l, \phi_l), \quad (12)$$

where $b_n(kR)$ is the normalized far-field mode strength and depends on the sphere boundary [8], $l = 1, 2, \dots, L$, and $d_1 = 1, 2, \dots, D_1$. The term $a_F(l, d_1, k)$ describes the transfer characteristics from the d_1 th far-field source to the l th sensor. For concise representation, define $\Phi_d = (\vartheta_d, \varphi_d)$ to denote the elevation and the azimuth of the d th source and define a $(N+1)^2 \times D_1$ spherical harmonic matrix including the DOA information of the far-field sources as

$$\mathbf{Y}(\Phi_F) = \begin{bmatrix} [Y_0^0(\Phi_1)]^* & [Y_0^0(\Phi_2)]^* & \cdots & [Y_0^0(\Phi_{D_1})]^* \\ [Y_1^{-1}(\Phi_1)]^* & [Y_1^{-1}(\Phi_2)]^* & \cdots & [Y_1^{-1}(\Phi_{D_1})]^* \\ [Y_1^0(\Phi_1)]^* & [Y_1^0(\Phi_2)]^* & \cdots & [Y_1^0(\Phi_{D_1})]^* \\ [Y_1^1(\Phi_1)]^* & [Y_1^1(\Phi_2)]^* & \cdots & [Y_1^1(\Phi_{D_1})]^* \\ \vdots & \vdots & \ddots & \vdots \\ [Y_N^N(\Phi_1)]^* & [Y_N^N(\Phi_2)]^* & \cdots & [Y_N^N(\Phi_{D_1})]^* \end{bmatrix}. \quad (13)$$

\mathbf{B}_F is defined as a $(N+1)^2 \times (N+1)^2$ diagonal matrix consisting of the far-field mode strength $b_n(kR)$, i.e.,

$$\mathbf{B}_F = \text{diag}(\underbrace{b_0(kR)}_{n=0}, \underbrace{b_1(kR), b_1(kR)}_{n=1}, \dots, \underbrace{b_N(kR), \dots, b_N(kR)}_{n=N}). \quad (14)$$

Therefore, by combining (9), (12), (13), and (14), the far-field steering matrix can be represented as

$$\mathbf{A}_F = \mathbf{Y}(\Omega) [\mathbf{B}_F \mathbf{Y}(\Phi_F)]. \quad (15)$$

If near-field sources impinge on the spherical array, the element $a_N(l, d_2, k)$ of the near-field steering matrix \mathbf{A}_N can be expressed using spherical harmonics as

$$a_N(l, d_2, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n b_n^{d_2}(kR, kr_{d_2}) [Y_n^m(\vartheta_{d_2}, \varphi_{d_2})]^* Y_n^m(\theta_l, \phi_l), \quad (16)$$

where $b_n^{d_2}(kR, kr_{d_2})$ is the normalized near-field mode strength and depends on the sphere boundary and the source distance r_{d_2} . Similarly, $a_N(l, d_2, k)$ represents the transfer characteristics from the d_2 th near-field source to the l th sensor. The relation of near-field and far-field mode strengths is

$$b_n^d(kR, kr_d) = i^{-(n-1)} k b_n(kR) h_n^{(2)}(kr_d), \quad (17)$$

where $h_n^{(2)}(kr_d)$ represents the spherical Hankel function of the second kind [35,36]. Similarly, $\mathbf{Y}(\Phi_N)$ is defined as a $(N+1)^2 \times D_2$ matrix made up of the spherical harmonics of the near-field sources, i.e.,

$$\mathbf{Y}(\Phi_N) = \begin{bmatrix} [Y_0^0(\Phi_{D_1+1})]^* & [Y_0^0(\Phi_{D_1+2})]^* & \cdots & [Y_0^0(\Phi_D)]^* \\ [Y_1^{-1}(\Phi_{D_1+1})]^* & [Y_1^{-1}(\Phi_{D_1+2})]^* & \cdots & [Y_1^{-1}(\Phi_D)]^* \\ [Y_1^0(\Phi_{D_1+1})]^* & [Y_1^0(\Phi_{D_1+2})]^* & \cdots & [Y_1^0(\Phi_D)]^* \\ [Y_1^1(\Phi_{D_1+1})]^* & [Y_1^1(\Phi_{D_1+2})]^* & \cdots & [Y_1^1(\Phi_D)]^* \\ \vdots & \vdots & \ddots & \vdots \\ [Y_N^N(\Phi_{D_1+1})]^* & [Y_N^N(\Phi_{D_1+2})]^* & \cdots & [Y_N^N(\Phi_D)]^* \end{bmatrix}. \quad (18)$$

\mathbf{B}_N is a $(N+1)^2 \times D_2$ near-field mode strength matrix:

$$\mathbf{B}_N = \begin{bmatrix} b_0^{D_1+1}(kR, kr_{D_1+1}) & b_0^{D_1+2}(kR, kr_{D_1+2}) & \cdots & b_0^D(kR, kr_D) \\ b_1^{D_1+1}(kR, kr_{D_1+1}) & b_1^{D_1+2}(kR, kr_{D_1+2}) & \cdots & b_1^D(kR, kr_D) \\ b_1^{D_1+1}(kR, kr_{D_1+1}) & b_1^{D_1+2}(kR, kr_{D_1+2}) & \cdots & b_1^D(kR, kr_D) \\ b_1^{D_1+1}(kR, kr_{D_1+1}) & b_1^{D_1+2}(kR, kr_{D_1+2}) & \cdots & b_1^D(kR, kr_D) \\ \vdots & \vdots & \ddots & \vdots \\ b_N^{D_1+1}(kR, kr_{D_1+1}) & b_N^{D_1+2}(kR, kr_{D_1+2}) & \cdots & b_N^D(kR, kr_D) \end{bmatrix}. \quad (19)$$

The near-field steering matrix is expressed by combining (9), (16), (18), and (19) as

$$\mathbf{A}_N = \mathbf{Y}(\Omega)[\mathbf{B}_N \odot \mathbf{Y}(\Phi_N)], \quad (20)$$

where \odot is the Hadamard product.

2.2 Array signal model in the spherical harmonics domain

By combining (10), (11), (15), and (20) into (4), the array signal model is written as

$$\begin{aligned} \mathbf{Y}(\Omega)\mathbf{x}_{nm}(k) &= \mathbf{Y}(\Omega)[\mathbf{B}_F\mathbf{Y}(\Phi_F)]\mathbf{s}_F(k) + \mathbf{Y}(\Omega) \\ &\quad \times [\mathbf{B}_N\odot\mathbf{Y}(\Phi_N)]\mathbf{s}_N(k) \\ &\quad + \mathbf{Y}(\Omega)\mathbf{v}_{nm}(k). \end{aligned} \quad (21)$$

According to the least squares criterion, the array signal model is constructed in the spherical harmonics domain as

$$\mathbf{x}_{nm}(k) = \mathbf{Y}_F(\Phi)\mathbf{s}_F(k) + \mathbf{Y}_N(\Phi)\mathbf{s}_N(k) + \mathbf{v}_{nm}(k), \quad (22)$$

where $\mathbf{Y}_F(\Phi) = \mathbf{B}_F\mathbf{Y}(\Phi_F)$ and $\mathbf{Y}_N(\Phi) = \mathbf{B}_N \odot \mathbf{Y}(\Phi_N)$ are the new steering matrix of the far-field and near-field in the spherical harmonics domain, respectively. Equation 22 can be expressed in a compact form as

$$\mathbf{x}_{nm}(k) = \mathbf{Y}_{FN}(\Phi)\mathbf{s}_{FN}(k) + \mathbf{v}_{nm}(k), \quad (23)$$

where $\mathbf{Y}_{FN}(\Phi) = [\mathbf{Y}_F(\Phi) \quad \mathbf{Y}_N(\Phi)] \in C^{(N+1)^2 \times D}$ is the new mixed steering matrix in the spherical harmonics domain and $\mathbf{s}_{FN}(k) = [\mathbf{s}_F^T(k) \quad \mathbf{s}_N^T(k)]^T$. In the spherical harmonics domain, the far-field and near-field mixed steering matrix is independent of the element positions of the sampled array. The following localization algorithm in Section 3 is developed based on the array signal model in (23).

3 Acoustic source localization algorithm

The source localization in mixed field aims to estimate 2D parameters $\{\vartheta_d, \varphi_d\}$ for far-field sources and 3D parameters $\{r_d, \vartheta_d, \varphi_d\}$ for near-field sources given the array observations $\mathbf{x}(k)$. Based on the spherical harmonic model in (23), the common characteristics of the far-field and near-field sources lie in that only spherical harmonics in the mixed steering matrix contain the DOA information. That is, the mixed steering matrix contains the DOAs of all sources. The difference between these sources is whether mode strength depends on the source distance or not. Therefore, DOAs can be estimated using the recursive relationship of spherical harmonics. Based on the estimated DOAs, near-field range can be easily computed by conventional MUSIC algorithm.

