

Article

An Efficient Framework for Estimating the Direction of Multiple Sound Sources Using Higher-Order Generalized Singular Value Decomposition

Bandhit Suksiri ¹  and Masahiro Fukumoto ^{2,*}

¹ Department of Engineering, Graduate School of Engineering, Kochi University of Technology, Kami Campus, Kochi 782-0003, Japan

² School of Information, Kochi University of Technology, Kami Campus, Kochi 782-0003, Japan

* Correspondence: fukumoto.masahiro@kochi-tech.ac.jp; Tel.: +81-887-53-1130

Received: 28 May 2019; Accepted: 3 July 2019; Published: 5 July 2019



Abstract: This paper presents an efficient framework for estimating the direction-of-arrival (DOA) of wideband sound sources. The proposed framework provides an efficient way to construct a wideband cross-correlation matrix from multiple narrowband cross-correlation matrices for all frequency bins. In addition, the proposed framework is inspired by the coherent signal subspace technique with further improvement of linear transformation procedure, and the new procedure no longer requires any process of DOA preliminary estimation by exploiting unique cross-correlation matrices between the received signal and itself on distinct frequencies, along with the higher-order generalized singular value decomposition of the array of this unique matrix. Wideband DOAs are estimated by employing any subspace-based technique for estimating narrowband DOAs, but using the proposed wideband correlation instead of the narrowband correlation matrix. It implies that the proposed framework enables cutting-edge studies in the recent narrowband subspace methods to estimate DOAs of the wideband sources directly, which result in reducing computational complexity and facilitating the estimation algorithm. Practical examples are presented to showcase its applicability and effectiveness, and the results show that the performance of fusion methods perform better than others over a range of signal-to-noise ratios with just a few sensors, which make it suitable for practical use.

Keywords: direction-of-arrival (DOA); higher-order generalized singular value decomposition (HOGSVD); wideband sources; sound source; array processing; subspace method; cross-correlation

1. Introduction

The fundamental competence of sound source localization has received much attention during the past decades, and has become an important part of navigation systems [1,2]. Direction-of-arrival (DOA) estimation in particular plays a critical role in navigation systems for the exploration of sources in widespread applications, including in acoustic signal processing [3–8]. Several approaches have been proposed as a potential way to estimate DOA. For instance, the time-difference-of-arrival-based DOA estimation is one of the most frequently used approaches, which is widely known as the generalized cross-correlation with phase transform (GCC-PHAT) [9]. In addition to this approach, a low computational requirement makes it attractive for practical applications; however, the major drawback is its low robustness in noisy and multipath environments. Another relevant approach is adopted from the independent component analysis (ICA) in blind source separation [10,11]. ICA searches independent components by measuring deviations from Gaussian distributions, such as maximization of negentropy or kurtosis. DOAs are estimated easily by using the separated components for all frequency bins, but it should be noted that the estimation accuracy of such a method is highly sensitive to the non-Gaussianity measures.

In an alternative approach to estimate narrowband DOAs, the subspace method has been proposed in an effort to improve estimation performance. The most prominent methods observe the signal and noise subspace for achieving more robust results, such as multiple signal classification (MUSIC) [12], estimation of signal parameters via rotational invariance techniques (ESPRIT) [13], and propagator method [14,15], which have been used frequently for one-dimensional (1D) DOA estimation along with the uniform linear array (ULA) of sensors. In case of a two-dimensional (2D) DOA estimation, a new geometrical structure of a sensor array is required, and it was previously found that the structure of an L-shaped array is considerably effective for estimating 2D DOAs [16]. Additionally, the L-shaped array allows for simple implementation, because it consists of two ULAs connected orthogonally at one end of each ULA. For these reasons, the L-shaped array is widely applied to the 2D DOA estimation method [17–26], and its practical applications can be found in the past researches [27,28]. Although the narrowband subspace method may be unable to directly estimate wideband DOAs, one possible way to solve this problem is to employ the narrowband subspace method in each temporal frequency intensively, and then the wideband DOA results can be estimated by interpolating the narrowband DOA results all frequency bins [29,30]. It should be noted again that intensive computational costs encountered in the above solution may be limited by practical considerations.

Several approaches were proposed to solve the problem of estimating wideband DOAs, for example, the incoherent MUSIC (IMUSIC) is one of the simplest methods for estimating wideband DOA [31]. There are two steps in IMUSIC: Firstly, a noise subspace model each temporal frequency is constructed. Then, wideband DOAs are obtained by minimizing the norm of orthogonal relation between a steering vector and the noise subspace of all frequency bins. Although accuracy performance of IMUSIC was demonstrated to be an effective method for estimating DOAs of multiple wideband signals in the high signal-to-noise ratio (SNR) region, a single small distortion of the noise subspace at any frequency can affect the whole DOA results. Many attempts were made recently to overcome this problem. For instance, the test of orthogonality of frequency subspaces (TOFS) was proposed to overcome this difficulty [32], but performance degradation caused by the small distortion still remains challenging. Another relevant approach is called the test of orthogonality of projected subspaces (TOPS) [33]. TOPS estimate DOA by constructing signal subspace of one reference frequency, and then measuring orthogonality of the previous signal subspace and noise subspace for all frequency bins. The simulations showed that TOPS is able to achieve higher accuracy than IMUSIC in mid SNR range, however, the undesirable false peaks still remain. The revised and greatly improved version of TOPSs were proposed recently to reduce these false peaks [34,35]. Obviously, computational complexities increased dramatically compared to the classical TOPS.

Another notable approach of wideband DOA estimation is the coherent signal subspace method (CSS) [36,37]. CSS specifically focuses a correlation matrix of received signals of each temporal frequency into a single matrix, which is called a universal correlation, associated with one focusing frequency via linear transformation procedure. Wideband DOAs are estimated by applying a single scheme of any narrowband subspace method on the universal correlation matrix. In addition to the transformation procedure of CSS [38–40], a process of DOA preliminary estimation is required before the wideband DOAs can be estimated. Therefore, a common shortcoming is clearly recognized as a requirement of DOA preliminary estimation, which means that any inferior initiation can lead to biased estimates. According to the literature [31–33,41], CSS demonstrates deficient performance than others such as TOPS; this is because the solutions of transformation procedure in CSS are solely focused on subspace between a temporal frequency and focusing frequency; to the best knowledge of the authors, it means that a fundamental component of the transformation matrix across all frequency bins may exhibit the different core component, which is clearly apparent when a narrowband DOA result at some frequency is not close enough to the true DOA. A single component distortion can definitely affect the whole DOA results. Therefore, the solutions have to exhibit the exact component even though power present in a received signal at that frequency is very weak; in other words, the solution

of transformation matrix have to be focused across all frequency bins instead of the pair of different frequencies.

