Broadband Sound Source Localisation via Non-synchronous Measurements for Service Robots: A Tensor Completion Approach

Long Chen[1], *Member, IEEE,* Weize Sun[2], *Member, IEEE,* Lei Huang[2], *Senior Member, IEEE,* and Liang Yu[3], *Member, IEEE*

***Abstract*—Constraint by the physical geometry, the lower and upper frequency bound and the scale of the scanning area of a microphone array are limited. Owing to its movable feature, for the service robots, achieving a wider working frequency range with a global view requires a virtually larger and denser array, which can be realised using non-synchronous measurements beamforming with a movable microphone array prototype. However, even when using the state-of-the-art method, it is challenging to localise multiple broadband sources, owing to the difficulty in selecting an appropriate operating frequency without any prior information about the target signal. Therefore, this letter proposes a tensor-completion-based non-synchronous measurements method for broadband multiple-sound-source localisation. The tensor data structure of the broadband signal is analysed, and an alternating direction method based on multiplier optimisation with a tensor multi-norm constraint is proposed. This algorithm can provide a sound map with a distinct global view of three different speech signal sources with high accuracy. Compared with the matrix-based optimisation method, the proposed method can significantly reduce the mean square error of the estimated source location.**

***Index Terms*—Localization, Service Robotics.**

I. Introduction

IT is essential for a service robot to locate sound sources by sound only via sound source localization (SSL) techniques [1]–[3]. The functionality of SSL on a robotic platform could be useful in several situations, for instance, locating human speakers without visual contact and mapping an unknown acoustic environment [4]. It has been achieved by various methodologies, such as head-related transfer function (HRTF) based time-difference-of-arrival (TDOA), space-domain distance (SDD), acoustic beamforming, etc [5]–[9]. Among those techniques, the acoustic beamforming is widely used to obtain the sound map of a measured field in industrial applications such as transporting pass-by noise localisation and machine fault detection [4], [10]–[13]. Especially, for the research area of robot audition, the broadband multiple signal classification (MUSIC) method has been successfully applied to robotics to detect and localize broadband sources [14]–[16]. On the other hand, the generalized cross-correlation with phase transformation (GCC-PHAT) method is one of the most popular TDOA-based method for time-delay as well as sound source location estimation [17]–[19]. Both of the two methods are very popular for SSL, and some comparisons of these two methods could be found in [20], [21]. Generally, MUSIC is known to be more robust to noise than GCC-PHAT, and it is more effective when using an array containing large numbers of microphones [20]. While the GCC-PHAT method is one of the most popular SSL method due to its robustness against moderate levels of reverberation by using the difference of arrival time between two microphone pairs and maximum time-lag yields the location of the sound sources [21], [22]. However, in practice, the working frequency range of a microphone array is typically limited by array size and the distance between two adjacent microphones. Consequently, many research attempts to improve the spatial distribution of microphones have been undertaken over the past decades [23]. The topological design of a microphone array, for example, sparse, random, and spiral, and its corresponding parameters such as the co-array and point spread function (PSF), have been thoroughly discussed in previous research for various potential applications [24]–[26].

To overcome this limitation, the non-synchronous measurements (NSM) method was proposed for acoustic imaging in near-field acoustical holography (NAH) [27]. Both the size of the microphone array and microphone density became flexible by sequentially moving a small prototype array during measurements. Actually, in order to access a global view under an unknown acoustic environment with a

Manuscript received: June 24, 2022; Revised: September 7, 2022; Accepted: October 3, 2022. This paper was recommended for publication by Editor Sven Behnke upon evaluation of the Associate Editor and Reviewers' comments. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 62101335, U1713217, 12074254, and 61925108, and in part by the Natural Science Foundation of Guangdong under Grant 2021A1515011706. *(Corresponding author: Weize Sun.)*

[1]Long Chen is with the School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an 710072, China, and also with the College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518000, China. chenmf767@foxmail.com

[2]Weize Sun and Lei Huang are with the College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518000, China. proton198601@hotmail.com; lhuang8sasp@hotmail.com

[3]Liang Yu is with the Institute of Vibration, Shock, and Noise, State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China. liang.yu@sjtu.edu.cn

Digital Object Identifier (DOI): see top of this page.

broader working frequency range, it is suitable to exploit a microphone array on a moving robotic platform. Thus, in one of our former works, the NSM method was subsequently implemented for acoustic beamforming using a cross-spectral matrix (CSM) completion approach [28]. From the perspective of optimisation, a low-rank matrix completion problem is equivalent to minimising the matrix rank, which can be relaxed to a nuclear norm minimisation problem because the rank minimisation is NP-hard [29], [30]. In our previous study, two different algorithms—the augmented Lagrange multiplier (ALM) and alternative direction method of multipliers (ADMM)—were applied to solve the raised optimisation problem; the algorithms were also compared and discussed [28]. Based on this study, the NSM approach was further improved via block Hermitian matrix completion (BHMC), variational Bayesian (VB) inference, and coprime positions (CP-NSM) for more accurate and robust source localisation results [31]–[33].

The working frequency range of acoustic beamforming can be broadened by the NSM method using a virtually larger and denser array. However, all of the abovementioned localisation approaches were conducted at one specific frequency. Therefore, it is essential to obtain the frequency-domain information of the target broadband sound signals before acoustic beamforming is conducted, which is typically impossible in practical applications. To solve the frequency bin selection problem for broadband signals, we propose a tensor singular value decomposition (t-SVD) and tensor completion (TC)-based NSM approach for broadband multiple-sound-source localisation (BM-SSL); this approach considers the complexity and multidimensional structural features of received signals in the frequency, time, and space domains [34]. In this case, the observed tensor structure can be fully utilised to complete the target data by minimising the defined tensor rank [35].

II. Tensor Singular Value Decomposition and Tensor Nuclear-norm

To easily understand the TC method, we first introduce the concept of t-product, t-SVD and tensor nuclear-norm (TNN) before applying the NSM signal model. The t-product operation $\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{B}}$ is a kind of matrix multiplication over the two first dimensions of tensor $\boldsymbol{\mathcal{A}}$ and tensor $\boldsymbol{\mathcal{B}}$ with the elementwise multiplication replaced by circular convolution along tubes [36]. Based on the t-product operation, the t-SVD of a three dimensional tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{D_1 \times D_2 \times D_3}$ is defined as [34], [37]:

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^H = \sum_{d=1}^{D_r} \boldsymbol{\mathcal{U}}(:,d,:) * \boldsymbol{\mathcal{S}}(d,d,:) * \boldsymbol{\mathcal{V}}(:,d,:)^{\boldsymbol{\mathcal{H}}}, \quad (1)$$

where $D_r \leq \min\{D_1, D_2\}$, $*$ is the t-product, $(\cdot)^{\boldsymbol{\mathcal{H}}}$ is the tensor complex conjugate transpose [34], [36], and $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{V}}$ are orthogonal tensors that satisfy

$$\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{U}}^{\boldsymbol{\mathcal{H}}} = \boldsymbol{\mathcal{U}}^{\boldsymbol{\mathcal{H}}} * \boldsymbol{\mathcal{U}} = \boldsymbol{\mathcal{I}} \in \mathbb{C}^{D_1 \times D_1 \times D_3}, \quad (2)$$

and

$$\boldsymbol{\mathcal{V}} * \boldsymbol{\mathcal{V}}^{\boldsymbol{\mathcal{H}}} = \boldsymbol{\mathcal{V}}^{\boldsymbol{\mathcal{H}}} * \boldsymbol{\mathcal{V}} = \boldsymbol{\mathcal{I}} \in \mathbb{C}^{D_2 \times D_2 \times D_3}, \quad (3)$$

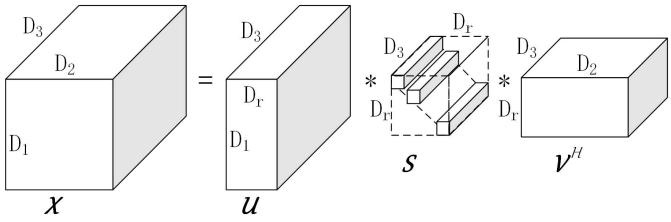

Fig. 1. Illustration of the t-SVD of an order-3 tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{D_1 \times D_2 \times D_3}$.

respectively, $\boldsymbol{\mathcal{I}}$ represents the identity tensor, and for any $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{D_1 \times D_2 \times D_3}$, $\boldsymbol{\mathcal{X}}^{\boldsymbol{\mathcal{H}}}$ takes the tensor complex conjugate transpose of each entry of $\boldsymbol{\mathcal{X}}$ [36]. Note that the identity tensor $\boldsymbol{\mathcal{I}} \in \mathbb{C}^{d \times d \times d_3}$ is the tensor with its first frontal slice being the $d \times d$ identity matrix, and other frontal slices being all zeros [34]. Here, we define $\tilde{\boldsymbol{\mathcal{X}}} = \text{fft}(\boldsymbol{\mathcal{X}}, [], 3)$ as the Discrete Fourier Transform (DFT) along the 3rd dimension, where $\text{fft}(\cdot)$ is the fast Fourier transform (FFT) operator. It is clear that $\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{I}} = \boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{I}} * \boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}$ given the appropriate dimensions. And $\tilde{\boldsymbol{\mathcal{I}}}$ is a tensor with each frontal slice being the identity matrix [36]. $\boldsymbol{\mathcal{S}}$ is an F-diagonal tensor, *i.e.*, its frontal slices are all diagonal matrices, and it contains all the D_r singular value tubes of tensor $\boldsymbol{\mathcal{X}}$ [36]. At last, the tensor average rank (TAR) of $\boldsymbol{\mathcal{X}}$ is defined as:

$$\|\boldsymbol{\mathcal{X}}\|_{TAR} = \frac{1}{D_3} \sum_{d_3=1}^{D_3} \text{rank}[\tilde{\boldsymbol{\mathcal{S}}}(:,:,d_3)]. \quad (4)$$

Considering the norms are often used as the surrogates of ranks in view of their convexity, the TNN [36], [37] defined as

$$\|\boldsymbol{\mathcal{X}}\|_{TNN} = \frac{1}{D_3} \sum_{d=1}^{D_r} \sum_{d_3=1}^{D_3} \tilde{\boldsymbol{\mathcal{S}}}(d,d,d_3) \quad (5)$$

is used instead of equation (4). An illustration of the t-SVD of an order-3 tensor with $D_r < \min\{D_1, D_2\}$, which is, at least one non-zero tube exists in the diagonal tubes of $\boldsymbol{\mathcal{S}}$, is shown in Fig. 1. Note that the t-SVD can be computed by using the FFT and the inverse fast Fourier transform (IFFT) [34], and the t-SVD pseudo-code is provided in Algorithm 1.

Algorithm 1: T-SVD
1: Start with a three dimensional tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{D_1 \times D_2 \times D_3}$;
2: **for** $i = 1 : D_3$
3: $[U, S, V] = svd(\tilde{\boldsymbol{\mathcal{X}}}(:,:,i))$;
4: $\tilde{\boldsymbol{\mathcal{U}}}(:,:,i) = U$; $\boldsymbol{\mathcal{U}}(:,:,i) = ifft(\tilde{\boldsymbol{\mathcal{U}}}, [], 3)$;
5: $\tilde{\boldsymbol{\mathcal{V}}}(:,:,i) = V$; $\boldsymbol{\mathcal{V}}(:,:,i) = ifft(\tilde{\boldsymbol{\mathcal{V}}}, [], 3)$;
6: $\tilde{\boldsymbol{\mathcal{S}}}(:,:,i) = S$; $\boldsymbol{\mathcal{S}}(:,:,i) = ifft(\tilde{\boldsymbol{\mathcal{S}}}, [], 3)$;
7: **end**
8: Output $\boldsymbol{\mathcal{U}}$, $\boldsymbol{\mathcal{V}}$, and $\boldsymbol{\mathcal{S}}$.

III. Signal Model and Former Solution

Consider a near-field acoustic signal propagating model with an NSM, as shown in Fig. 2. The distance between the planar microphone array and the measurement plane is h. In the NSM process, a prototype array consisting of M

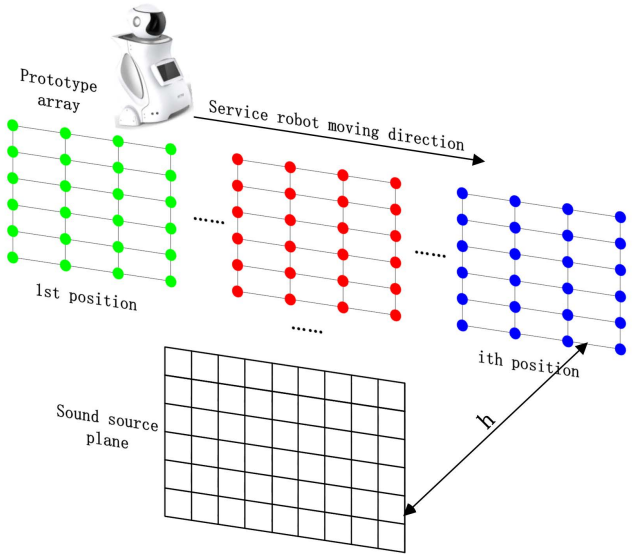

Fig. 2. Near-field acoustic signal propagating model with NSM.

microphones moves N times from the first position to the Nth position along the arrow direction.