3.1 DOA estimation

In order to exploit the recurrent relation between spherical harmonics and avoid complex search in the 3D parameter space, we develop a spherical ESPRIT-like algorithm, automatically estimating paired azimuth and elevation angles for multiple mixed source signals. Define $\mu = \tan \vartheta e^{i\varphi}$ only containing the DOA information, where ϑ and φ are the elevation and azimuth angles of a source, respectively. According to the recursive relation for the associated Legendre polynomials of the adjacent three degrees ($m-1$, m , and $m+1$) [38] and the spherical harmonics definition in (6), we get the following relation that the DOA estimation depends on

$$2mY_n^m(\vartheta, \varphi) + \lambda_{nm}^+ Y_n^{m+1}(\vartheta, \varphi)\mu^* + \lambda_{nm}^- Y_n^{m-1}(\vartheta, \varphi)\mu = 0, \quad (24)$$

where $\lambda_{nm}^\pm = \sqrt{(n \mp m)(n \pm m + 1)}$ [42]. For a source whether in far-field or in near-field, the recurrent relationship in (24) is independent of the corresponding mode strength $b_n(kR)$ or $b_n^d(kR, kr_d)$. That is, the relation is still satisfied as follows:

$$\begin{aligned} 2mb_n(kR)Y_n^m(\vartheta, \varphi) + \lambda_{nm}^+ b_n(kR)Y_n^{m+1}(\vartheta, \varphi)\mu^* \\ + \lambda_{nm}^- b_n(kR)Y_n^{m-1}(\vartheta, \varphi)\mu = 0, \end{aligned} \quad (25)$$

or

$$\begin{aligned} 2mb_n^d(kR, kr_d)Y_n^m(\vartheta, \varphi) + \lambda_{nm}^+ b_n^d(kR, kr_d)Y_n^{m+1}(\vartheta, \varphi)\mu^* \\ + \lambda_{nm}^- b_n^d(kR, kr_d)Y_n^{m-1}(\vartheta, \varphi)\mu = 0. \end{aligned} \quad (26)$$

For a fixed order n , we choose all rows from $\mathbf{Y}_{FN}(\Phi)$ consisting of elements Y_n^m , $m = -n, -n+1, \dots, n-2$, to construct a $(2n-1) \times D$ matrix $\mathbf{B}_n^{(-1)}$, select $m = -n+1, -n+2, \dots, n-1$ to construct the second $(2n-1) \times D$ matrix $\mathbf{B}_n^{(0)}$, and choose $m = -n+2, -n+3, \dots, n$ to construct the third

$(2n - 1) \times D$ matrix $\mathbf{B}_n^{(1)}$. When the order n varies from 1 to N , $\mathbf{Y}_{\text{FN}}^{(-1)}(\Phi) = \left[\left(\mathbf{B}_1^{(-1)} \right)^T \quad \left(\mathbf{B}_2^{(-1)} \right)^T \quad \cdots \quad \left(\mathbf{B}_N^{(-1)} \right)^T \right]^T \in \mathbb{C}^{N^2 \times D}$, the second chosen $N^2 \times D$ sub-matrix is $\mathbf{Y}_{\text{FN}}^{(0)}(\Phi) = \left[\left(\mathbf{B}_1^{(0)} \right)^T \quad \left(\mathbf{B}_2^{(0)} \right)^T \quad \cdots \quad \left(\mathbf{B}_N^{(0)} \right)^T \right]^T$, and the third sub-matrix is $\mathbf{Y}_{\text{FN}}^{(1)}(\Phi) = \left[\left(\mathbf{B}_1^{(1)} \right)^T \quad \left(\mathbf{B}_2^{(1)} \right)^T \quad \cdots \quad \left(\mathbf{B}_N^{(1)} \right)^T \right]^T$. To exploit the recursive relationship of the three sub-matrices $\mathbf{Y}_{\text{FN}}^{(-1)}(\Phi)$, $\mathbf{Y}_{\text{FN}}^{(0)}(\Phi)$, and $\mathbf{Y}_{\text{FN}}^{(1)}(\Phi)$ including all spherical harmonics up to order N , we define four diagonal matrices as follows:

$$\Theta = \text{diag}(\mu_1, \mu_2, \dots, \mu_D), \quad (27)$$

$$\Gamma = \text{diag} \left(\underbrace{0}_{n=1}, \underbrace{-1, 0, 1}_{n=2}, \dots, \underbrace{-N+1, \dots, N-1}_{n=N} \right), \quad (28)$$

$$\Lambda^\pm = \text{diag} \left(\underbrace{\lambda_{1,0}^\pm}_{n=1}, \underbrace{\lambda_{2,-1}^\pm, \lambda_{2,0}^\pm, \lambda_{2,1}^\pm}_{n=2}, \dots, \underbrace{\lambda_{N,-N+1}^\pm, \dots, \lambda_{N,N-1}^\pm}_{n=N} \right), \quad (29)$$

where Θ contains the DOA information of all incident sources, Γ and Λ^\pm are the three $N^2 \times N^2$ diagonal matrices. The recurrent relationship of the three sub-matrices $\mathbf{Y}_{\text{FN}}^{(q)}(\Phi)$ ($q = -1, 0, 1$) is described as

$$2\Gamma \mathbf{Y}_{\text{FN}}^{(0)}(\Phi) + \Lambda^+ \mathbf{Y}_{\text{FN}}^{(1)}(\Phi) \Theta^* + \Lambda^- \mathbf{Y}_{\text{FN}}^{(-1)}(\Phi) \Theta = \mathbf{0}. \quad (30)$$

We cannot solve (30) directly to estimate the DOAs because $\mathbf{Y}_{\text{FN}}^{(q)}(\Phi)$ is unknown. The available data are the sensor observations. The covariance matrix $\mathbf{R}_{nm}(k)$ of the transform coefficient vector $\mathbf{x}_{nm}(k)$ can be constructed as

$$\begin{aligned} \mathbf{R}_{nm}(k) &= E[\mathbf{x}_{nm}(k) \mathbf{x}_{nm}^H(k)] \\ &= \mathbf{Y}_{\text{FN}}(\Phi) \mathbf{R}_s(k) \mathbf{Y}_{\text{FN}}^H(\Phi) \\ &\quad + E[\mathbf{v}_{nm}(k) \mathbf{v}_{nm}^H(k)], \end{aligned} \quad (31)$$

where $\mathbf{R}_s(k) = E[\mathbf{s}(k) \mathbf{s}^H(k)]$. It can be estimated by the sample covariance matrix from the sensor samples. The eigenvalue decomposition (EVD) of \mathbf{R}_{nm} results in two orthogonal subspaces:

$$\mathbf{R}_{nm} = \mathbf{U} \Sigma \mathbf{U}^H = \mathbf{U}_s \Sigma_s \mathbf{U}_s^H + \mathbf{U}_v \Sigma_v \mathbf{U}_v^H, \quad (32)$$

where $\mathbf{U}_s \in \mathbb{C}^{(N+1)^2 \times D}$ contains D eigenvectors spanning the signal subspace of \mathbf{R}_{nm} , and the diagonal matrix Σ_s contains the corresponding eigenvalues. Similarly, \mathbf{U}_v denotes the noise subspace, and Σ_v is built from the remaining $(N+1)^2 - D$ eigenvalues of \mathbf{R}_{nm} .

According to (31) and (32), \mathbf{U}_s spans the same range as that of the mixed steering matrix $\mathbf{Y}_{\text{FN}}(\Phi)$. Therefore, the signal subspace \mathbf{U}_s can be transformed into the mixed steering matrix $\mathbf{Y}_{\text{FN}}(\Phi)$, that is, $\mathbf{U}_s = \mathbf{Y}_{\text{FN}}(\Phi) \mathbf{T}$, where \mathbf{T} is a unique non-singular $D \times D$ matrix, called similarity transform matrix. Three sub-matrices $\mathbf{U}_s^{(q)}$ choosing from the signal subspace \mathbf{U}_s satisfy the same recurrent relationship as

$$2\Gamma \mathbf{U}_s^{(0)} + \Lambda^+ \mathbf{U}_s^{(1)} \Psi^* + \Lambda^- \mathbf{U}_s^{(-1)} \Psi = \mathbf{0}, \quad (33)$$

where $\Psi = \mathbf{T}^{-1} \Theta \mathbf{T}$. We can rewrite this equation in block matrix form:

$$2\Gamma \mathbf{U}_s^{(0)} + \mathbf{E} \Psi = \mathbf{0}, \quad (34)$$

where $\mathbf{E} = [\Lambda^- \mathbf{U}_s^{(-1)}; \Lambda^+ \mathbf{U}_s^{(1)}] \in \mathbb{C}^{N^2 \times 2D}$, $\underline{\Psi} = [\Psi^T; \Psi^H]^T \in \mathbb{C}^{2D \times D}$ has the block conjugate structure. Equation 34 has the following solution:

$$\underline{\Psi} = -2(\mathbf{E}^H \mathbf{E})^{-1} \mathbf{E}^H \Gamma \mathbf{U}_s^{(0)}. \quad (35)$$

The elevation angle $\hat{\theta}_d$ and the azimuth angle $\hat{\phi}_d$ are easily estimated from the eigenvalues $\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_D$ of either the top or the bottom $D \times D$ sub-block of $\underline{\Psi}$ as follows:

$$\hat{\theta}_d = \arctan |\hat{\mu}_d|, \quad \hat{\phi}_d = \arg(\hat{\mu}_d), \quad (36)$$

where $d = 1, 2, \dots, D$. When $N^2 < 2D$, this procedure fails. Hence, the maximum number of sources that can be accurately estimated by this algorithm is $D = \lfloor N^2/2 \rfloor$, where the operation $\lfloor \cdot \rfloor$ is the flooring operation. The three classical spatial sampling schemes for a spherical microphone array are equiangular, Gaussian, and nearly uniform sampling schemes. For a spherical microphone array with a given order N , the equiangular sampling scheme requires $4(N+1)^2$ sensors, the Gaussian sampling scheme demands $2(N+1)^2$ sensors, while the uniform sampling scheme only needs $(N+1)^2$ sensors [43]. Therefore, when L sensors collect the information of acoustic sources, the dimension of the spherical harmonics space $(N+1)^2 \leq L$, that is, the dimension of localization in the spherical harmonics domain is lower than that of the element space. The maximal number of sources that can be uniquely estimated is the nearest integer less than or equal to $N^2/2$.