Therefore, the purpose of this paper is to investigate an alternative for estimating wideband 2D DOAs in a more efficient way. We consider wideband sources as sound sources, such as human speeches and musical sounds. In order to estimate the wideband DOAs, we address the issue of transforming multiple narrowband cross-correlation matrices for all frequency bins into a wideband cross-correlation matrix. Additionally, our study is inspired by a computational model of CSS with further improvement of a linear transformation procedure [36–40]. Since the transformation procedures of CSS are only focused on subspace between current and reference frequency as previously mentioned, we propose a new transformation procedure which focus all frequency bins simultaneously and efficiently. The higher-order generalized singular value decomposition (HOGSVD) is firstly used to achieve this important issue [42]. By employing HOGSVD of arrays of the new unique cross-correlation matrix, where elements in the row and column positions are a sample cross-correlation matrix between received signal and itself on two distinct frequencies, the new transformation procedure no longer require any process of DOA preliminary estimation. Finally, the wideband cross-correlation matrix is constructed via the proposed transformation procedure, and the wideband DOAs can be estimated by employing any subspace-based technique for estimating narrowband DOAs, but using this wideband correlation matrix instead of the narrowband correlation matrix. Therefore, the proposed framework enables cutting-edge studies in the recent narrowband subspace methods to estimate DOA of the wideband sources directly, which result in reducing computational complexity and facilitating the estimation algorithm. Practical examples, such as 2D-MUSIC and ESPRIT with an L-shaped array, are presented to showcase its applicability and effectiveness.

The rest of this paper is organized as follows. Section 2 presents the array signal model, basic assumptions and problem formulation for transforming narrowband sample cross-correlation matrices for all frequency bins into a single matrix, which is called wideband cross-correlation matrix. Description of the new transformation procedure is introduced in Section 3.1 and its effective solution via HOGSVD in Section 3.2. Section 3.3 provide a description of the proposed framework for estimating wideband DOAs by combining the proposed transformation procedure along with a scheme of estimating DOAs in a recent narrowband subspace method, and its practical examples are presented in Sections 3.3.1 and 3.3.2. The simulation and experimental results are compared with the several existing methods in Sections 4 and 5. Finally, Section 6 concludes this paper.

2. Preliminaries

2.1. Data Model

The proposed method presented in this paper considers far-field sound sources. Received signals are a composition of the multiple sources, each one consisting of an angle in a spherical coordinate system. The received signals are transformed into a time-frequency representation via the short-time Fourier transform (STFT), and are given by

$$\mathbf{r}(t, f) = \mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f) \mathbf{s}(t, f) + \mathbf{w}(t, f), \quad (1)$$

where $\mathbf{r}(t, f) \in \mathbb{C}^M$ is the summation of a received signal, $\mathbf{s}(t, f) \in \mathbb{C}^K$ is a source signal, $\mathbf{w}(t, f) \in \mathbb{C}^M$ is an additive noise, the constant M is the number of microphone elements, and K is the number of incident sources. The matrix $\mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f) \in \mathbb{C}^{M \times K}$ stands for the array manifold where $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are phase angle components of the source on x and z axes in the spherical coordinate system. Note that the elements in $\mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f)$ depend on an array geometry.

Consider the L-shaped array structure consisting of two ULAs as illustrated in Figure 1, the received signals are simplified as

$$\begin{bmatrix} \mathbf{x}(t, f) \\ \mathbf{z}(t, f) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_x(\boldsymbol{\phi}, f) \\ \mathbf{A}_z(\boldsymbol{\theta}, f) \end{bmatrix} \mathbf{s}(t, f) + \begin{bmatrix} \mathbf{w}_x(t, f) \\ \mathbf{w}_z(t, f) \end{bmatrix}, \quad (2)$$

where

$$\begin{aligned} \mathbf{A}_x(\boldsymbol{\phi}, f) &= [\mathbf{a}_x(\phi_1, f) \quad \mathbf{a}_x(\phi_2, f) \quad \dots \quad \mathbf{a}_x(\phi_K, f)], \\ \mathbf{A}_z(\boldsymbol{\theta}, f) &= [\mathbf{a}_z(\theta_1, f) \quad \mathbf{a}_z(\theta_2, f) \quad \dots \quad \mathbf{a}_z(\theta_K, f)], \\ \mathbf{a}_x(\phi_k, f) &= [e^{\alpha_x(\phi_k, f)j} \quad e^{2\alpha_x(\phi_k, f)j} \quad \dots \quad e^{N\alpha_x(\phi_k, f)j}]^T, \\ \mathbf{a}_z(\theta_k, f) &= [1 \quad e^{\alpha_z(\theta_k, f)j} \quad \dots \quad e^{(N-1)\alpha_z(\theta_k, f)j}]^T, \\ \alpha_x(\phi_k, f) &= \left(\frac{f}{f_0}\right) \cdot \left(\frac{2\pi d \cos \phi_k}{\lambda}\right), \\ \alpha_z(\theta_k, f) &= \left(\frac{f}{f_0}\right) \cdot \left(\frac{2\pi d \cos \theta_k}{\lambda}\right). \end{aligned} \quad (3)$$

From the above definitions, $\mathbf{x}(t, f), \mathbf{w}_x(t, f) \in \mathbb{C}^N, \mathbf{A}_x(\boldsymbol{\phi}, f) \in \mathbb{C}^{N \times K}$ and a subscript x are belonged to x subarray, and likewise, $\mathbf{z}(t, f), \mathbf{w}_z(t, f) \in \mathbb{C}^N, \mathbf{A}_z(\boldsymbol{\theta}, f) \in \mathbb{C}^{N \times K}$ and a subscript z are belonged to z subarray where N is the number of microphone elements each subarray with $M = 2N$. The variable t is time, f is a source frequency, d is the spacing of microphone elements, λ is a wavelength with respect to $\lambda = \frac{c}{f_0}$ where c is the speed of sound in current medium, and f_0 is a reference frequency.

2.2. Basic Assumptions

Based on the recent reviews, the following assumptions are required on the proposed framework:

Assumption 1: The number of sources is known or predicted in advance [43,44].

Assumption 2: The spacing between adjacent elements of each subarray and spacing between x_1 and z_1 should be set to $d = \frac{\lambda}{2}$ for avoiding the angle ambiguity in array structure radiation [1,2,16].

Assumption 3: The source $\mathbf{s}(t, f)$ is assumed to be Gaussian complex random variable as suggested by the literature [12,16,31]. However, we consider wideband sources as sound sources such as human speech; therefore, $\mathbf{s}(t, f)$ can also be Super-Gaussian complex random variable, and it is not stationary signals for the most general case when giving an appropriate period of time.