It is assumed that p_m^n is one of the measured complex amplitude in the N measurements by the array at one specific frequency. The measured complex acoustic field can be expressed in vector form as:

$$\mathbf{p}^n = [p_1^n \quad p_2^n \quad \cdots \quad p_M^n]^T, \tag{6}$$

and the CSM $\mathbf{R}^n \in \mathbb{C}^{M\times M}$ is

$$\mathbf{R}^n = E[\mathbf{p}^n \mathbf{p}^{n^H}], \tag{7}$$

where $E[\cdot]$ denotes the mathematical expectation and $n = (1, 2, \cdots, N)$.

Considering that the entire NSM procedure consists of N times measurements, the N CSMs $\mathbf{R}^n \in \mathbb{C}^{M\times M}$ can be regarded as the diagonal blocks in a full CSM $\mathbf{R} \in \mathbb{C}^{MN\times MN}$ obtained via a synchronous measurement; these CSMs consist of all the microphone arrays in different positions. With this understanding, the observed $\mathbf{R}_\Omega$ by the NSM can be expressed as

$$\mathbf{R}_\Omega = \begin{bmatrix} \mathbf{R}^1 & & \\ & \ddots & \\ & & \mathbf{R}^N \end{bmatrix}, \tag{8}$$

where Ω is a sampling operator that extracts elements from the diagonal blocks. To obtain a fully estimated CSM $\hat{\mathbf{R}} \in \mathbb{C}^{MN\times MN}$ from N CSMs $\mathbf{R}^n$, the SSL problem is reformulated to a low-rank diagonal block matrix completion problem, which was solved by minimising the rank of the matrix in our previous study [28]. The sound-source locations can be obtained using the MUSIC method [38]–[40]:

$$\beta_M(r_m) = \frac{1}{\mathbf{w}(r_m)^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{w}(r_m)}, \tag{9}$$

where $\mathbf{w}(r_m)$ is the steering vector from the signal subspace and r_m is the distance between the sound source and the mth receiver microphone. r_m could be calculated by $w(r_m) = \sqrt{(x-x_m)^2 + (y-y_m)^2 + (z-z_m)^2}$, in which the source location is assumed at (x, y, z) and the coordinates of the mth microphone are (x_m, y_m, z_m). And $\mathbf{U}_n$ are the eigenvectors from the noise subspace after the eigen-decomposition of $\hat{\mathbf{R}}$.

IV. Proposed Solution

The localisation approach using the proposed ADMM-based NSM method was conducted at one specific frequency in a previous study [28]. However, an improper selection of the frequency bin would lead to an inaccurate or even incorrect source location in the pseudo-spectra. To mitigate this challenge, we define $\boldsymbol{\mathcal{R}}_\Omega \in \mathbb{C}^{MN\times MN\times S}$ as the observed tensor, which comprises $N \times S$ known submatrices $\mathbf{R}^n \in \mathbb{C}^{M\times M}$ at every frequency bin; and $\boldsymbol{\mathcal{R}} \in \mathbb{C}^{MN\times MN\times S}$ is the target cross-spectral tensor (CST). $\boldsymbol{\mathcal{R}}_\Omega$ can be expressed as

$$\boldsymbol{\mathcal{R}}_\Omega = R_{1\Omega} \sqcup_3 R_{2\Omega} \sqcup_3 \cdots \sqcup_3 R_{S\Omega}, \tag{10}$$

where $f = 1, 2, \cdots, S$ is the number of frequency slices, and $\sqcup_3$ is the concatenation operation along the third dimension. Similarly, to seek a fully estimated CST $\hat{\boldsymbol{\mathcal{R}}}$ from $N \times S$ CSMs $\mathbf{R}^n$, the BM-SSL problem is reformulated as a tensor completion problem, which can be solved by minimising the TNN:

$$\begin{aligned} \min_{\hat{\boldsymbol{\mathcal{R}}}, \boldsymbol{\mathcal{M}}} \quad & \lambda \|\hat{\boldsymbol{\mathcal{R}}}\|_{TNN} + \frac{1}{2}\|\boldsymbol{\mathcal{M}}_\Omega - \boldsymbol{\mathcal{R}}_\Omega\|_F^2 \\ s.t. \quad & \hat{\boldsymbol{\mathcal{R}}} = \boldsymbol{\mathcal{M}}, \quad \hat{\boldsymbol{\mathcal{R}}}_f \succeq 0, \end{aligned} \tag{11}$$

where λ is a regularisation parameter, which was set to be 20 in the following operations, $\hat{\boldsymbol{\mathcal{R}}}_f$ means the f-th frontal slice of $\hat{\boldsymbol{\mathcal{R}}}$, and $\|\cdot\|_F$ denotes the Frobenius norm. Its augmented Lagrangian function is

$$\begin{aligned} L(\hat{\boldsymbol{\mathcal{R}}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{Y}}) = \lambda \|\hat{\boldsymbol{\mathcal{R}}}\|_{TNN} + \frac{1}{2}\|\boldsymbol{\mathcal{M}}_\Omega - \boldsymbol{\mathcal{R}}_\Omega\|_F^2 \\ + \langle \boldsymbol{\mathcal{Y}}, \hat{\boldsymbol{\mathcal{R}}} - \boldsymbol{\mathcal{M}} \rangle + \frac{\mu}{2}\|\hat{\boldsymbol{\mathcal{R}}} - \boldsymbol{\mathcal{M}}\|_F^2. \end{aligned} \tag{12}$$

Here, we define $\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle = vec\{\boldsymbol{\mathcal{A}}\}^{\boldsymbol{\mathcal{H}}} vec\{\boldsymbol{\mathcal{B}}\}$, where $vec\{\boldsymbol{\mathcal{A}}\}$ is the vectorisation operation of the tensor $\boldsymbol{\mathcal{A}}$. $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{MN\times MN\times S}$ is the Lagrange multiplier, and μ is a positive penalty parameter, which was set to be $\frac{24.5}{M}$ in the following operations. Then, $\hat{\boldsymbol{\mathcal{R}}}$, $\boldsymbol{\mathcal{M}}$, and $\boldsymbol{\mathcal{Y}}$ can be updated iteratively using ADMM. By fixing $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{Y}}$, $\hat{\boldsymbol{\mathcal{R}}}$ can be updated by

$$\hat{\boldsymbol{\mathcal{R}}}^{(k+1)} = \arg\min_{\hat{\boldsymbol{\mathcal{R}}}_f \succeq 0} \lambda \|\hat{\boldsymbol{\mathcal{R}}}\|_{TNN} + \frac{\mu}{2}\|\hat{\boldsymbol{\mathcal{R}}} - \boldsymbol{\mathcal{M}}^{(k)} + \frac{1}{\mu}\boldsymbol{\mathcal{Y}}^{(k)}\|_F^2, \tag{13}$$

where the superscript (k) denotes the k-th iteration. According to [34] and [37], equation (13) is equivalent to its FFT format, and can be further reduced to