3.2 Range estimation

The above DOA estimator can provide azimuth and elevation estimates of both far-field and near-field sources. However, it cannot discriminate far-field or near-field sources [37]. Only for near-field sources, range information must be estimated. When the source is in the near-

field, the steering vector is dependent on the range in the spherical harmonics domain:

$$\mathbf{y}_N(r_d, \Phi_d) = [b_0^d(kR, kr_d)[Y_0^0(\Phi_d)]^* b_1^d(kR, kr_d) [Y_1^{-1}(\Phi_d)]^* \cdots b_N^d(kR, kr_d)[Y_N^N(\Phi_d)]^*]^T. \quad (37)$$

Therefore, the near-field MUSIC spectrum [44] is

$$p(r_d, \hat{\Phi}_d) = \frac{1}{\mathbf{y}_N^H(r_d, \hat{\Phi}_d) (\mathbf{I} - \mathbf{U}_s \mathbf{U}_s^H) \mathbf{y}_N(r_d, \hat{\Phi}_d)}. \quad (38)$$

The near-field search extent is $r_d \in (R, r_N)$. The range estimate of the d th source is obtained by

$$\hat{r}_d = \max_{r_d} p(r_d, \hat{\Phi}_d). \quad (39)$$

The far-field steering vector in the spherical harmonics domain is independent of the source range:

$$\mathbf{y}_F(\Phi_d) = [b_0(kR)[Y_0^0(\Phi_d)]^* b_1(kR)[Y_1^{-1}(\Phi_d)]^* \cdots b_N(kR)[Y_N^N(\Phi_d)]^*]^T. \quad (40)$$

The far-field MUSIC spectrum [44] is

$$p(\hat{\Phi}_d) = \frac{1}{\mathbf{y}_F^H(\hat{\Phi}_d) (\mathbf{I} - \mathbf{U}_s \mathbf{U}_s^H) \mathbf{y}_F(\hat{\Phi}_d)}. \quad (41)$$

With the DOA estimates $\hat{\Phi}_d = (\hat{\vartheta}_d, \hat{\varphi}_d)$ ($d = 1, 2, \dots, D$) in (36), the range estimator calculates the MUSIC spectra according to (38) and (41). For a DOA estimate $\hat{\Phi}_d$, we compare $p(\hat{\Phi}_d)$ and the peak of $p(r_d, \hat{\Phi}_d)$; if the former is larger than the latter, the source is the far-field one. Otherwise, the source is the near-field one and the estimated \hat{r}_d in (39) automatically pairs with the DOA estimate $\hat{\Phi}_d$.

3.3 Algorithm summarization

The proposed two-stage algorithm can be summarized as follows:

- Step 1. *Array signal modeling*: Apply spherical harmonic representation and construct mixed field array signal model in the spherical harmonics domain in (23).
- Step 2. *DOA estimation*: Perform the EVD of \mathbf{R}_{nm} in (32) and choose three sub-matrices from the signal subspace. Construct the recurrent relation of these sub-matrices in (33) and estimate DOA information for all sources in (36).
- Step 3. *Range estimation*: Based on the estimated DOAs, compute the far-field MUSIC spectrum in

(41) and search the near-field MUSIC spectrum in (38) to discriminate far-field and near-field sources and obtain the pairing ranges for the near-field sources.

Remarks

1. For computational complexity, we mainly consider the implementation of EVD and one-dimensional (1D) MUSIC spectral search. In the spherical harmonics domain, the dimension of covariance matrix \mathbf{R}_{nm} is $(N + 1)^2 \times (N + 1)^2$. The computational complexity of the proposed localization algorithm includes (a) the eigendecomposition of \mathbf{R}_{nm} , of order $\mathcal{O}((N + 1)^6)$, and (b) 1D spectral search, of order $\mathcal{O}((N + 1)^4 g_r)$, where g_r is the search number conducted along the range axis [45].
2. Whether the incident sources are far-field ones, near-field ones, or their mixture, the spherical ESPRIT-like algorithm in the first stage can estimate all DOAs. When all incident sources are far-field ones, only the DOA information is enough. If all sources locate in near-field, 1D MUSIC spectral search can find out the pairing range parameters. When the mixed sources include far-field and near-field, based on the near-field spectral search, compute MUSIC spectra in (38) and (41) and distinguish the source being far-field or near-field one.
3. The proposed algorithm can localize mixed sources without parameter pairing and multidimensional search. In the first stage, the proposed algorithm estimates DOAs of mixed far-field and near-field sources. In the second stage, the DOA estimates are used to compute the 1D MUSIC spectra according to (38) and (41). The spectral peak corresponds to the pairing ranges. Therefore, parameter pairing can be avoided in the proposed method.

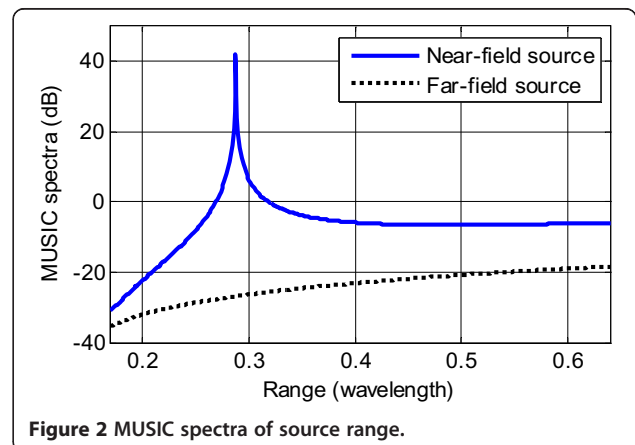
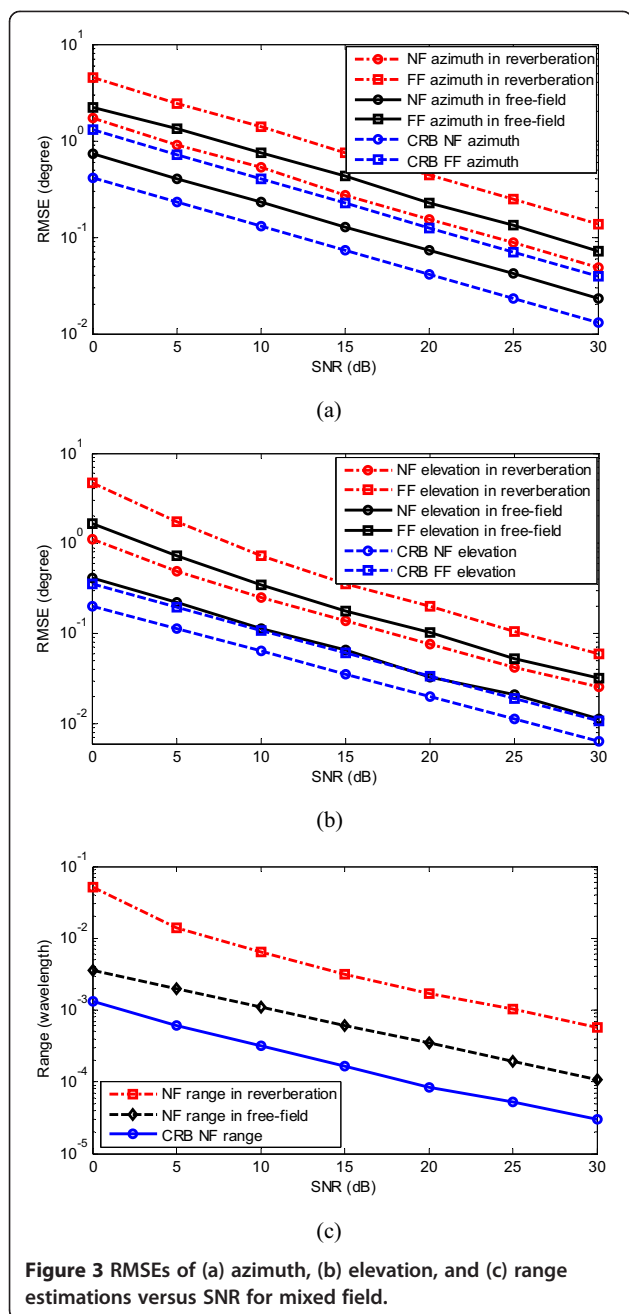


Figure 2 MUSIC spectra of source range.