Assumption 4: According to acoustic theory of speech, frequency dependence of the sound source, especially a human speech, is existed [45]; it means that a cross-covariance between the source and itself with distinct frequencies is not zero; $\text{cov}(s_k(\mathbf{t}, f), s_k(\mathbf{t}, f')) = c_{s_k\{f, f'\}}$, where $c_{s_k\{f, f'\}} \in \mathbb{C}$. Next, suppose that $\mathbf{s}(\mathbf{t}, f)$ are uncorrelated, which implies that $s_k(\mathbf{t}, f)$ and $s_{k'}(\mathbf{t}, f')$ are statistically independent of each other when $k \neq k'$; $\text{cov}(s_k(\mathbf{t}, f), s_{k'}(\mathbf{t}, f')) = 0$. When $k = k'$, the sources can take to be partially dependent by the following literature [45]; therefore, a sample cross-covariance matrix of the incident sources over two different frequencies is given by

$$\begin{aligned} \mathbf{S}_{\{f, f'\}} &= E \left\{ \mathbf{s}(\mathbf{t}, f) \mathbf{s}^H(\mathbf{t}, f') \right\} \\ &= \text{diag} \left(c_{s_1\{f, f'\}}, c_{s_2\{f, f'\}}, \dots, c_{s_K\{f, f'\}} \right). \end{aligned} \quad (4)$$

Remark that $c_{s_k\{f, f'\}}$ is equal to $\sigma_{s_k\{f\}}^2$, and $\sigma_{s_k\{f\}}^2 \in \mathbb{R}_{\geq 0}$ is a variance at frequency f of the source.

Assumption 5: An additive white Gaussian noise is considered in this paper, which is modeled as Gaussian random variable as well as the past studies. A noise cross-covariance matrix over two different frequencies is given by

$$\begin{aligned} \mathbf{W}_{\{f,f'\}} &= E \left\{ \mathbf{w}(t, f) \mathbf{w}^H(t, f') \right\} \\ &= c_{w\{f,f'\}} \mathbf{I}_M, \end{aligned} \quad (5)$$

where $c_{w\{f,f'\}} \in \mathbb{C}$, and \mathbf{I}_i is a i -by- i identity matrix. Note again that $c_{w\{f,f\}} = \sigma_{w\{f\}}^2$ where $\sigma_{w\{f\}}^2 \in \mathbb{R}_{\geq 0}$ is a variance of the noise at frequency f . In case of the L-shaped array structure in Equation (2), we have

$$\begin{bmatrix} \mathbf{W}_{xx\{f,f'\}} & \mathbf{W}_{xz\{f,f'\}} \\ \mathbf{W}_{zx\{f,f'\}} & \mathbf{W}_{zz\{f,f'\}} \end{bmatrix} = \begin{bmatrix} c_{w\{f,f'\}} \mathbf{I}_N & \mathbf{O}_{N \times N} \\ \mathbf{O}_{N \times N} & c_{w\{f,f'\}} \mathbf{I}_N \end{bmatrix}, \quad (6)$$

where $\mathbf{O}_{i \times j}$ is a i -by- j null matrix.

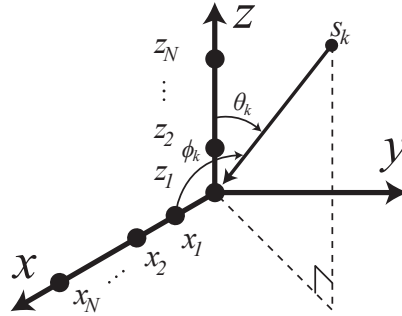


Figure 1. L-shaped microphone array configuration for 2D DOA estimation.

2.3. Transformation Problem

Under the data model and assumptions in Sections 2.1 and 2.2, a cross-correlation matrix of the received signals is defined as

$$\begin{aligned} \mathbf{R}_{\{f,f'\}} &= E \left\{ \mathbf{r}(t, f) \mathbf{r}^H(t, f') \right\} \\ &= \mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f) \mathbf{S}_{\{f,f'\}} \mathbf{A}^H(\boldsymbol{\phi}, \boldsymbol{\theta}, f') + \mathbf{W}_{\{f,f'\}}, \end{aligned} \quad (7)$$

where $\mathbf{R}_{\{f,f'\}} \in \mathbb{C}^{M \times M}$. In order to transform $\mathbf{R}_{\{f,f'\}}$ over the available frequency range into a single smoothed matrix, which is named as a wideband cross-correlation, a transformation procedure is required as mentioned previously [36], which is expressed as

$$\begin{aligned} \mathbf{R} &= \frac{1}{P} \sum_{i=1}^P \mathbf{T}_{\{f_i\}} \mathbf{R}_{\{f_i,f_i\}} \mathbf{T}_{\{f_i\}}^H \\ &= \mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f_0) \left(\frac{1}{P} \sum_{i=1}^P \mathbf{S}_{\{f_i,f_i\}} \right) \mathbf{A}^H(\boldsymbol{\phi}, \boldsymbol{\theta}, f_0) + \frac{1}{P} \sum_{i=1}^P \mathbf{T}_{\{f_i\}} \mathbf{W}_{\{f_i,f_i\}} \mathbf{T}_{\{f_i\}}^H, \end{aligned} \quad (8)$$

where

$$\mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f_0) = \mathbf{T}_{\{f_i\}} \mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f_i), \quad (9)$$

$\mathbf{R} \in \mathbb{C}^{M \times M}$ is the wideband cross-correlation matrix, and P is the number of STFT frequency bins. $\mathbf{T}_{\{f_i\}} \in \mathbb{C}^{M \times M}$ is a transformation matrix, which was originally designed by using the ordinary beamforming technique [36], or by minimizing the Frobenius norm of array manifold matrices [37]. The objective of $\mathbf{T}_{\{f\}}$ is to transform any given f of the array manifold $\mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f)$ into $\mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f_0)$. All previous solutions of $\mathbf{T}_{\{f\}}$ are solely based on subspace between pair of distinct frequencies $\{f, f_0\}$,

as emphasized in the introduction [36–40]. When power of the source at some frequency is weak or less than noise power, the matrix $\mathbf{T}_{\{f\}}$ may not share any common angle of $\boldsymbol{\phi}, \boldsymbol{\theta}$ because its non-zero eigenvalues are not full rank, which is resulted in a performance degradation for estimating both $\mathbf{T}_{\{f\}}$ and wideband DOAs. If the transformation matrix can be focused by all frequency bins instead of the pair of frequencies, a good estimate of DOAs in Equation (8) might be expected. Based on this hypothesis, a new concept and scheme are presented in next section.

3. Proposed Method

This section introduces a new procedure for estimating a transformation matrix, its alternative solution by using the higher-order generalized singular value decomposition (HOGSVD), and practical examples of wideband DOA estimation scheme.

3.1. Problem for Estimating the Transformation Matrix and Its Solution

We start by introducing the following lemma that will be useful for obtaining a solution of transformation matrix.