$$\begin{aligned} \tilde{\boldsymbol{\mathcal{R}}}_f^{(k+1)} &= \arg\min_{\tilde{\boldsymbol{\mathcal{R}}}_f \succeq 0} \lambda' \|\tilde{\boldsymbol{\mathcal{R}}}_f\|_* + \frac{\mu'}{2}\|\tilde{\boldsymbol{\mathcal{R}}}_f - \tilde{\boldsymbol{\mathcal{M}}}_f^{(k)} + \frac{1}{\mu}\tilde{\boldsymbol{\mathcal{Y}}}_f^{(k)}\|_F^2 \\ &= \tilde{\boldsymbol{\mathcal{U}}}_f^{(k)} \cdot \max\{\tilde{\boldsymbol{\mathcal{W}}}_f^{(k)} - \frac{\lambda'}{\mu'}, 0\} \cdot \tilde{\boldsymbol{\mathcal{U}}}_f^{(k)\boldsymbol{\mathcal{H}}}, \end{aligned} \tag{14}$$

where $\boldsymbol{\mathcal{M}}^{(k)} - \frac{1}{\mu}\boldsymbol{\mathcal{Y}}^{(k)} = \boldsymbol{\mathcal{U}}^{(k)} * \boldsymbol{\mathcal{W}}^{(k)} * \boldsymbol{\mathcal{U}}^{(k)\boldsymbol{\mathcal{H}}}$ is the t-SVD decomposition, $\|\cdot\|_*$ denotes the nuclear norm, and λ' and

μ' are the FFT formats of λ and μ. Note that although the actual operation of t-SVD on a tensor is carried out in its matrix slices after FFT one by one, it equals to processing a rearranged matrix of this tensor [34] [37], thus TC method is different from calculating a series of matrix completion problems. And $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{Y}}$ can be updated as follows:

$$\boldsymbol{\mathcal{M}}^{(k+1)} = \arg\min_{\boldsymbol{\mathcal{M}}} \frac{1}{2}\|\boldsymbol{\mathcal{M}}_{\Omega}-\boldsymbol{\mathcal{R}}_{\Omega}\|_F^2+\frac{\mu}{2}\|\boldsymbol{\mathcal{M}}-\frac{1}{\mu}\boldsymbol{\mathcal{Y}}^{(k)}-\hat{\boldsymbol{\mathcal{R}}}^{(k+1)}\|_F^2 \tag{15}$$

$$\boldsymbol{\mathcal{M}}_{\Omega}^{(k+1)} = \frac{1}{\mu+1}[\boldsymbol{\mathcal{R}}_{\Omega} + \mu\hat{\boldsymbol{\mathcal{R}}}_{\Omega}^{(k+1)} + \boldsymbol{\mathcal{Y}}_{\Omega}^{(k)}], \tag{16}$$

$$\boldsymbol{\mathcal{M}}_{\bar{\Omega}}^{(k+1)} = \hat{\boldsymbol{\mathcal{R}}}_{\bar{\Omega}}^{(k+1)} + \frac{1}{\mu}\boldsymbol{\mathcal{Y}}_{\bar{\Omega}}^{(k)}, \tag{17}$$

$$\boldsymbol{\mathcal{Y}}^{(k+1)} = \boldsymbol{\mathcal{Y}}^{(k)} + \gamma\mu[\hat{\boldsymbol{\mathcal{R}}}^{(k+1)} - \boldsymbol{\mathcal{M}}^{(k+1)}], \tag{18}$$

where $\bar{\Omega}$ is a sampling operator that extracts elements from non-diagonal blocks, and γ is a relaxation parameter, which was set to be 2.6 in the following operations. Additionally, the stopping criterion is:

$$\frac{\|\hat{\boldsymbol{\mathcal{R}}}_{\Omega} - \boldsymbol{\mathcal{R}}_{\Omega}\|_F}{\|\boldsymbol{\mathcal{R}}_{\Omega}\|_F} \leq \varepsilon. \tag{19}$$

And N_m is the maximum number of iteration steps.

To obtain a good TNN optimisation, the shrinkage factor $\frac{\lambda'}{\mu'}$ should decrease as the number of iterations increases. Therefore, we set a reducing parameter $0 < \alpha < 1$ to update λ' as $\lambda'^{(k+1)} = \lambda'^{(k)}\alpha$ in each iteration and it will stop decreasing till it reaches $\lambda'_0 = 0.0001$. In the ADMM optimization process, all the singular values less than the thresholding $\frac{\lambda'}{\mu'}$ would be rejected. And this approach is conducted after t-SVD to ensure that we have considered all the frequency bins as far as possible. After that, the noise introduced by the incomplete CST would be reduced, and then the sidelobe levels could be reduced in the sound maps. In other words, in the sound maps, the broadband ADMM-based NSM could achieve an energy concentration at every frequency bin independently. However, the TC-based method could achieve an energy concentration at all frequency bins together. In addition, to ensure the spatial continuity of the acoustic field, for each frequency slice, $\hat{\boldsymbol{\mathcal{R}}}_f = \boldsymbol{\Psi}_f\hat{\boldsymbol{\mathcal{R}}}_f\boldsymbol{\Psi}_f^H$ is set in each iteration by introducing the proposed projection basis $\boldsymbol{\Psi} \in \mathbb{C}^{MN\times MN}$, which satisfies $\boldsymbol{\Psi} = \boldsymbol{\Phi}\boldsymbol{\Phi}^{\dagger}$ [28]. Note that the spatial basis $\boldsymbol{\Phi} \in \mathbb{C}^{MN\times K}$ is a plane Fourier basis satisfies $\boldsymbol{\Phi}(x_m, y_m) = e^{i(\kappa_{x_m}x_m+\kappa_{y_m}y_m)}$, in which $(\cdot)^{\dagger}$ is the pseudo-inverse of the spatial basis and $K = M^{\frac{1}{2}}N$ is the dimension of the spatial basis in our case. Furthermore, the wavenumbers $(\kappa_{x_m}, \kappa_{y_m})$ and the coordinates of the microphones (x_m, y_m) are required to be discretized for constructing this basis $\boldsymbol{\Phi}$ [41], [42]. Introducing the proposed projection basis here is to smooth the spectral matrix and to ensure the spatial continuity of the acoustical field. The orthogonal projection basis $\boldsymbol{\Psi}$ imposes a specific structure that encodes the information on microphone positions into $\hat{\boldsymbol{\mathcal{R}}}_f$. This can be seen as another constraint to be added in the optimization [42], [43]. Therefore, the TC-based NSM problem can be solved using the ADMM algorithm, which is summarised in Algorithm 2.

Algorithm 2: TC-ADMM-based NSM.