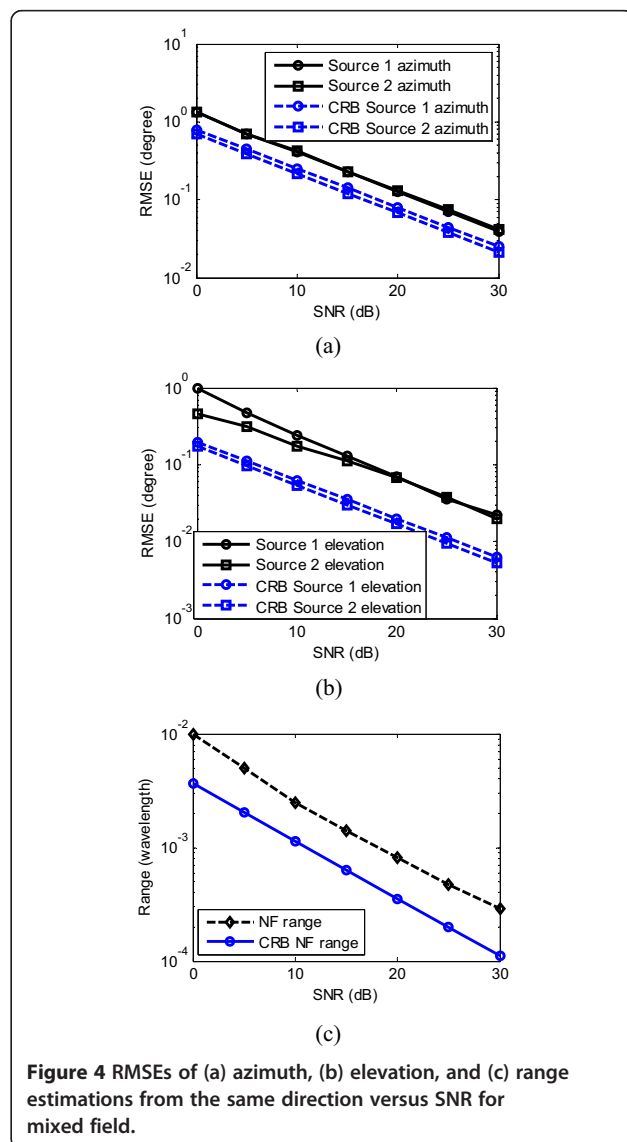
4 Simulations

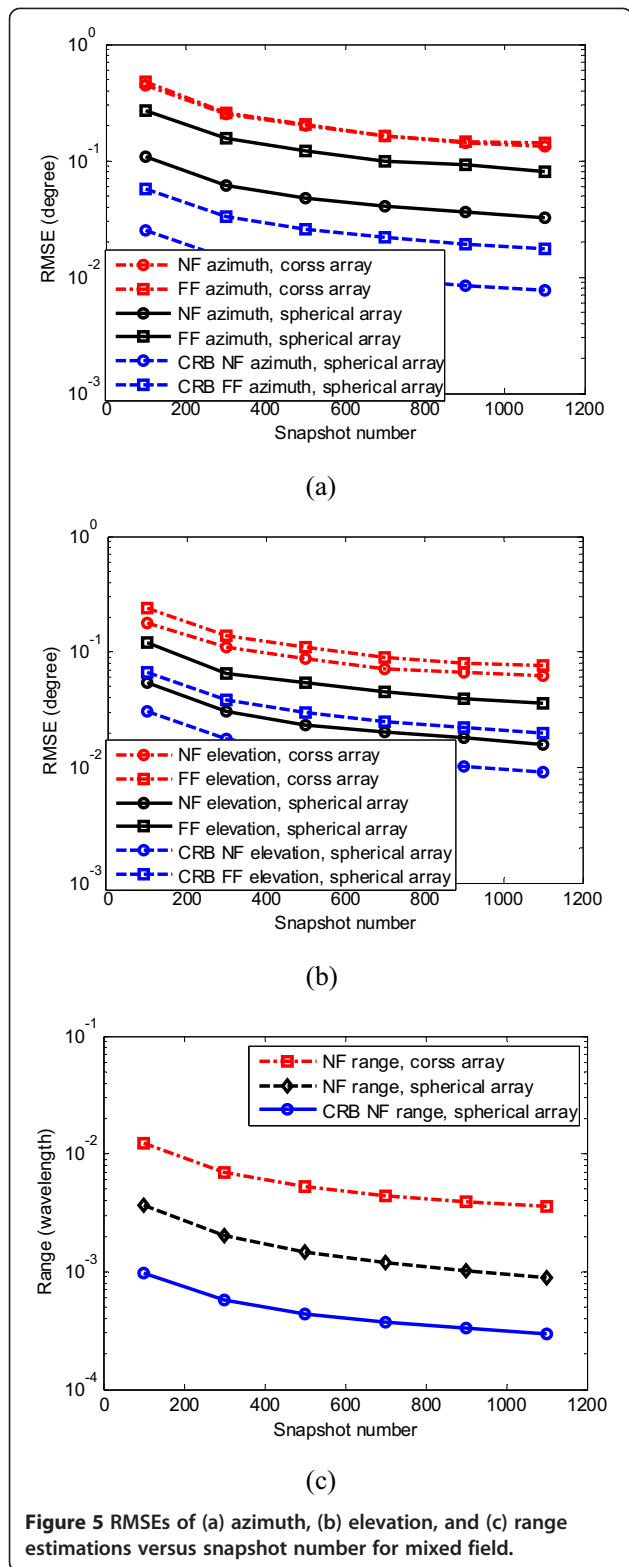
In this section, we conduct some simulations in free-field environment to evaluate the proposed localization algorithm for narrowband and wideband sources. A 32-element uniform sampling [43] of spherical microphone array is chosen to estimate source locations. Its radius is assumed to be 10 cm. The highest spherical harmonics order is $N=4$, so the near-field extent of the array is $(R, 0.64\lambda)$, where λ is the wavelength. The maximum number of sources that can be detected by the algorithm is $D=N^2/2=8$. The DOA (azimuth and elevation) and range estimations are scaled in units of degree and

wavelength (or meter), respectively. The performance of the localization estimation is measured by the root-mean-square error (RMSE) of 1,000 independent Monte Carlo trials. In addition, the Bayesian Cramer-Rao bound (CRB) provides a lower bound on the variance of any estimated parameter and defines the ultimate accuracy. The CRB analysis in [46] assumed all the sources were from far-field. The CRB analysis in [47] assumed the incident sources were all from near-field. When both far-field and near-field sources coexisted, the CRB analysis was provided in [6].

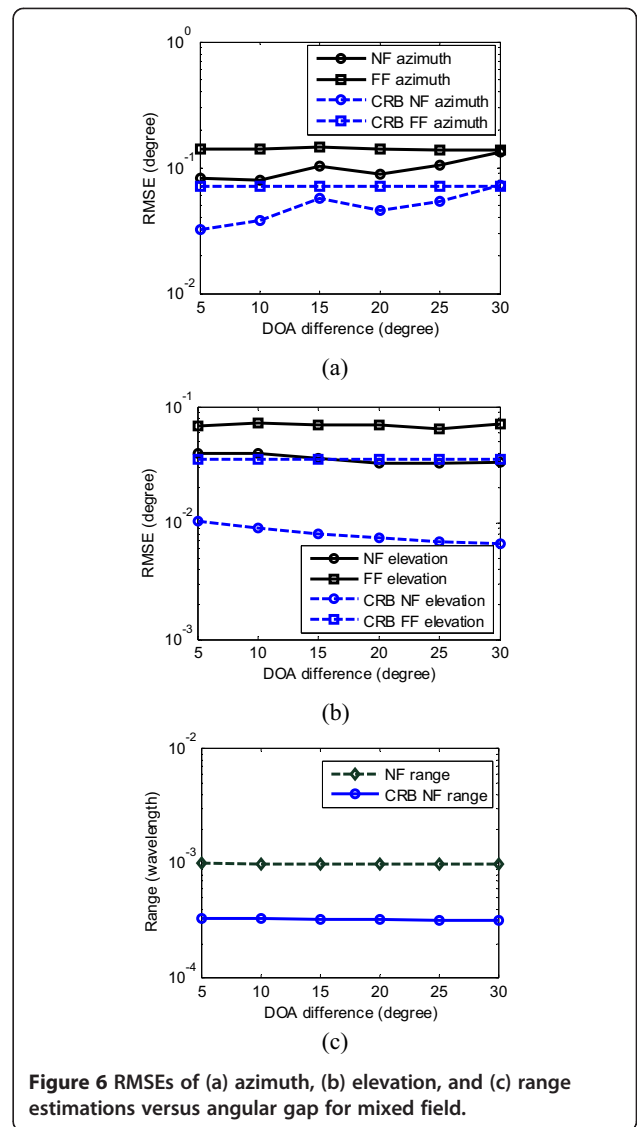
4.1 Narrowband source localization

The first simulation demonstrates the performance of the proposed algorithm in localizing far-field and near-