Lemma 1. Given a set of two distinct frequencies by $\{f, f_o\}$ into Equation (7), and given a transformation matrix $\mathbf{T}_{\{f\}}$ which satisfy the property in Equation (9), assume that $K < M$, the cross-correlation $\mathbf{R}_{\{f, f_o\}}$ can be factorized into the singular value decomposition (SVD) form;

$$\mathbf{R}_{\{f, f_o\}} = \mathbf{U}_{\{f, f_o\}_s} \boldsymbol{\Sigma}_{\{f, f_o\}_s} \mathbf{V}_{\{f, f_o\}_s}^H + \mathbf{U}_{\{f, f_o\}_n} \boldsymbol{\Sigma}_{\{f, f_o\}_n} \mathbf{V}_{\{f, f_o\}_n}^H \quad (10)$$

where $\mathbf{U}_{\{f, f_o\}_s}, \mathbf{V}_{\{f, f_o\}_s} \in \mathbb{C}^{M \times K}$, $\boldsymbol{\Sigma}_{\{f, f_o\}_s} \in \mathbb{R}^{K \times K}$ are the matrix of left and right singular vectors and diagonal matrix of singular values in signal subspace, and likewise, $\mathbf{U}_{\{f, f_o\}_n}, \mathbf{V}_{\{f, f_o\}_n} \in \mathbb{C}^{M \times M-K}$, $\boldsymbol{\Sigma}_{\{f, f_o\}_n} \in \mathbb{C}^{M-K \times M-K}$ are with noise subspace. If the K largest singular values of $\mathbf{T}_{\{f\}} \mathbf{R}_{\{f, f_o\}}$ and $\mathbf{R}_{\{f, f_o\}}$ are equal, then $\mathbf{T}_{\{f\}} \mathbf{U}_{\{f, f_o\}_s}$ is a matrix with orthonormal columns.

Proof. Since the transformation procedure of $\mathbf{R}_{\{f, f_o\}}$ is expressed by $\mathbf{T}_{\{f\}} \mathbf{R}_{\{f, f_o\}}$ and the array manifold $\mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f)$ and $\mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f_o)$ are full rank matrices [36], Lemma 1 is valid if and only if the K largest singular values of $\mathbf{T}_{\{f\}} \mathbf{R}_{\{f, f_o\}}$ and $\mathbf{R}_{\{f, f_o\}}$ are equal; therefore, $\mathbf{U}_{\{f, f_o\}_s}^H \mathbf{T}_{\{f\}}^H \mathbf{T}_{\{f\}} \mathbf{U}_{\{f, f_o\}_s} = \mathbf{I}_K$. Considering the $M - K$ smallest singular values of $\mathbf{R}_{\{f, f_o\}}$ are close to zeros by assuming a noise-free signal and using solely the signal subspace $\mathbf{U}_{\{f, f_o\}_s} \boldsymbol{\Sigma}_{\{f, f_o\}_s} \mathbf{V}_{\{f, f_o\}_s}^H$, we have

$$\begin{aligned} & \left(\mathbf{T}_{\{f\}} \left(\mathbf{R}_{\{f, f_o\}} - \mathbf{W}_{\{f, f_o\}} \right) \right)^H \left(\mathbf{T}_{\{f\}} \left(\mathbf{R}_{\{f, f_o\}} - \mathbf{W}_{\{f, f_o\}} \right) \right) \\ & = \mathbf{V}_{\{f, f_o\}_s} \boldsymbol{\Sigma}_{\{f, f_o\}_s} \boldsymbol{\Sigma}_{\{f, f_o\}_s}^H \mathbf{V}_{\{f, f_o\}_s}^H \end{aligned} \quad (11)$$

Performing the Eigenvalue decomposition (EVD) to Equation (11), square roots of the non-zero eigenvalues of above matrix is equal to $\boldsymbol{\Sigma}_{\{f, f_o\}_s}$ [46,47]. This completes the proof of the lemma. \square

Lemma 1 shows that $\mathbf{R}_{\{f, f_o\}}$ and $\mathbf{T}_{\{f\}} \mathbf{R}_{\{f, f_o\}}$ share the common components on the singular values and right singular vectors, whereas the both left singular vectors may be different. Since $\mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f)$ and $\mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f_o)$ are full rank, its remaining components are given by [48]:

$$\begin{aligned} \mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f) &= \mathbf{U}_{\{f, f_o\}_s} \mathbf{F}_{\{f, f_o\}_s}, \\ \mathbf{T}_{\{f\}} \mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f) &= \left(\mathbf{T}_{\{f\}} \mathbf{U}_{\{f, f_o\}_s} \right) \mathbf{F}_{\{f, f_o\}_s}, \\ \mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f_o) &= \mathbf{V}_{\{f, f_o\}_s} \mathbf{G}_{\{f, f_o\}_s}, \end{aligned} \quad (12)$$

where

$$\boldsymbol{\Sigma}_{\{f, f_o\}_s} = \mathbf{F}_{\{f, f_o\}_s} \mathbf{S}_{\{f, f_o\}_s} \mathbf{G}_{\{f, f_o\}_s}^H \quad (13)$$

$\mathbf{F}_{\{f,f_0\}}, \mathbf{G}_{\{f,f_0\}} \in \mathbb{C}^{K \times K}$ are full rank matrices and have invertibility. From Equations (9) and (12), we have

$$\mathbf{T}_{\{f\}} \mathbf{U}_{\{f,f_0\}_s} = \mathbf{V}_{\{f,f_0\}_s} \mathbf{G}_{\{f,f_0\}} \mathbf{F}_{\{f,f_0\}}^{-1}, \quad (14)$$

which mean that the right singular vectors of $\mathbf{R}_{\{f,f_0\}}$ and the left singular vectors of $\mathbf{T}_{\{f\}} \mathbf{R}_{\{f,f_0\}}$ share the common subspace when $\mathbf{G}_{\{f,f_0\}} \mathbf{F}_{\{f,f_0\}}^{-1}$ has unitary property.

Since the left singular vectors of $\mathbf{T}_{\{f\}} \mathbf{R}_{\{f,f_0\}}$ exist, we continue to introduce a new transformation procedure. The matrix $\mathbf{T}_{\{f\}}$ can be found as a solution to

$$\begin{aligned} & \underset{\mathbf{T}_{\{f\}}}{\text{minimize}} \quad \left\| \mathbf{R}_{\{f_0,f_0\}} - \mathbf{T}_{\{f\}} \mathbf{R}_{\{f,f_0\}} \right\|_{\text{F}}^2 \\ & \text{subject to} \quad \sum_{k=1}^K \sigma_k^2 \left(\mathbf{T}_{\{f\}} \mathbf{R}_{\{f,f_0\}} \right) = \sum_{k=1}^K \sigma_k^2 \left(\mathbf{R}_{\{f,f_0\}} \right), \end{aligned} \quad (15)$$

where $\|\cdot\|_{\text{F}}$ is the Frobenius norm, and $\sum_{k=1}^K \sigma_k^2(\mathbf{A})$ is the sum-of-squares K largest singular values of \mathbf{A} . If the constraint in Equation (15) is not imposed, then one of the possible choices is obtained by the least squares problem [49,50]; the solution is derived by observing the point where the derivative of cost function with respect to $\mathbf{T}_{\{f\}}$ is zero, then we can have $\mathbf{T}_{\{f\}}^{\text{LS}} = \mathbf{R}_{\{f_0,f_0\}} \mathbf{R}_{\{f,f_0\}}^H (\mathbf{R}_{\{f,f_0\}} \mathbf{R}_{\{f,f_0\}}^H)^{-1}$, and $\mathbf{\Sigma}_{\{f_0,f_0\}_s} \mathbf{\Sigma}_{\{f,f_0\}_s}^{-1} = \mathbf{I}_K$, which is difficult in practice. To solve the problem much more practically, an alternative solution is introduced, which is based on the constraint in Equation (15) and Lemma 1:

Theorem 1. Let $\mathbf{U}_{\psi\{f\}}, \mathbf{V}_{\psi\{f\}} \in \mathbb{C}^{M \times K}$ are the matrices in signal subspace containing the left and right singular vectors of $\mathbf{R}_{\{f,f_0\}} \mathbf{R}_{\{f_0,f_0\}}^H$. Imposing the constraint in Equation (15) and Lemma 1, along with the modification of orthogonal Procrustes problem (MOP), an alternative solution to Equation (15) is given by

$$\mathbf{T}_{\{f\}_{\text{MOP}}} = \mathbf{V}_{\psi\{f\}} \mathbf{U}_{\psi\{f\}}^{\dagger}, \quad (16)$$

where \dagger stands for the pseudo-inverse of a matrix. Defining the square matrix $\mathbf{\Omega}_{\{f\}} \in \mathbb{C}^{K \times K}$ as the matrix containing error corrections, the error of transformation remains consistent with the following equation;

$$\varepsilon_{\{f\}_{\text{MOP}}} = \left| 2 \cdot \Re \left(\text{tr} \left(\mathbf{\Sigma}_{\psi\{f\}} \left(\mathbf{\Omega}_{\{f\}} - \mathbf{I}_K \right) \right) \right) + \text{tr} \left(\mathbf{\Sigma}_{\{f,f_0\}_n}^2 \left(\mathbf{U}_{\psi\{f\}}^{\dagger} \mathbf{U}_{\varepsilon\{f\}} \right)^H \mathbf{U}_{\psi\{f\}}^{\dagger} \mathbf{U}_{\varepsilon\{f\}} \right) \right| \quad (17)$$

where $\mathbf{\Sigma}_{\psi\{f\}} \in \mathbb{R}^{K \times K}$ and $\mathbf{U}_{\varepsilon\{f\}} \in \mathbb{C}^{M \times M-K}$ are the diagonal matrix of the K largest singular values, and the noise subspace left singular vectors of $\mathbf{R}_{\{f,f_0\}} \mathbf{R}_{\{f_0,f_0\}}^H$, respectively.

Proof. See Appendix A. \square

Theorem 1 provides an efficient way to construct $\mathbf{T}_{\{f\}}$ without any process of DOA preliminary estimation, but the solution are still solely based on subspace between pair of distinct frequencies. In order to observe the solution across all frequency bins, we will present an alternative for constructing $\mathbf{T}_{\{f\}}$ by using HOGSVD along with Theorem 1, which the next section will address further.

3.2. Estimation of the Transformation Matrices by HOGSVD

Suppose we have a set of P complex matrices $\mathbf{E}_{\{f_i\}} \in \mathbb{C}^{M \times M}$ and all of them have a full rank;

$$\begin{aligned} \mathbf{E}_{\{f_1\}} &= \mathbf{R}_{\{f_1,f_0\}} \mathbf{R}_{\{f_0,f_0\}}^H, \\ \mathbf{E}_{\{f_2\}} &= \mathbf{R}_{\{f_2,f_0\}} \mathbf{R}_{\{f_0,f_0\}}^H, \\ &\vdots \\ \mathbf{E}_{\{f_P\}} &= \mathbf{R}_{\{f_P,f_0\}} \mathbf{R}_{\{f_0,f_0\}}^H, \end{aligned} \quad (18)$$

where $\{f_1, f_2, \dots, f_P\}$ is a set of frequency intervals, and the cross-correlation matrices $\mathbf{R}_{\{f_i, f_o\}}$ and $\mathbf{R}_{\{f_o, f_o\}}$ are obtained from Equation (7). The definition of HOGSVD of these P matrices are given by the generalized singular value decomposition (GSVD) of $P \geq 2$ datasets and its right singular vectors are identical in all decomposition [42], as follows:

$$\begin{bmatrix} \mathbf{E}_{\{f_1\}} \\ \mathbf{E}_{\{f_2\}} \\ \vdots \\ \mathbf{E}_{\{f_P\}} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_{e\{f_1\}_s} \boldsymbol{\Sigma}_{e\{f_1\}_s} \\ \mathbf{U}_{e\{f_2\}_s} \boldsymbol{\Sigma}_{e\{f_2\}_s} \\ \vdots \\ \mathbf{U}_{e\{f_P\}_s} \boldsymbol{\Sigma}_{e\{f_P\}_s} \end{bmatrix} \mathbf{V}_{e_s}^H + \begin{bmatrix} \mathbf{U}_{e\{f_1\}_n} \boldsymbol{\Sigma}_{e\{f_1\}_n} \\ \mathbf{U}_{e\{f_2\}_n} \boldsymbol{\Sigma}_{e\{f_2\}_n} \\ \vdots \\ \mathbf{U}_{e\{f_P\}_n} \boldsymbol{\Sigma}_{e\{f_P\}_n} \end{bmatrix} \mathbf{V}_{e_n}^H, \quad (19)$$

where $\mathbf{U}_{e\{f_i\}_s} \in \mathbb{C}^{M \times K}$, $\mathbf{U}_{e\{f_i\}_n} \in \mathbb{C}^{M \times M-K}$ are the matrix of left singular vectors, $\mathbf{V}_{e_s} \in \mathbb{C}^{M \times K}$, $\mathbf{V}_{e_n} \in \mathbb{C}^{M \times M-K}$ are the matrix of right singular vectors, and $\boldsymbol{\Sigma}_{e\{f_i\}_s} \in \mathbb{R}^{K \times K}$, $\boldsymbol{\Sigma}_{e\{f_i\}_n} \in \mathbb{R}^{M-K \times M-K}$ are the diagonal matrix of singular values. Note that subscripts s and n denote subspace of signal and noise, respectively. Unlike the left singular vectors $\mathbf{U}_{\{f_i, f_o\}_s}$ and $\mathbf{U}_{\{f_i, f_o\}_n}$ that have orthonormal columns by performing SVD, $\mathbf{U}_{e\{f_i\}_s}$ and $\mathbf{U}_{e\{f_i\}_n}$ now have unit 2-norm columns instead.