Start with $\boldsymbol{\mathcal{Y}}^{(1)} = 0 \in \mathbb{C}^{MN\times MN\times S}$, randomly assigned $\boldsymbol{\mathcal{M}}^{(1)} \in \mathbb{C}^{MN\times MN\times S}$, and input $\hat{\boldsymbol{\mathcal{R}}} = 0 \in \mathbb{C}^{MN\times MN\times S}$. λ, μ, λ', μ', γ, and α are given parameters.
for $k = 1 : N_m$ (N_m is the maximum number of iteration steps)
for $f = 1 : S$
$\hat{\boldsymbol{\mathcal{R}}}_f^{(k+1)} = \boldsymbol{\Psi}_f\hat{\boldsymbol{\mathcal{R}}}_f^{(k)}\boldsymbol{\Psi}_f^H$;
end
update $\hat{\boldsymbol{\mathcal{R}}}$ following (12);
update $\boldsymbol{\mathcal{M}}$ following (14) and (15);
update $\boldsymbol{\mathcal{Y}}$ following (16);
if stopping criterion is reached
break
end for if
end
Output $\hat{\boldsymbol{\mathcal{R}}}^{(k+1)}$.

Now the SSL problems have been reformulated to the low-rank tensor completion problems, and the outputs of this approach is CST. Thus, without the reverberation, utilizing the subspace-based MUSIC method could be a more straightforward way to conduct the SSL approach by the eigen-decompositions with relatively large number of microphones in our case. Finally, by utilising the completed CST $\hat{\boldsymbol{\mathcal{R}}}$ along the frequency dimension from f_0 to f_S, the broadband format of equation (9) can be used:

$$\beta_B(r_m) = \frac{1}{\sum_{f=f_0}^{f_S}\mathbf{w}(f,r_m)^H\mathbf{U}(f)_n\mathbf{U}(f)_n^H\mathbf{w}(f,r_m)}. \tag{20}$$

Note that the NSM broadband MUSIC output $\beta_B(r_m)$ can also be obtained by estimating the full CSM $\hat{\mathbf{R}}$ at every frequency bin, and the corresponding approach will be mentioned as broadband ADMM-based NSM below. In both the broadband ADMM-based NSM and TC-ADMM-based NSM approaches, all the frequency slices are used. And the frequency selection problem could be avoided. The two solutions are compared in the following section.

V. Simulations

As shown in Fig. 3, a 24-element rectangular array is considered. The number of measurements is $N = 5$. Three different speech signal sources are set in the sound field. The distance between adjacent microphones is 0.1 m, and the distance between the microphone array and the measurement plane is 0.5 m. Considering the common frequency range of a speech signal and the computing consumption, the operating frequency range is 1000– 2200 Hz with a 100 Hz step length. The simulation results are shown in Fig. 4.

Localisation results from different positions are shown in Fig. 4(1a) and (1b). The prototype array moved from the left to the right side along the x-axis. The actual locations of the three sources are marked as a circle, square, and diamond. Clearly, the sound sources at the farther side could not be located at the single-handed position. Owing to the limited size of the prototype array, it is difficult to identify a proper position with all three accurate locations of the sources.

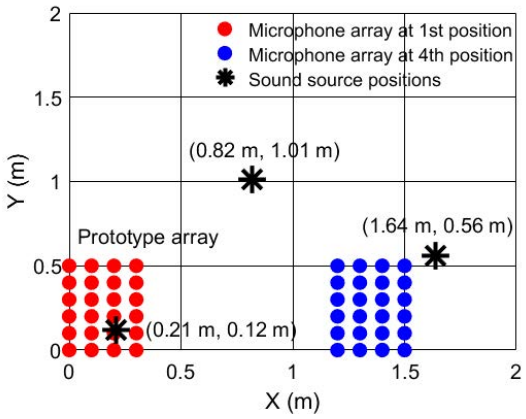

Fig. 3. Geometry of the prototype array at 1st and 4th position, and the location of the sound sources.

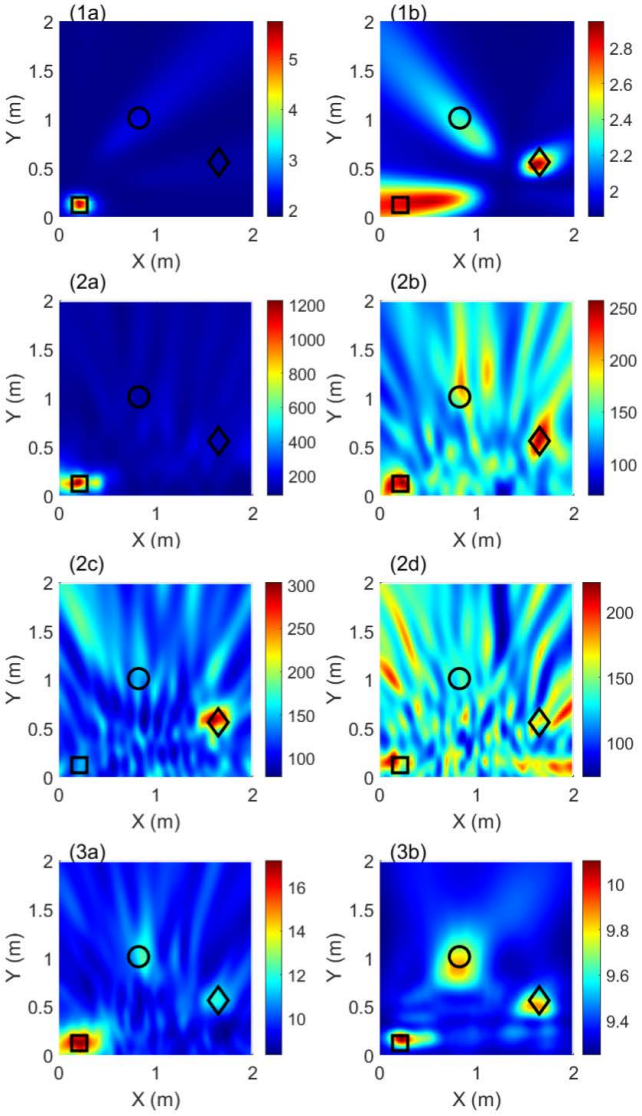

Fig. 4. Simulation results by broadband MUSIC at (1a) first position and (1b) fourth position of the prototype array, by ADMM-based NSM at (2a) 1200 Hz, (2b) 1500 Hz, (2c) 1800 Hz, and (2d) 2000 Hz, and by (3a) broadband ADMM-based NSM and (3b) TC-ADMM-based NSM.

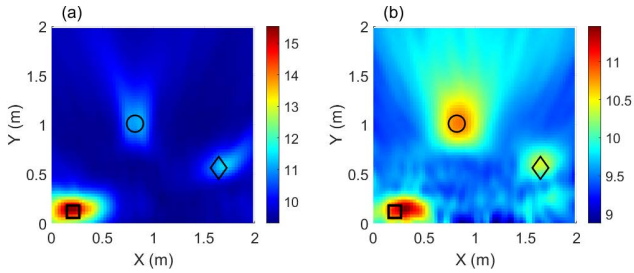

Fig. 5. Simulation results by TC-ADMM-based NSM: (a) $\gamma = 0.6$ and (b) $\gamma = 20$.