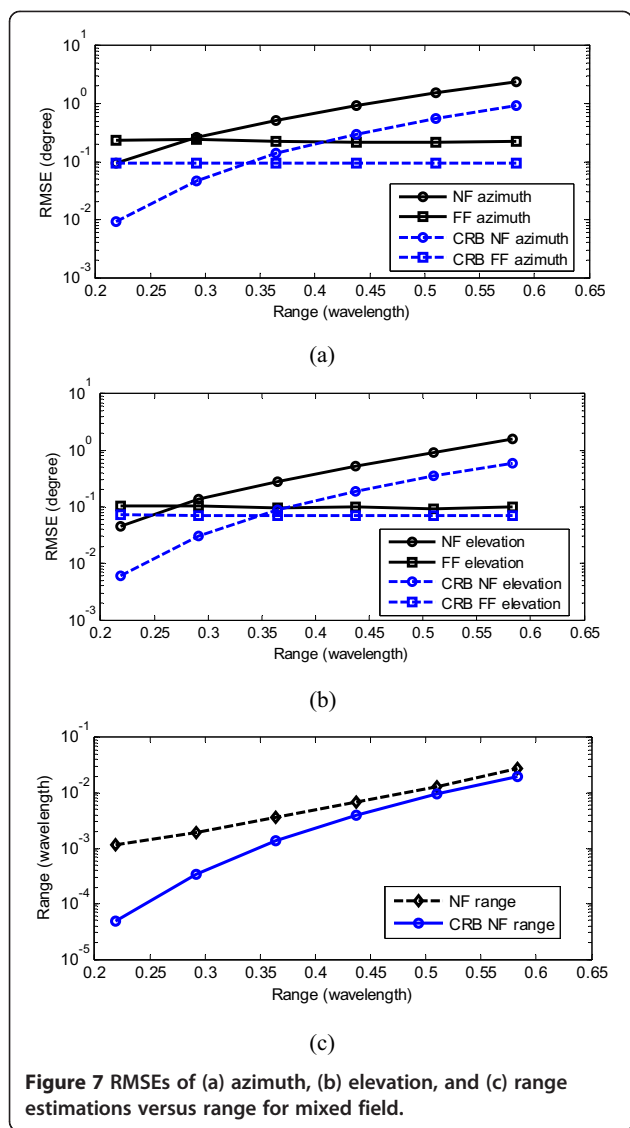




field sources. One far-field source and one near-field source are located at $(r_1, \Phi_1) = (\infty, 45^\circ, 68^\circ)$ and $(r_2, \Phi_2) = (0.29\lambda, 60^\circ, 122^\circ)$, respectively. The number of snapshots and SNR are fixed at 128 and 15 dB, respectively.



Firstly, the azimuths and elevations of the two sources are estimated. The RMSEs of the azimuth and elevation estimations are 0.07, 0.22, 0.03, and 0.10, respectively. Based on the DOA estimations and suppose the two sources both located in the near-field, MUSIC spectra for the two sources using (38) of the proposed algorithm are shown in Figure 2. Assume the two sources both located in the far-field, the MUSIC spectra of the two sources calculated by (41) are -12.39 and -4.73 dB, respectively. From the maximums of the MUSIC spectra, we can discriminate that one is the near-field source and the other is the far-field source. The range of the near-field source can be estimated from the MUSIC spectra. Therefore, the proposed algorithm can distinguish the near-field and far-field sources and performs well in localizing them.



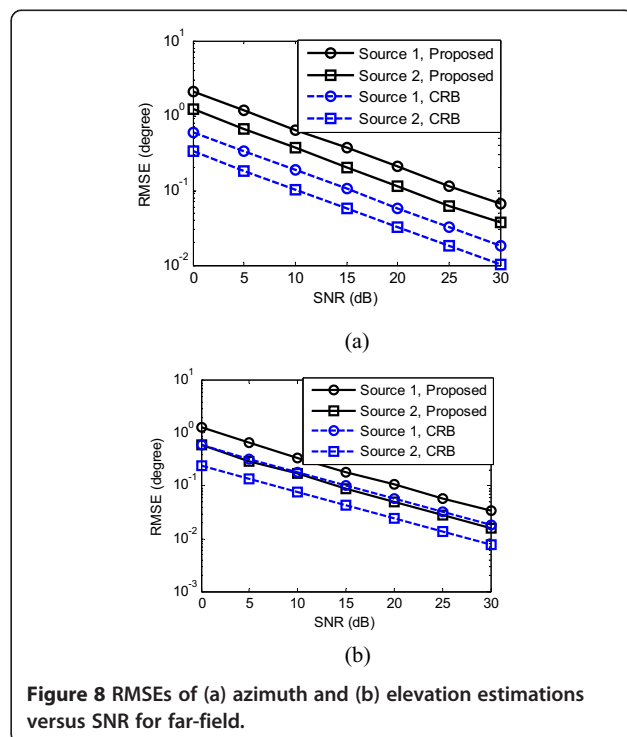
1. *RMSE versus SNR*: The number of snapshots is set equal to 128. The two sources are localized in free-field and a rectangular room with the floor area 71 m², the ceiling height 3 m, and the reverberation time 0.7 s [48], respectively. The array is placed in the center of the rectangular room. When SNR varies from 0 to 30 dB, the RMSEs of the azimuths, elevations, and range estimations are shown in Figure 3. When SNR increases, the RMSEs of the azimuths, elevations, and range estimations decrease. The localization performance degrades in reverberant environment compared with that in free-field. When the two sources are incident from the same direction (45°, 122°), the RMSEs of azimuths, elevations, and range estimations versus SNR in free-field are shown in Figure 4. From the figure, we can see that the proposed algorithm can localize the far-field and

Table 1 RMSEs of azimuth, elevation, and range estimations for mixed field

| Source | Azimuth (deg) | Elevation (deg) | Range (wavelength) |
|---------------------|---------------|-----------------|--------------------|
| Near-field source 1 | 0.25 | 0.01 | 0.002 |
| Near-field source 2 | 0.03 | 0.01 | 0.002 |
| Near-field source 3 | 0.42 | 0.01 | 0.010 |
| Near-field source 4 | 0.01 | 0.02 | 0.026 |
| Far-field source 5 | 0.13 | 0.01 | - |
| Far-field source 6 | 0.11 | 0.05 | - |
| Far-field source 7 | 1.47 | 0.03 | - |
| Far-field source 8 | 0.04 | 0.38 | - |

near-field source incident from the same direction. Moreover, the proposed algorithm has approximated estimation accuracy for the far-field and near-field sources with respect to azimuths and elevations. The elevation angle estimation accuracy is higher than the azimuth estimation accuracy. This is because spherical arrays can provide a larger aperture in the elevation plane than that in the azimuth plane.

2. *RMSE versus snapshot*: A crossed array placed in the X-Z plane is used in the simulation. Each ULA branch of the array consists of 15 ($M = 7$) uniformly spaced omni-directional sensors with the inter-sensor spacing R / M . First, the elevation angles are obtained based on the generalized ESPRIT method. Then, the range parameters are estimated



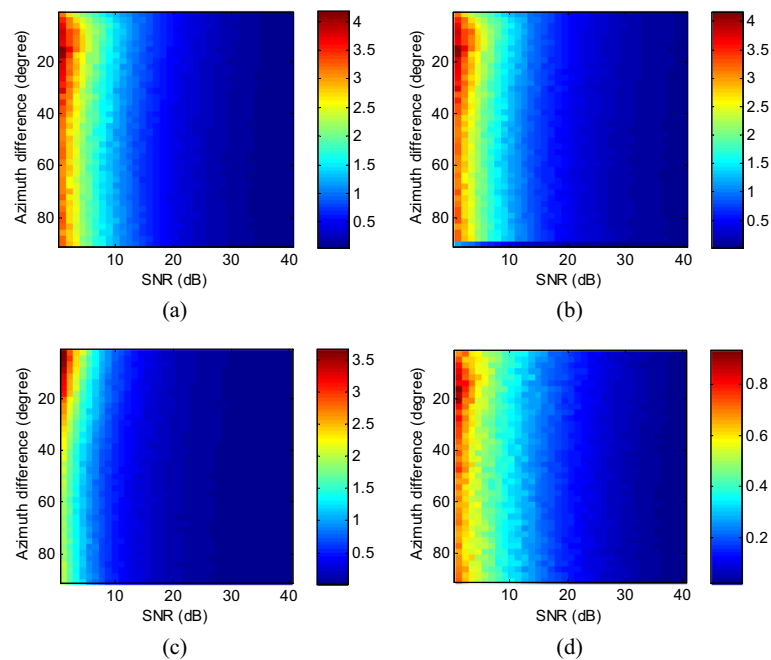


Figure 9 RMSEs of azimuth and elevation estimations versus azimuth variation and SNR for far-field. (a) Azimuth RMSE of source 1. (b) Azimuth RMSE of source 2. (c) Elevation RMSE of source 1. (d) Elevation RMSE of source 2.

using the root MUSIC method based on the elevation estimates [9]. Finally, a 1D MUSIC pseudo-spectrum is used to search the azimuth angles with the elevation and range estimates. SNR is fixed at 15 dB. When the number of snapshots varies from 100 to 1,100, the average performances

of 1,000 Monte Carlo runs are shown in Figure 5. The RMSEs of the azimuth, elevation, and range estimations decrease as the number of snapshots increases. The estimation performance of the spherical array is better than that of the cross array. This may be due to the symmetrical structure of