To show that \mathbf{V}_{e_s} is equal to $\mathbf{V}_{\psi\{f\}_s}$ for all frequency bins, let us start from brief description of HOGSVD benchmark. The matrix \mathbf{V}_{e_s} is obtained by performing EVD on the following matrix;

$$\mathbf{S} = \frac{1}{P(P-1)} \sum_{i=1}^P \sum_{j=i+1}^P \left(\left(\mathbf{E}_{\{f_i\}}^H \mathbf{E}_{\{f_i\}} \right) \left(\mathbf{E}_{\{f_j\}}^H \mathbf{E}_{\{f_j\}} \right)^{-1} + \left(\mathbf{E}_{\{f_j\}}^H \mathbf{E}_{\{f_j\}} \right) \left(\mathbf{E}_{\{f_i\}}^H \mathbf{E}_{\{f_i\}} \right)^{-1} \right). \quad (20)$$

Let us redefine

$$\mathbf{E}_{\{f_i\}} = \mathbf{U}_{o\{f_i\}} \boldsymbol{\Sigma}_{o\{f_i\}} \mathbf{V}_{o\{f_i\}}^H, \quad (21)$$

where

$$\begin{aligned} \mathbf{U}_{o\{f_i\}} &= \begin{bmatrix} \mathbf{U}_{\psi\{f_i\}} & \mathbf{U}_{\varepsilon\{f_i\}} \end{bmatrix}, \\ \mathbf{V}_{o\{f_i\}} &= \begin{bmatrix} \mathbf{V}_{\psi\{f_i\}} & \mathbf{V}_{\varepsilon\{f_i\}} \end{bmatrix}, \\ \boldsymbol{\Sigma}_{o\{f_i\}} &= \begin{bmatrix} \boldsymbol{\Sigma}_{\psi\{f_i\}} & \mathbf{O}_{K \times M-K} \\ \mathbf{O}_{M-K \times K} & \boldsymbol{\Sigma}_{\varepsilon\{f_i\}} \end{bmatrix}, \end{aligned} \quad (22)$$

$\boldsymbol{\Sigma}_{\varepsilon\{f_i\}} \in \mathbb{R}^{M-K \times M-K}$ is the matrix of the $M-K$ smallest singular values of $\mathbf{R}_{\{f_i, f_o\}} \mathbf{R}_{\{f_o, f_o\}}^H$ and $\mathbf{V}_{\psi\{f_i\}} = \mathbf{Q}_{\{f_o\}_s}$, $\mathbf{V}_{\varepsilon\{f_i\}} = \mathbf{Q}_{\{f_o\}_n}$ by employing Theorem 1 (For details, see Appendix A). Substituting Equations (21) and (22) into Equation (20), we have

$$\begin{aligned} \left(\mathbf{E}_{\{f_i\}}^H \mathbf{E}_{\{f_i\}} \right) \left(\mathbf{E}_{\{f_j\}}^H \mathbf{E}_{\{f_j\}} \right)^{-1} &= \mathbf{V}_{o\{f_i\}} \boldsymbol{\Sigma}_{o\{f_i\}} \boldsymbol{\Sigma}_{o\{f_i\}}^{-1} \boldsymbol{\Sigma}_{o\{f_j\}}^{-1} \boldsymbol{\Sigma}_{o\{f_j\}} \mathbf{V}_{o\{f_j\}}^{-1}, \\ \left(\mathbf{E}_{\{f_j\}}^H \mathbf{E}_{\{f_j\}} \right) \left(\mathbf{E}_{\{f_i\}}^H \mathbf{E}_{\{f_i\}} \right)^{-1} &= \mathbf{V}_{o\{f_j\}} \boldsymbol{\Sigma}_{o\{f_j\}} \boldsymbol{\Sigma}_{o\{f_j\}}^{-1} \boldsymbol{\Sigma}_{o\{f_i\}}^{-1} \boldsymbol{\Sigma}_{o\{f_i\}} \mathbf{V}_{o\{f_i\}}^{-1}. \end{aligned} \quad (23)$$

Since $\mathbf{V}_{\psi\{f_i\}} = \mathbf{Q}_{\{f_o\}_s} = \mathbf{V}_{\psi\{f_j\}}$, $\mathbf{V}_{\varepsilon\{f_i\}} = \mathbf{Q}_{\{f_o\}_n} = \mathbf{V}_{\varepsilon\{f_j\}}$ for all frequency bins, therefore

$$\mathbf{S} = \mathbf{V}_e \left(\frac{1}{P(P-1)} \sum_{i=1}^P \sum_{j=i+1}^P \left(\boldsymbol{\Sigma}_{o\{f_i\}} \boldsymbol{\Sigma}_{o\{f_i\}}^{-1} \boldsymbol{\Sigma}_{o\{f_j\}}^{-1} \boldsymbol{\Sigma}_{o\{f_j\}} + \boldsymbol{\Sigma}_{o\{f_j\}} \boldsymbol{\Sigma}_{o\{f_j\}}^{-1} \boldsymbol{\Sigma}_{o\{f_i\}}^{-1} \boldsymbol{\Sigma}_{o\{f_i\}} \right) \right) \mathbf{V}_e^{-1}. \quad (24)$$

where

$$\mathbf{V}_e = \begin{bmatrix} \mathbf{Q}_{\{f_o\}_s} & \mathbf{Q}_{\{f_o\}_n} \end{bmatrix}. \quad (25)$$

Performing EVD in Equation (24), we can obtain V_{e_s} , which reveal that V_{e_s} is equal to $V_{\psi\{f\}}$ for all frequency bins. In addition, it can be seen that the matrix V_{e_s} or $V_{\psi\{f\}}$ is estimated by focusing all frequency bins simultaneously; when power of the source at some frequency is weak or less than noise power, the matrices $V_{\psi\{f\}}$ still share common angle of ϕ, θ across all frequency bands effectively and identically.

After obtaining the right singular vectors of $E_{\{f_i\}}$, we then moved forward to find its left singular vectors. We start by considering the following equations based in Equations (19) and (25);

$$\begin{bmatrix} E_{\{f_1\}} \\ E_{\{f_2\}} \\ \vdots \\ E_{\{f_P\}} \end{bmatrix} V_e = \begin{bmatrix} \mathbf{U}_{e\{f_1\}_s} \Sigma_{e\{f_1\}_s} \\ \mathbf{U}_{e\{f_2\}_s} \Sigma_{e\{f_2\}_s} \\ \vdots \\ \mathbf{U}_{e\{f_P\}_s} \Sigma_{e\{f_P\}_s} \end{bmatrix} \begin{bmatrix} \mathbf{I}_K & \mathbf{O}_{K \times M-K} \end{bmatrix} + \begin{bmatrix} \mathbf{U}_{e\{f_1\}_n} \Sigma_{e\{f_1\}_n} \\ \mathbf{U}_{e\{f_2\}_n} \Sigma_{e\{f_2\}_n} \\ \vdots \\ \mathbf{U}_{e\{f_P\}_n} \Sigma_{e\{f_P\}_n} \end{bmatrix} \begin{bmatrix} \mathbf{O}_{M-K \times K} & \mathbf{I}_{M-K} \end{bmatrix}. \quad (26)$$