TABLE I
RMSE OF DIFFERENT LOCALISATION APPROACHES

Localisation approach	RMSE (m)
Broadband MUSIC at first position	0.341
Broadband MUSIC at fourth position	0.345
ADMM-based NSM at 1200 Hz	0.399
ADMM-based NSM at 1500 Hz	0.091
ADMM-based NSM at 1800 Hz	0.788
ADMM-based NSM at 2000 Hz	0.366
Broadband ADMM-based NSM	0.085
TC-ADMM-based NSM	0.067

The localisation results of the ADMM-based NSM approach in [28] at different frequencies are shown in Fig. 4(2a), (2b), (2c) and (2d). The sound sources are broadband speech signals; therefore, the selection of the operating frequency significantly affects the localisation results in this approach. At 1200 Hz, only the square source is indicated in Fig. 4(2a), whereas at 1800 Hz, only the diamond source appears in Fig. 4(2c). Fig. 4(2b) shows an indistinct sound map containing all three sources at 1500 Hz. Finally, an improper selection directly leads to an incorrect result at 2000 Hz, as shown in Fig. 4(2d).

Compared with the results above, the two NSM approaches virtually utilise all the $M \times N = 120$ microphones, and the three sound sources can be accurately located by these two broadband methods with a global view, as shown in Fig. 4(3a) and (3b). Furthermore, Fig. 4(3b) shows that the proposed method can achieve a more distinct image of the three sound sources. The simulation results illustrate that for a BM-SSL problem by NSM, completing the CST directly would be a better choice than completing the CSM at every frequency.

Additionally, we found that the relaxation parameter γ could actually influence the level of energy concentration of the pseudo-spectra for multiple sound sources in the sound field. For example, when $\gamma = 0.6$, two sound sources with longer distance from the array become indistinct in Fig. 5(a) due to the energy concentration. On the other hand, when $\gamma = 20$, more sidelobes are introduced in the sound map in Fig. 5(b). And $\gamma = 2.6$ could be a point of balance in this case.

The root mean square error (RMSE) is calculated by comparing the actual locations with the coordinates of three reasonable peak values in Table I, and the number of Monte Carlo trials is 50 in this letter. The RMSE of the proposed method is the smallest among all the approaches. As the

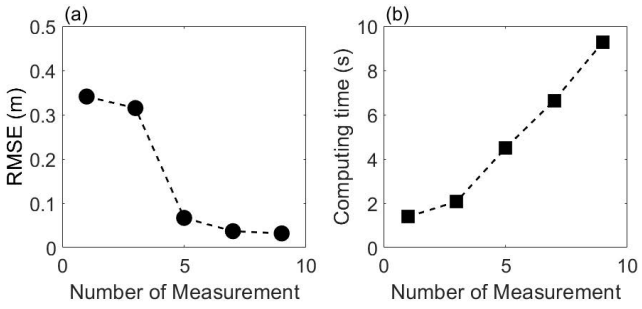

Fig. 6. Influence of the number of measurements on (a) RMSE and (b) computing time.

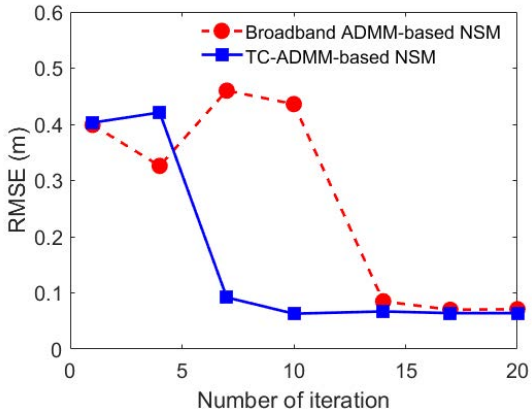

Fig. 7. RMSE versus the number of iteration.

number of measurements increases, the pseudo-spectra tends to provide more accurate localisation results in Fig. 6(a) when the total moving distance remains at 1.6 m. Fig. 6(b) shows the influence of the number of measurements on computing time with a Intel i7-6700 central processing unit (CPU) and 44 GB random access memory (RAM). Considering the time-consumption and computing burden required to obtain more measurements, we suggest an acceptable number of measurements in every specific case when the NSM is applied (i.e., $N = 5$ in this letter). At last, to justify the constringency of the proposed method in this paper by an empirical study at this stage, the RMSE versus the number of iteration in this case is shown in Fig. 7. It could be noticed that the RMSE of two ADMM-based methods converge well when the number of iteration is larger than 15.

To make the proposed method more statistically significant and convincible, the simulation was further extended by including five sound sources with the same settings as the former case, and the actual locations of the two newly added sources are marked as $+$ and $*$. As shown in Fig. 8(a), two additional speech sources could be found in the sound field. The localisation results by two NSM approaches are shown in Fig. 8(b) and (c). Compared with the results in Fig. 4(3a), the broadband ADMM-based NSM method performs worse in Fig. 8(b). The sidelobes introduced by five different speech sources superpose in the sound map. As a result, it is difficult to recognize all the five sound sources in

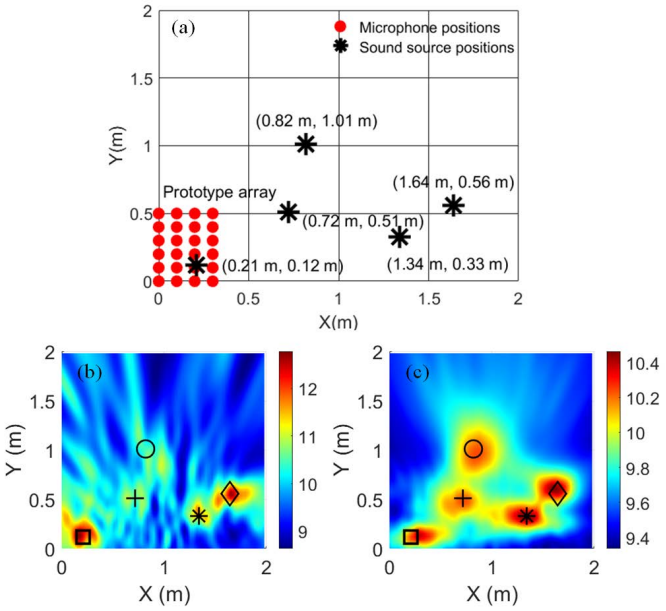

Fig. 8. Geometry of the prototype array and the locations of the five sources (a) with the SSL results by broadband ADMM-based NSM (b) and TC-ADMM-based NSM (c).