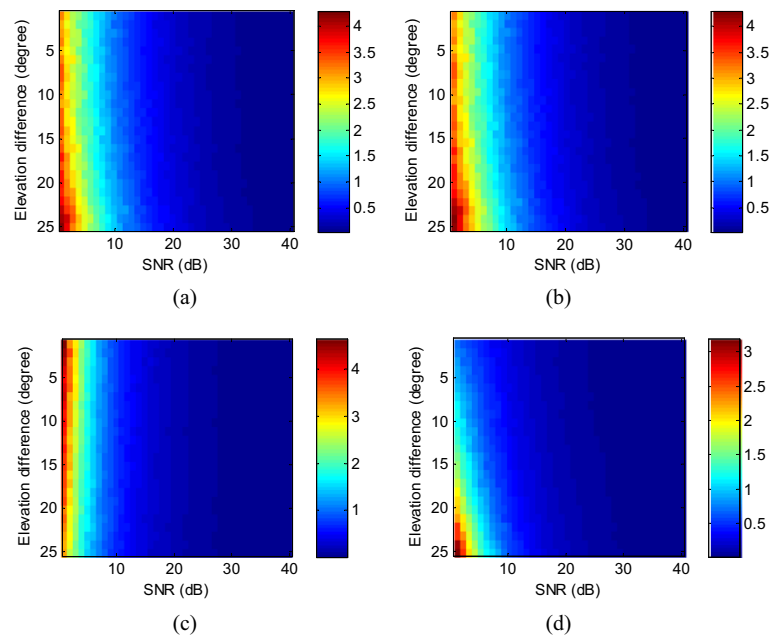


Figure 10 RMSEs of azimuth and elevation estimations versus elevation variation and SNR for far-field. (a) Azimuth RMSE of source 1. (b) Azimuth RMSE of source 2. (c) Elevation RMSE of source 1. (d) Elevation RMSE of source 2.

the spherical array in 3D space and the simultaneous estimation for azimuth and elevation angles of the proposed method. Both our proposed algorithm and the three-step estimation method in [9] are based on the eigendecomposition of the array covariance matrix. The estimate of the covariance matrix affects the localization performance. Vershynin proposed that the sample size $Q \sim (N + 1)^2$ suffices to estimate the covariance by the sample covariance matrix [49]. Therefore, the localization performance of the two methods gets more stable with the larger number of the snapshots. The CRBs decrease proportional to the number of the snapshots. When the number of snapshots increases, the CRBs get more stable too.

3. *RMSE versus angular gap*: The snapshot number is set equal to 128. The SNR is fixed at 15 dB. When the direction of the far-field source is fixed, the azimuth and elevation of the near-field source both vary from 5° to 30° . The RMSE of directions and range estimations are demonstrated in Figure 6. The DOA estimation of the far-field source and the range estimation of the near-field source are insensitive to the angular gap. The azimuth estimation performance for the near-field source of the proposed algorithm changes with the increase of the angular gap. The RMSE of the elevation estimation for the near-field source gets slightly smaller.
4. *RMSE versus range*: Let the number of snapshots and SNR be 128 and 10 dB, respectively. When the range of the near-field source varies from 0.22λ to 0.58λ , the RMSEs of DOAs and range estimations are shown in Figure 7. From the results, we can see that both of DOAs and range estimations of the near-field source are very sensitive to the varied range. The RMSEs of the azimuth, elevation, and range estimations for the near-field source, which is closer to the spherical array, are smaller than those of the source which is farther to the array. However, the location estimation of the far-field source is insensitive to the varied range of the near-field source.

The maximal number of sources which can be uniquely estimated by the proposed algorithm is 8. Four near-field sources are located at $(r_1, \Phi_1) = (0.16\lambda, 102^\circ, 45^\circ)$, $(r_2, \Phi_2) = (0.29\lambda, 122^\circ, 60^\circ)$, $(r_3, \Phi_3) = (0.44\lambda, 60^\circ, 10^\circ)$, and $(r_4, \Phi_4) = (0.51\lambda, 155^\circ, 75^\circ)$, respectively. Four far-field sources are located at $(r_5, \Phi_5) = (\infty, 40^\circ, 15^\circ)$, $(r_6, \Phi_6) = (\infty, 168^\circ, 40^\circ)$, $(r_7, \Phi_7) = (\infty, 10^\circ, 4^\circ)$, and $(r_8, \Phi_8) = (\infty, 140^\circ, 70^\circ)$, respectively. When the SNR is 40 dB, the estimated RMSEs associated with the eight sources can be seen in Table 1. They are lower than 1° except for the azimuth of source 7.

In the second simulation, our proposed algorithm is used to localize pure far-field sources. Two sources are located at $(r_1, \Phi_1) = (\infty, 60^\circ, 122^\circ)$ and $(r_2, \Phi_2) = (\infty,$

$45^\circ, 68^\circ)$, respectively. The number of snapshots is set equal to 128. When SNR varies from 0 to 30 dB, the RMSEs of azimuths and elevations are shown in Figure 8. When SNR increases, the estimation performance of azimuth and elevation gets better.

We consider two sources which are located at $(r_1, \Phi_1) = (\infty, 45^\circ, 22^\circ)$ and $(r_2, \Phi_2) = (\infty, 45^\circ, 22^\circ + \Delta\varphi)$, with $\Delta\varphi$ varying from 0° to 90° . When SNR varies from 0 to 40 dB, the RMSEs of azimuths and elevations are shown in Figure 9. When the SNR is low, the angular estimation RMSEs are larger for all the azimuth differences than those in high SNRs. Moreover, the azimuth difference of the two sources has more effect in low SNRs than high SNRs. In low SNRs, when the two sources are close, the algorithm estimates the azimuth and elevation with larger RMSEs. In high SNRs, the azimuth difference has slight effect on the localization performance. Therefore, the algorithm can estimate two close sources accurately in high SNRs. When the second source has the same azimuth

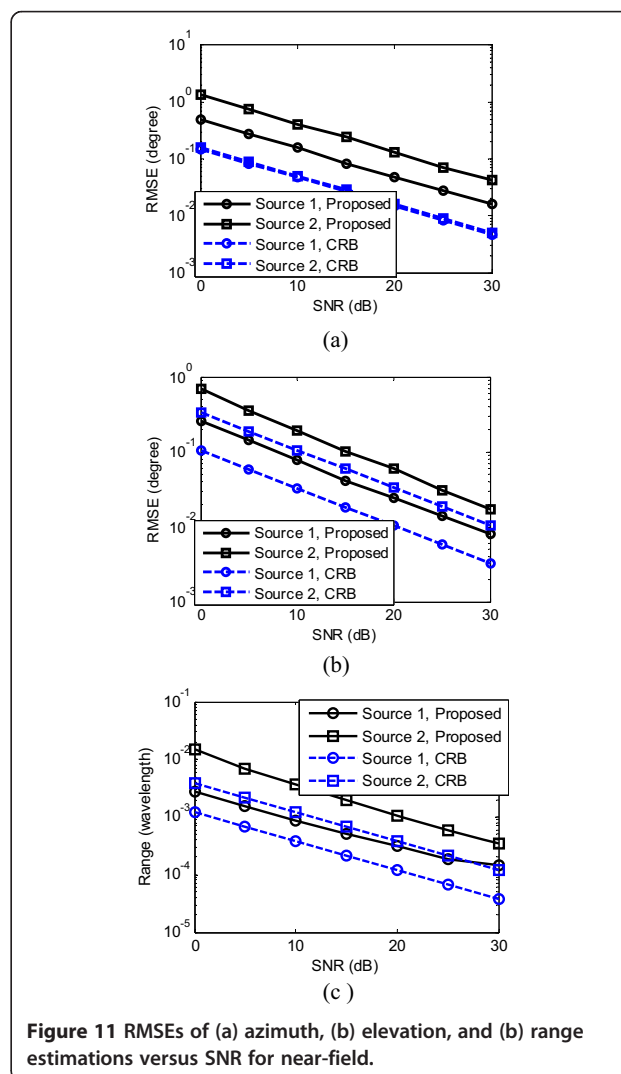


Figure 11 RMSEs of (a) azimuth, (b) elevation, and (b) range estimations versus SNR for near-field.

with the first one, its elevation is $45^\circ + \Delta\theta$, with $\Delta\theta$ varying from 0° to 25° . The localization performance is shown in Figure 10. For different SNRs, the localization performance with the varying elevation is similar to that with varying azimuth. However, when the elevation of the second source is close to 90° in low SNRs, the estimation errors get larger because the elevation estimation is determined by the tangent function.

In the third simulation, the proposed method is adopted to localize pure near-field sources. Two near-field sources are located at $(r_1, \Phi_1) = (0.22\lambda, 45^\circ, 68^\circ)$ and $(r_2, \Phi_2) = (0.44\lambda, 60^\circ, 122^\circ)$, respectively. When the ranges of the sources are within the near-field extent of the array, the RMSEs of the azimuths, elevations, and ranges are shown in Figure 11. From Figure 11, it can be seen that the RMSEs of the azimuths, elevations, and ranges for the first source,

which is closer to the spherical array, are smaller than those of the second source. When the range of the second source varies from 0.2λ to 0.5λ , the angular and range estimations are shown in Figure 12. The angular and range estimation errors of the source closer to the array are smaller than those of the other source. When the range of the second source is closer to the boundary of the near-field extent, the azimuth, elevation, and range estimation errors get larger for almost all SNRs from 0 to 30 dB.