We remark again that $\mathbf{U}_{e\{f_i\}_s}, \mathbf{U}_{e\{f_i\}_n}$ have unit 2-norm columns instead of orthonormal columns [42];

$$\begin{bmatrix} \mathbf{U}_{e\{f_i\}_s}^H \\ \mathbf{U}_{e\{f_i\}_n}^H \end{bmatrix} \begin{bmatrix} \mathbf{U}_{e\{f_i\}_s} & \mathbf{U}_{e\{f_i\}_n} \end{bmatrix} = \begin{bmatrix} 1 & \zeta_{12} & \cdots & \zeta_{1M} \\ \zeta_{21} & 1 & \cdots & \zeta_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \zeta_{M1} & \zeta_{M2} & \cdots & 1 \end{bmatrix}, \quad (27)$$

where $\zeta_{jk} \in \mathbb{C}, \forall j \in M, \forall k \in M : j \neq k$. Then, the singular values are obtained as follows:

$$\begin{aligned} \Sigma_{e\{f_i\}_s} &= \text{diag} (\|e_1\|_2, \|e_2\|_2, \dots, \|e_K\|_2), \\ \Sigma_{e\{f_i\}_n} &= \text{diag} (\|e_{K+1}\|_2, \|e_{K+2}\|_2, \dots, \|e_M\|_2), \end{aligned} \quad (28)$$

where $\|\cdot\|_2$ is the Euclidean norm, and $e_j \in \mathbb{C}^M$ is a j^{th} column of $E_{\{f_i\}} V_e$. Finally, the matrices $\mathbf{U}_{e\{f_i\}_s}, \mathbf{U}_{e\{f_i\}_n}$ are obtained by solving Equation (26) with Equation (28), which also satisfy the condition in Equation (27).

After performing HOGSVD of Equation (18) to obtain the left and right singular vectors of $\mathbf{R}_{\{f_i, f_o\}} \mathbf{R}_{\{f_o, f_o\}}^H$, the transformation matrices $T_{\{f_i\}_{\text{MOP}}}$ can be assembled as follows:

$$\begin{bmatrix} T_{\{f_1\}_{\text{MOP}}} \\ T_{\{f_2\}_{\text{MOP}}} \\ \vdots \\ T_{\{f_P\}_{\text{MOP}}} \end{bmatrix} = V_{e_s} \begin{bmatrix} \mathbf{U}_{e\{f_1\}_s}^\dagger \\ \mathbf{U}_{e\{f_2\}_s}^\dagger \\ \vdots \\ \mathbf{U}_{e\{f_P\}_s}^\dagger \end{bmatrix}. \quad (29)$$

Note that since orthonormal columns have not yet been assumed on the matrix $\mathbf{U}_{\psi\{f\}}$ in Theorem 1, the transformation procedure via HOGSVD is still compatible with Theorem 1 without requiring any modifications (For details, see Equations (A13) and (A14) in Appendix A).

We now consider the computational complexity of HOGSVD. It is not surprising that HOGSVD has a heavy computational burden; that is because matrix inversions are intensively used in Equation (20). To avoid the computational burden caused by the matrix inversions, Equation (20) is reformulated by the following technique [51]. It begins by performing the economy-sized QR decomposition of Equation (19);

$$\begin{bmatrix} E_{\{f_1\}} \\ E_{\{f_2\}} \\ \vdots \\ E_{\{f_P\}} \end{bmatrix} = \begin{bmatrix} Q_{\zeta_1} \\ Q_{\zeta_2} \\ \vdots \\ Q_{\zeta_P} \end{bmatrix} R_\zeta, \quad (30)$$

where $R_{\zeta_i} \in \mathbb{C}^{M \times M}$ is the upper triangular matrix, and $Q_{\zeta_i} \in \mathbb{C}^{M \times M}$ is a one portion of the $(M \times P)$ -by- M matrix resulting from the QR decomposition of Equation (19). Next, S is simplified as

$$S_{\zeta} = \frac{1}{P(P-1)} (D_{\zeta} - PI_M), \tag{31}$$

where

$$D_{\zeta} = \sum_{i=1}^P (Q_{\zeta_i}^H Q_{\zeta_i})^{-1}. \tag{32}$$

Performing EVD of Equation (32), then we have $D_{\zeta} = Z_{\zeta} \Lambda_{\zeta} Z_{\zeta}^H$, where $Z_{\zeta} \in \mathbb{C}^{M \times M}$ and $\Lambda_{\zeta} \in \mathbb{R}^{M \times M}$ are the matrix of eigenvectors and matrix of eigenvalues, respectively. Finally, the alternative computation of V_e is expressed as $R_{\zeta}^H Z_{\zeta}$, where the K smallest eigenvalues of D_{ζ} are belonged to signal subspace.

Computational complexity of conventional HOGSVD in Equation (20) and optimized HOGSVD in Equation (32) are investigated by applying the following scenario: an $M \times M$ matrix addition, subtraction, multiplication and element-wise multiplication follow the traditional way, whereas an $M \times M$ matrix inversion and QR decomposition of $PM \times M$ matrix are implemented by using Gauss–Jordan elimination algorithm and Householder transformation, respectively. Comparing computation costs of Equations (20) and (32) from Tables 1 and 2, it is clearly seen that the technique in Equations (31) and (32) simplifies the mathematical model, reduces the matrix operations and improves the speed of V_e computation. When $P = M^i \quad \forall i : i > 0$, the optimized HOGSVD has arithmetic complexity of $\mathcal{O}(M^{i+3})$, which exhibits remarkably less computational complexity than the conventional HOGSVD that is presented as $\mathcal{O}(M^{2i+3})$. Since P in most cases is much greater than M , therefore, the cost of the optimized HOGSVD can logically be less than the conventional HOGSVD.

Table 1. Command used in HOGSVD.

Command Name	Command Counts	
	HOGSVD in Equation (20)	Optimized HOGSVD in Equation (32)
Matrix Addition/Subtraction	$P(P-1) - 1$	$P - 1$
Element-wise Multiplication	1	0
Matrix Multiplication	$3P(P-1)$	$P + \{1\}^*$
Matrix Inversion	$P(P-1)$	P
QR Decomposition	0	1
Eigenvalue Decomposition (EVD)	1	1

Remark: $\{\dots\}^*$ is caused by a matrix multiplication of $R_{\zeta}^H Z_{\zeta}$.

Table 2. Computational complexities.