TABLE II
RMSE FROM FIVE SOURCES CASE

Localisation approach	RMSE (m)
Broadband MUSIC at first position	0.799
Broadband MUSIC at fourth position	0.732
ADMM-based NSM at 1200 Hz	0.793
ADMM-based NSM at 1500 Hz	0.720
ADMM-based NSM at 1800 Hz	0.950
ADMM-based NSM at 2000 Hz	0.472
Broadband ADMM-based NSM	0.461
TC-ADMM-based NSM	0.063

this figure. On the other hand, all the five sound sources could still be localised by the proposed TC-ADMM-based NSM in Fig. 8(c). Due to the short distances between the adjacent sources, in Fig. 8(c), the area of the mainlobes are bigger compared with Fig. 4(3b). Whlie the accuracy of the SSL are still promising. This conclusion are further confirmed by the RMSE results shown in Table II. Compared with the former case, the RMSE of the TC-ADMM-based NSM in this five sources case shows little change, while the RMSE of the ADMM-based NSM raises from 0.085 m to 0.461 m.

VI. FIELD DATA

To verify the BM-SSL performance of the proposed method, an experimental validation was carried out in the semi-anechoic chamber of Institute of Vibration, Shock, and Noise, State Key Laboratory, Shanghai Jiao Tong University (SJTU). The background noise level of the semi-anechoic chamber is 15.6 dB(A), and the cut-off frequency is 100 Hz. As it is shown in Fig. 9, a 56-channel spiral microphone array was placed 0.6 m in front of the measurement plane in the experiments. The prototype array was moved from left to right side along the x-axis to imitate the movement of

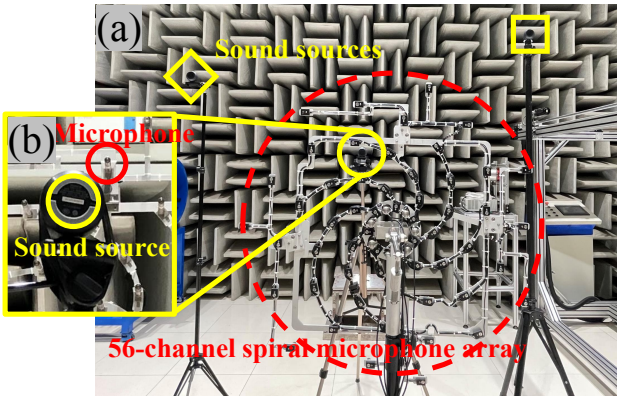

Fig. 9. (a) Three loudspeakers and the 56-channel spiral microphone array in a semi-anechoic chamber and (b) a zoomed in view from the back.

TABLE III
RMSE FROM EXPERIMENTAL STUDY

Localisation approach	RMSE (m)
Broadband MUSIC at first position	0.541
Broadband MUSIC at fourth position	0.567
Broadband MUSIC at seventh position	0.374
Broadband ADMM-based NSM	0.046
TC-ADMM-based NSM	0.035

a service robot during measurements. To imitate the human speakers, three Philips BT25 loudspeakers were arranged in the measurement plane as sound sources with the coordinates (-0.31 m, 0.95 m), (-0.92 m, 0.25 m) and (-1.71 m, 0.65 m), respectively. The speech signals were randomly selected from the Surfing-Tech Chinese Mandarin Corpus (ST-CMDS) data testing set, and the broadcasting acoustic signals were captured by a Mueller-BBM MKII sound measurement system with 56 $Brüel\&Kjær$ 4944A microphones, which were calibrated by a $BrüelKjær$ 4231 94 dB calibrator before the measurements. The sampling rate was set to be 16384 Hz, and the measurements duration was 1.875 seconds.

The direct localisation results from different measurement positions are shown in Fig. 10(a), (b), and (c) by broadband MUSIC. Obviously, the prototype array is too small to localize all the three sound sources. It is observed that the three sound sources can not be distinguished, and only the partial sound field can be visualized with a small single array. In order to have a global view of all the three sound sources, more microphones are required. As a contrast, the T-SVD based NSM approach by seven measurements in Fig. 10(d) could provide a distinct and accurate global view of the three speech sources, which is difficult to be achieved by the conventional measurements with limited microphones. It is concluded that the proposed approach could virtually utilize all the microphones with a wide frequency range since the the complexity and multidimensional structural information has been recovered by the TC algorithm.

At last, the RMSE from experimental study is calculated in Table III. Compared with the conventional measurements,

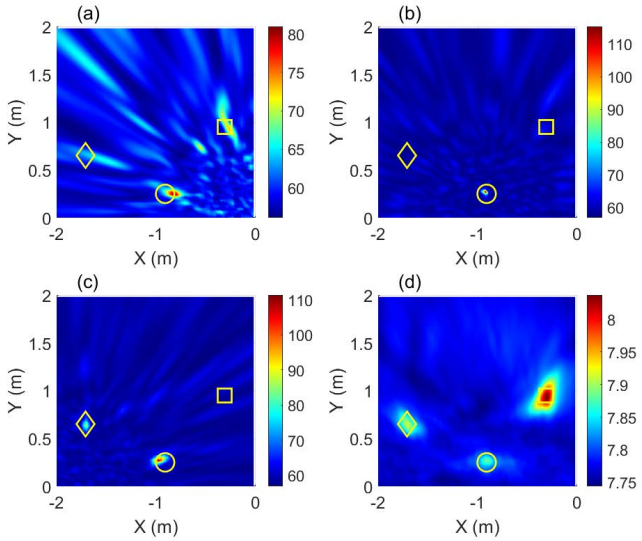

Fig. 10. Localisation results of the experimental study by broadband MUSIC at (a) first position, (b) fourth position, (c) seventh position, and (d) TC-ADMM-based NSM.

the RMSE of the proposed TC-ADMM-based NSM method is much smaller.

VII. CONCLUSION

A trade-off always exists between the aperture of a microphone array and the spatial separation of the microphones in microphone array design for service robots. The NSM method provides a potential solution to this problem. In this study, we extended the NSM method and developed a conceivable approach for BM-SSL with simulations and experiments. The basic ideas are to use the tensor structure of the observed broadband signals and to employ a TC-based ADMM approach, instead of matrix-based one to perform the tensor data recovery. By this mean, a full CST could be obtained and utilized for BM-SSL. The problem of selecting the operating frequency for the BM-SSL by NSM was suitably solved. The corresponding SSL approach could provide a distinct global view of three different speech signal sources with high accuracy. Compared with the state-of-the-art algorithms, the proposed method can provide higher locating accuracy. With the movable microphone array, it could be a potential option for designing the service robots SLL system to image a large-scale broadband sound radiation.

REFERENCES

[1] R. A. Brooks, "Elephants don't play chess," *Robotics and autonomous systems*, vol. 6, no. 1-2, pp. 3–15, 1990.

[2] R. E. Irie, "Robust sound localization: An application of an auditory perception system for a humanoid robot," tech. rep., MASSACHUSETTS INST OF TECH CAMBRIDGE DEPT OF ELECTRICAL ENGINEERING AND , 1995.

[3] J. Huang, T. Supaongprapa, I. Terakura, N. Ohnishi, and N. Sugie, "Mobile robot and sound localization," in *Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robot and Systems. Innovative Robotics for Real-World Applications. IROS'97*, vol. 2, pp. 683–689, IEEE, 1997.