4.2 Wideband source localization

The proposed algorithm can localize multiple wideband mixed field sources. A female signal and a male signal randomly chosen from TIMIT database [50] are incident from the direction $(45^\circ, 68^\circ)$ and $(60^\circ, 122^\circ)$. The male signal source is located in the far-field of the array. The female

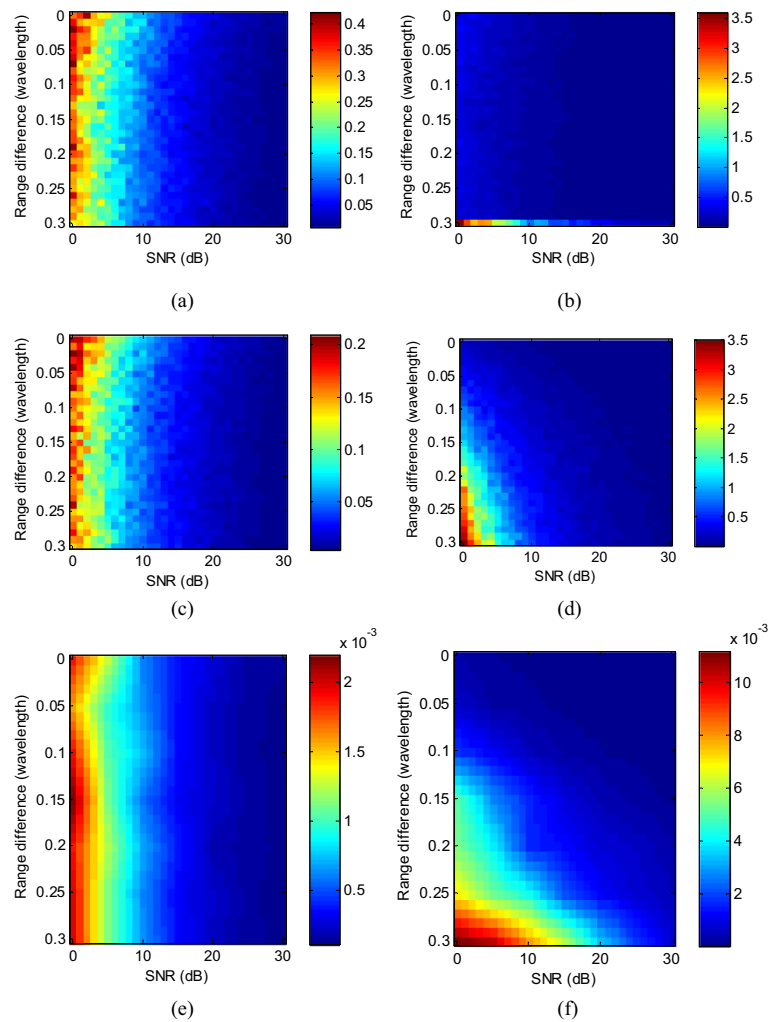


Figure 12 RMSEs of azimuth, elevation, and range estimations versus range difference and SNR for near-field. (a) Azimuth RMSE of source 1. (b) Azimuth RMSE of source 2. (c) Elevation RMSE of source 1. (d) Elevation RMSE of source 2. (e) Range RMSE of source 1. (f) Range RMSE of source 2.

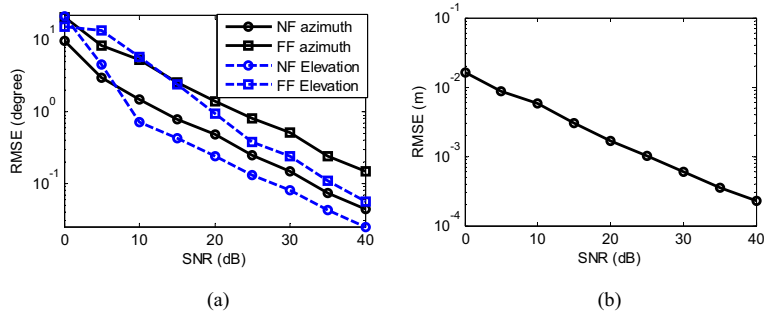


Figure 13 RMSEs of angular and range estimations versus SNR for wideband mixed field sources. (a) Angular estimation RMSE versus SNR. (b) Range estimation RMSE versus SNR.

signal source moves to the near-field of the array with range $r = 0.2 m$. The signal sampling frequency is 16 kHz with 16 bits per sample. The observed data are decomposed into frequency bins using a short-time Fourier transform (STFT) of length 512 with a rectangular window. For each frequency bin, the sample covariance matrix $\mathbf{R}_{nm}(k)$ is estimated with Q frequency snapshots as follows:

$$\mathbf{R}_{nm}(k) = \frac{1}{Q} \sum_{q=1}^Q \mathbf{R}_{nm}(k, q) = \frac{1}{Q} \sum_{q=1}^Q \hat{\mathbf{x}}_{nm}(k, q) \hat{\mathbf{x}}_{nm}^H(k, q), \quad (42)$$

where $\hat{\mathbf{x}}_{nm}(k, q)$ is the q th snapshot of $\mathbf{x}_{nm}(k)$. The operating frequency bandwidth is limited by aliasing at the higher frequencies and measurement errors at the lower frequencies [43]. Therefore, the frequency bins we used are limited to around $2 \leq kr \leq N$ due to errors and spatial aliasing. The SNR is 10 dB. When the value of kr is within the extent (2, 4), the RMSEs of azimuth and elevation for two sources are smaller. Therefore, we choose the frequency bins satisfying the constraint $2 \leq kr \leq N$ to localize wideband sources.

The wideband source localization results for different SNRs are denoted in Figure 13. The near-field source angular estimates are more accurate than those of the far-field source. The length of the observed signals affects the

calculation of the sample covariance matrix in (42) [51]. When the size of the available data is small, the number of snapshots Q for each frequency bin is small and estimating the covariance matrix correctly is harder. When the signal size is large, the correct sample covariance matrix estimation leads to better localization performance shown in Figure 14. The SNR is 10 dB. The RMSE changes of the near-field range estimation are consistent with those of azimuth and elevation estimates, because the range estimate is based on the angular estimates.

5 Conclusions

In this paper, we developed a two-stage source localization algorithm jointly estimating elevation, azimuth angles, and range for the mixed far-field and near-field sources using spherical array. In the first stage, the 3D localization algorithm estimated the DOAs of all mixed sources. In the second stage, 1D MUSIC method was used to distinguish far-field and near-field sources and provided the ranges of the near-field sources based on the estimated DOAs. The algorithm had good performance for azimuth, elevation, and range estimations. It had low computational cost because it avoided multidimensional search and did not require parameter pairing procedure. The estimation performance of the far-field sources was not sensitive to the varied range of the near-field sources. However, the

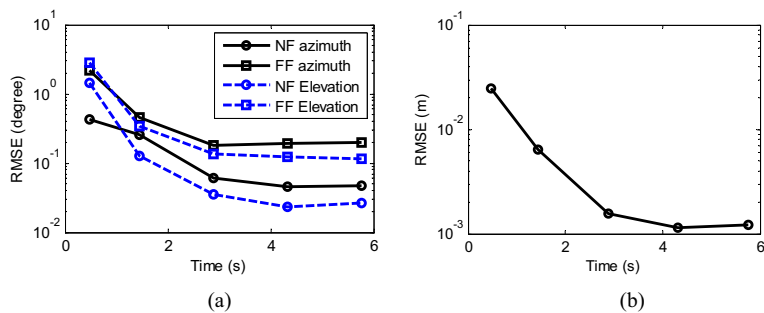


Figure 14 RMSEs of angular and range estimations versus time for wideband mixed field sources. (a) Angular estimation RMSE versus time. (b) Range estimation RMSE versus time.

RMSEs of the azimuths, elevations, and ranges for the near-field source, which was closer to the spherical array, were smaller than those of the source which was farther to the array. Spherical array had better performance for elevation estimation due to its larger aperture in the elevation plane than azimuth estimation. In our future work, we will develop range estimation algorithm without 1D search and incorporate a reverberated signal model to localize multiple sources in reverberant environment.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The authors would like to thank the editor and anonymous reviewers for their valuable comments. The work was supported by the National Natural Science Foundation (61001160), Innovation Program of Shanghai Municipal Education Commission (12YZ23), and Visiting Scholar Funding of Shanghai Municipal Education Commission of China.

Author details

¹School of Communication and Information Engineering, Shanghai University, Shanghai 200072, People's Republic of China. ²Institute of Computer and Information, Shanghai Second Polytechnic University, Shanghai 201209, People's Republic of China.

Received: 30 October 2013 Accepted: 22 May 2014

Published: 14 June 2014

References

1. H Krim, M Viberg, Two decades of array signal processing research: the parametric approach. *IEEE Signal Process Mag* **13**(4), 67–94 (1996)
2. JC Chen, K Yao, RE Hudson, Acoustic source localization and beamforming: theory and practice. *EURASIP J Appl Signal Process* **4**, 359–370 (2003)
3. L Liang, D Liu, Passive localization of near-field sources using cumulant. *IEEE Sensors J* **9**(8), 953–960 (2009)
4. B Wang, J Liu, X Sun, Mixed sources localization based on sparse signal reconstruction. *IEEE Signal Process Lett* **19**(8), 487–490 (2012)
5. L Liang, D Liu, Passive localization of mixed near-field and far-field sources using two-stage MUSIC algorithm. *IEEE Trans Signal Process* **58**(1), 108–120 (2010)
6. J He, MNS Swamy, MO Ahmad, Efficient application of MUSIC algorithm under the coexistence of far-field and near-field sources. *IEEE Trans Signal Process* **60**(4), 2066–2070 (2012)
7. B Wang, Y Zhao, J Liu, Mixed-order MUSIC algorithm for localization of far-field and near-field sources. *IEEE Signal Process Lett* **20**(4), 311–314 (2013)
8. J Jiang, F Duan, J Chen, Y Li, X Hua, Mixed near-field and far-field sources localization using the uniform linear sensor array. *IEEE Sensors J* **13**(8), 3136–3143 (2013)
9. J Jiang, F Duan, J Chen, Three-dimensional localization algorithm for mixed near-field and far-field sources based on ESPRIT and MUSIC method. *Prog Electromagnetics Res* **136**, 435–456 (2013)
10. CP Mathews, MD Zoltowski, Eigenstructure techniques for 2-D angle estimation with uniform circular arrays. *IEEE Trans Signal Process* **42**(9), 2395–2407 (1994)
11. D Pavlidi, A Griffin, M Puigt, A Mouchtaris, Real-time multiple sound source localization and counting using a circular microphone array. *IEEE Trans Audio Speech Lang Process* **21**(10), 2193–2206 (2013)
12. G Sommerkorn, D Hampicke, R Klukas, A Richter, A Schneider, R Thomä, Uniform rectangular antenna array design and calibration issues for 2-D ESPRIT application, in *The 4th European Personal Mobile Communications Conference, Vienna*, ed. by (Morgan Kaufman, San Francisco, 2001), pp. 1–8
13. P Ioannides, CA Balanis, Uniform circular and rectangular arrays for adaptive beamforming applications. *IEEE Antennas Wireless Propagation Lett* **4**, 351–354 (2005)
14. P Smaragdīs, P Boufounos, Position and trajectory learning for microphone arrays. *IEEE Trans Audio Speech Lang Process* **15**(1), 358–368 (2007)
15. M Costa, V Koivunen, A Richter, Low complexity azimuth and elevation estimation for arbitrary array configurations, in *IEEE Int. Conf. Acoust., Speech, Signal Process, Taipei*, ed. by (IEEE, Piscataway, 2009), pp. 2185–2188
16. F Belloni, A Richter, V Koivunen, D. O. A. Estimation, via manifold separation for arbitrary array structures. *IEEE Trans. Signal. Process.* **55**(10), 4800–4810 (2007)
17. T Filik, TE Tuncer, A fast and automatically paired 2-dimensional direction-of-arrival estimation using arbitrary array geometry, in *IEEE 17th Signal Process and Communications Applications Conference, Antalya*, ed. by (IEEE, Piscataway, 2009), pp. 556–559
18. M Park, B Rafaely, Sound-field analysis by plane-wave decomposition using spherical microphone array. *J Acoust Soc Am* **118**(5), 3094–3103 (2005)
19. S Yan, H Sun, UP Svensson, X Ma, JM Hovem, Optimal modal beamforming for spherical microphone arrays. *IEEE Trans Audio Speech Lang Process* **19**(2), 361–371 (2012)
20. D Khaykin, B Rafaely, Acoustic analysis by spherical microphone array processing of room impulse response. *J Acoust Soc Am* **132**(1), 261–270 (2012)
21. H Teutsch, *Modal Array Signal Processing – Principles and Applications of Acoustic Wavefield Decomposition* (Springer-Verlag, Heidelberg, 2007), pp. 134–146
22. I Cohen, J Benesty, S Gannot, *Speech Processing in Modern Communication* (Springer-Verlag, Heidelberg, 2010), p. 281
23. R Goossens, H Rogier, Unitary spherical ESPRIT: 2-D angle estimation with spherical arrays for scalar fields. *IET Signal Process* **3**(2), 221–231 (2009)
24. H Sun, W Kellermann, E Mabande, Robust localization of multiple sources in reverberant environments using EB-ESPRIT with spherical microphone arrays, in *IEEE Int. Conf. Acoust., Speech, Signal Process, Prague*, ed. by (IEEE, Piscataway, 2011), pp. 117–120
25. S Argentieri, P Danes, Broadband variations of the MUSIC high-resolution method for sound source localization robotics, in *IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego*, ed. by (IEEE, Piscataway, 2007), pp. 2009–2014
26. H Sun, E Mabande, K Kowalczyk, W Kellermann, Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing. *J Acoust Soc Am* **131**(4), 2828–2840 (2012)
27. PKT Wu, N Epain, C Jin, A super-resolution beamforming algorithm for spherical microphone arrays using a compressive sensing approach, in *Conf. Acoust., Speech, Signal Process, Vancouver*, ed. by (IEEE, Piscataway, 2013), pp. 649–653
28. H Sawada, R Mukai, S Araki, S Malcino, Multiple source localization using independent component analysis, in *IEEE Antennas and Propagation Society International Symposium, Washington DC*, ed. by (IEEE, Piscataway, 2005), pp. 81–84
29. N Epain, C Jin, Independent component analysis using spherical microphone arrays. *Acta Acustica United Acustica* **98**(1), 91–102 (2012)
30. T Noohi, N Epain, C Jin, Direction of arrival estimation for spherical microphone arrays by combination of independent component analysis and sparse recovery, in *Conf. Acoust., Speech, Signal Process, Vancouver*, ed. by (IEEE, Piscataway, 2013), pp. 346–349
31. D Malioutov, M Cetin, AS Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans Signal Process* **53**(8), 3010–3022 (2005)
32. X Wei, Y Yuan, Q Ling, DOA estimation using a greedy block coordinate descent algorithm. *IEEE Trans Signal Process* **60**(4), 2066–2070 (2012)
33. M Swartling, B Sallberg, N Grbic, Source localization for multiple speech sources using low complexity non-parametric source separation and clustering. *Signal Process* **91**, 1781–1788 (2011)
34. J Meyer, GW Elko, Position independent close-talking microphone. *Signal Process* **86**(6), 1254–1259 (2006)
35. E Fisher, B Rafaely, The nearfield spherical microphone array, in *IEEE Int. Conf. Acoust., Speech, Signal Process, Las Vegas*, ed. by (IEEE, Piscataway, 2008), pp. 5272–5275
36. E Fisher, B Rafaely, Near-field spherical microphone array processing with radial filtering. *IEEE Trans Audio Speech Lang Process* **19**(2), 256–265 (2011)
37. Q Huang, T Song, DOA estimation of mixed near-field and far-field sources using spherical array, in *The 11th Int. Conf. on Signal Process, Beijing*, ed. by (IEEE, Piscataway, 2012), pp. 382–385
38. EG Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, New York, 1999), p. 183
39. B Rafaely, B Weiss, E Bachmat, Spatial aliasing in spherical microphone arrays. *IEEE Trans Signal Process* **55**(3), 1003–1010 (2007)

40. J Meyer, G Elko, A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield, in *IEEE Int. Conf. Acoust., Speech, Signal Process, Orlando*, ed. by (IEEE, Piscataway, 2002), pp. II-1781–II-1784
41. TD Abhayapala, DB Ward, Theory and design of high order sound field microphones using spherical microphone array, in *IEEE Int. Conf. Acoust., Speech, Signal Process Orlando*, ed. by (IEEE, Piscataway, 2002), pp. II-1949–II-1952
42. H Teutsch, W Kellermann, Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures, in *IEEE Int Conf Acoust, Speech, Signal Process, Las Vegas*, ed. by (IEEE, Piscataway, 2008), pp. 5276–5279
43. B Rafaely, Analysis and design of spherical microphone arrays. *IEEE Trans Speech Audio Process* **13**(1), 135–143 (2005)
44. RO Schmidt, Multiple emitter location and signal parameter estimation. *IEEE Trans Antennas Propag* **AP-34**, 276–280 (1986)
45. Y Wang, J Chen, W Fang, TST-MUSIC for joint DOA-delay estimation. *IEEE Trans Signal Process* **49**(4), 721–729 (2001)
46. P Stoica, EG Larsson, AB Gershman, The stochastic CRB for array processing: a textbook derivation. *IEEE Signal Process Lett* **8**(5), 148–150 (2001)
47. MN El Korso, R Boyer, A Renaux, S Marcos, Conditional and unconditional Cramér–Rao bounds for near-field source localization. *IEEE Trans Signal Process* **58**(5), 2901–2907 (2010)
48. B Gardner, A realtime multichannel room simulator. *J Acoust Soc Am* **92**(4), 2395 (1992)
49. R Vershynin, How close is the sample covariance matrix to the actual covariance matrix? *J Theor Probabil* **25**, 655–686 (2012)
50. LF Lamel, RH Kasel, S Senneff, Speech database development: design and analysis of the acoustic-phonetic corpus, in *Proc. of the DARPA Speech Recognition Workshop*, ed. by (IET, Glasgow, 1986), pp. 100–109
51. M Puigt, E Vincent, Y Deville, Validity of the independence assumption for the separation of instantaneous and convolutive mixtures of speech and music sources, in *Int. Conf. ICA, LNCS, Brazil*, ed. by (Springer, New York, 2009), pp. 613–620

doi:10.1186/1687-6180-2014-90

Cite this article as: Huang and Wang: Acoustic source localization in mixed field using spherical microphone arrays. *EURASIP Journal on Advances in Signal Processing* 2014 **2014**:90.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