[4] C. Rascon and I. Meza, "Localization of sound sources in robotics: A review," *Rob Auton Syst*, vol. 96, pp. 184–210, 2017.

[5] D. S. Talagala, W. Zhang, T. D. Abhayapala, and A. Kamineni, "Binaural sound source localization using the frequency diversity of the head-related transfer function," *The Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 1207–1217, 2014.

[6] J. Wang, J. Wang, K. Qian, X. Xie, and J. Kuang, "Binaural sound localization based on deep neural network and affinity propagation clustering in mismatched hrtf condition," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2020, no. 1, pp. 1–16, 2020.

[7] V. Tourbabin and B. Rafaely, "Speaker localization by humanoid robots in reverberant environments," in *2014 IEEE 28th Convention of Electrical & Electronics Engineers in Israel (IEEEI)*, pp. 1–5, IEEE, 2014.

[8] D. Salvati, C. Drioli, and G. L. Foresti, "Power method for robust diagonal unloading localization beamforming," *IEEE Signal Processing Letters*, vol. 26, no. 5, pp. 725–729, 2019.

[9] L. Huang, S. Wu, and X. Li, "Reduced-rank mdl method for source enumeration in high-resolution array processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5658–5667, 2007.

[10] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wideband sources," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 823–831, 1985.

[11] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. II–1949, IEEE, 2002.

[12] T. Padois and A. Berry, "Application of acoustic imaging techniques on snowmobile pass-by noise," *J Acoust Soc Am*, vol. 141, no. 2, pp. EL134–EL139, 2017.

[13] L. Chen, Y. S. Choy, T. G. Wang, and Y. K. Chiang, "Fault detection of wheel in wheel/rail system using kurtosis beamforming method," *Struct Health Monit*, vol. 19, no. 2, pp. 495–509, 2020.

[14] S. Argentieri and P. Danes, "Broadband variations of the music high-resolution method for sound source localization in robotics," in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2009–2014, IEEE, 2007.

[15] P. Danes and J. Bonnal, "Information-theoretic detection of broadband sources in a coherent beamspace music scheme," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1976–1981, IEEE, 2010.

[16] K. Nakamura, K. Nakadai, and G. Ince, "Real-time super-resolution sound source localization for robots," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 694–699, IEEE, 2012.

[17] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[18] M. Cobos, M. Garcła-Pineda, and M. Arevalillo-Herrez, "Steered response power localization of acoustic passband signals," *IEEE Signal Processing Letters*, vol. 24, no. 5, pp. 717–721, 2017.

[19] H. Sundar, T. V. Sreenivas, and C. S. Seelamantula, "Tdoa-based multiple acoustic source localization without association ambiguity," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 11, pp. 1976–1990, 2018.

[20] M. Strauss, P. Mordel, V. Miguet, and A. Deleforge, "Dregon: Dataset and methods for uav-embedded sound source localization," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1–8, 2018.

[21] W. Manamperi, T. D. Abhayapala, J. Zhang, and P. N. Samarasinghe, "Drone audition: Sound source localization using on-board microphones," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 508–519, 2022.

[22] K.-C. Kwak and S.-S. Kim, "Sound source localization with the aid of excitation source information in home robot environments," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 852–856, 2008.

[23] P. Chiariotti, M. Martarelli, and P. Castellini, "Acoustic beamforming for noise source localization–reviews, methodology and applications," *Mech Syst Signal Process*, vol. 120, pp. 422–448, 2019.

[24] R. A. Haubrich, "Array design," *Bulletin of the Seismological Society of America*, vol. 58, no. 3, pp. 977–991, 1968.

[25] Z. Prime and C. Doolan, "A comparison of popular beamforming arrays," *Proceedings of the Australian Acoustical Society AAS2013 Victor Harbor*, vol. 1, p. 5, 2013.

[26] E. Sarradj, "A generic approach to synthesize optimal array microphone arrangements," in *Proceedings of the 6th Berlin Beamforming Conference*, p. 4, 2016.

[27] J. Antoni, "Synthetic aperture acoustical holography," in *International Conference on Noise and Vibration Engineering and International, Conference on Uncertainty in Structural Dynamics, ISMA2012, Leuven, Belgium*, 2012.

[28] L. Yu, J. Antoni, H. Wu, Q. Leclere, and W. Jiang, "Fast iteration algorithms for implementing the acoustic beamforming of non-synchronous measurements," *Mech Syst Signal Process*, vol. 134, p. 106309, 2019.

[29] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans Inf Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.

[30] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.

[31] N. Chu, Y. Ning, L. Yu, Q. Huang, and D. Wu, "A fast and robust localization method for low-frequency acoustic source: Variational bayesian inference based on nonsynchronous array measurements," *IEEE Trans Instrum Meas*, vol. 70, pp. 1–18, 2020.

[32] F. Ning, J. Song, J. Hu, and J. Wei, "Sound source localization of non-synchronous measurements beamforming with block hermitian matrix completion," *Mech Syst Signal Process*, vol. 147, p. 107118, 2021.

[33] N. Chu, Q. Liu, L. Yu, Y. Ning, and P. Hou, "Non-synchronous measurements of a microphone array at coprime positions," *IEEE Signal Processing Letters*, vol. 28, pp. 1420–1424, 2021.

[34] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Appl*, vol. 435, no. 3, pp. 641–658, 2011.

[35] W. Z. Sun, P. Zhang, and B. Zhao, "Rank revealing-based tensor completion using improved generalized tensor multi-rank minimization," *IET Signal Processing*, vol. 15, no. 8, pp. 483–499, 2021.

[36] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 4, pp. 925–938, 2019.

[37] Z. Zhang and S. Aeron, "Exact tensor completion using t-svd," *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1511–1526, 2016.

[38] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans Antennas Propag*, vol. 34, no. 3, pp. 276–280, 1986.

[39] L. Huang, T. Long, and S. Wu, "Source enumeration for high-resolution array processing using improved gerschgorin radii without eigendecomposition," *IEEE Transactions on Signal Processing*, vol. 56, no. 12, pp. 5916–5925, 2008.

[40] L. Huang and H.-C. So, "Source enumeration via mdl criterion based on linear shrinkage estimation of noise subspace covariance matrix," *IEEE Transactions on Signal Processing*, vol. 61, no. 19, pp. 4806–4821, 2013.

[41] D. Hu, J. Ding, H. Zhao, and L. Yu, "Spatial basis interpretation for implementing the acoustic imaging of non-synchronous measurements," *Applied Acoustics*, vol. 182, p. 108198, 2021.

[42] L. Yu, J. Antoni, and Q. Leclere, "Spectral matrix completion by cyclic projection and application to sound source reconstruction from non-synchronous measurements," *Journal of Sound and Vibration*, vol. 372, pp. 31–49, 2016.

[43] L. Yu, *Acoustical source reconstruction from non-synchronous sequential measurements*. PhD thesis, INSA de Lyon, 2015.