Command Name	Complex Floating Point Operations per Command
Matrix Addition/Subtraction	M^2
Element-wise Multiplication	M^2
Matrix Multiplication	$2M^3 - M^2$
Matrix Inversion (Gauss-Jordan elimination)	$\frac{2}{3}M^3 + \frac{3}{2}M^2 - \frac{7}{6}M$
QR Decomposition (Householder transformation)	$(2P - \frac{2}{3})M^3 - 2PM^2 + \frac{2}{3}M$
HOGSVD in Equation (20) without counting EVD	$\frac{20P(P-1)}{3}M^3 - \frac{P(P-1)}{2}M^2 - \frac{7P(P-1)}{6}M$
Optimized HOGSVD in Equation (32) without counting EVD	$\frac{(14P+4)}{3}M^3 - \frac{(P+4)}{2}M^2 - \frac{(7P-4)}{6}M$

3.3. DOA Estimation Scheme

After the transformation matrices are formed by using HOGSVD, we now proceed to describe a framework for estimating the wideband DOAs. We start by simplifying the wideband cross-correlation matrix in Equation (8) with EVD form and substituting with $T_{\{f_i\}_{\text{MOP}}}$, as follows:

$$\begin{aligned} & \frac{1}{P} \sum_{i=1}^P T_{\{f_i\}_{\text{MOP}}} \mathbf{R}_{\{f_i, f_i\}} T_{\{f_i\}_{\text{MOP}}}^H \\ &= \frac{1}{P} \sum_{i=1}^P \left(\mathbf{V}_{e_s} \mathbf{U}_{e\{f_i\}_s}^\dagger \right) \left(\mathbf{Q}_{\{f_i\}_s} \mathbf{\Lambda}_{\{f_i\}_s} \mathbf{Q}_{\{f_i\}_s}^H + \mathbf{Q}_{\{f_i\}_n} \mathbf{\Lambda}_{\{f_i\}_n} \mathbf{Q}_{\{f_i\}_n}^H \right) \left(\mathbf{V}_{e_s} \mathbf{U}_{e\{f_i\}_s}^\dagger \right)^H \\ &= \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H + \mathbf{\Pi}, \end{aligned} \quad (33)$$

where

$$\begin{aligned} \mathbf{\Lambda} &= \mathbf{L}^H \left(\frac{1}{P} \sum_{i=1}^P \left(\mathbf{U}_{e\{f_i\}_s}^\dagger \mathbf{Q}_{\{f_i\}_s} \right) \mathbf{\Lambda}_{\{f_i\}_s} \left(\mathbf{U}_{e\{f_i\}_s}^\dagger \mathbf{Q}_{\{f_i\}_s} \right)^H \right) \mathbf{L}, \\ \mathbf{\Pi} &= \mathbf{V}_{e_s} \left(\frac{1}{P} \sum_{i=1}^P \left(\mathbf{U}_{e\{f_i\}_s}^\dagger \mathbf{Q}_{\{f_i\}_n} \right) \mathbf{\Lambda}_{\{f_i\}_n} \left(\mathbf{U}_{e\{f_i\}_s}^\dagger \mathbf{Q}_{\{f_i\}_n} \right)^H \right) \mathbf{V}_{e_s}^H, \end{aligned} \quad \mathbf{Q} = \mathbf{V}_{e_s} \mathbf{L}. \quad (34)$$

Here, $\mathbf{\Lambda} \in \mathbb{C}^{K \times K}$ and $\mathbf{Q} \in \mathbb{C}^{M \times K}$ are the diagonal matrix of eigenvalues and matrix of eigenvectors of Equation (33) in signal subspace, and $\mathbf{L} \in \mathbb{C}^{K \times K}$ possess unitary property by the fact that $\mathbf{Q}, \mathbf{V}_{e_s}$ are the matrices with orthonormal columns [46,47]. Remark that $\mathbf{R}_{\{f_i, f_i\}}$ is also derived by performing EVD; the matrices $\mathbf{Q}_{\{f_i\}_s} \in \mathbb{C}^{M \times K}$, $\mathbf{\Lambda}_{\{f_i\}_s} \in \mathbb{R}^{K \times K}$ are the eigenvectors and diagonal matrix of eigenvalues in signal subspace, and likewise, $\mathbf{Q}_{\{f_i\}_n} \in \mathbb{C}^{M \times M-K}$, $\mathbf{\Lambda}_{\{f_i\}_n} \in \mathbb{R}^{M-K \times M-K}$ are with noise subspace. Furthermore, considering only the signal subspace by focusing on the K largest singular values $\mathbf{\Lambda}$, we can expect that Equation (33) is equivalent to Equation (8);

$$\mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H \equiv \mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f_o) \left(\frac{1}{P} \sum_{i=1}^P \mathbf{S}_{\{f_i, f_i\}} \right) \mathbf{A}^H(\boldsymbol{\phi}, \boldsymbol{\theta}, f_o), \quad (35)$$

which can be proved by employing Lemma 1, Equations (12)–(14), and Equations (A4)–(A6) on Appendix A (We omit the proof since the result is easily obtained by performing straightforward substitution). In this state, $T_{\{f\}_{\text{MOP}}}$ provides an efficient way to transform any given f into f_o by observing the solution across frequency bands without loss of generality; it means that the transformation is no longer biased by the pair of distinct frequencies $\{f, f_o\}$. Furthermore, it is clearly seen that the wideband cross-correlation matrix in Equation (33) is the combination of narrowband sample cross-correlation matrices across all frequency bins, but its array manifolds are focused on the single reference frequency by using $T_{\{f\}_{\text{MOP}}}$, which is now feasible to estimate the wideband DOAs by employing any recent subspace-based technique for estimating narrowband DOAs [18,20–26], but using this wideband correlation matrix instead of the narrowband correlation matrix. Practical examples, such as MUSIC and ESPRIT, will be presented to showcase its applicability and effectiveness in the next section.

Remarks: In case of the L-shaped array structure in Equation (2), we can repeat the proposed transformation procedure to find the solution for x subarray in Equation (2) and (3); starting from Equation (7) by replacing $\mathbf{r}(t, f)$ with $\mathbf{x}(t, f)$, the solution for the x subarray can be given by:

$$\mathbf{T}_{x\{f_i\}_{\text{MOP}}} = \mathbf{V}_{x, e_s} \mathbf{U}_{x, e\{f_i\}_s}^\dagger, \quad (36)$$

$$\frac{1}{P} \sum_{i=1}^P \mathbf{T}_{x\{f_i\}_{\text{MOP}}} \mathbf{R}_{x\{f_i, f_i\}} \mathbf{T}_{x\{f_i\}_{\text{MOP}}}^H = \mathbf{Q}_x \mathbf{\Lambda}_x \mathbf{Q}_x^H + \mathbf{\Pi}_x, \quad (37)$$

$$\mathbf{Q}_x \mathbf{\Lambda}_x \mathbf{Q}_x^H \equiv \mathbf{A}_x(\boldsymbol{\phi}, f_0) \left(\frac{1}{P} \sum_{i=1}^P \mathbf{S}_{\{f_i, f_i\}} \right) \mathbf{A}_x^H(\boldsymbol{\phi}, f_0). \quad (38)$$

By performing the same procedure, the solution for z subarray is likewise given by replacing $\mathbf{x}(t, f)$, $\mathbf{A}_x(\boldsymbol{\phi}, f_0)$ with $\mathbf{z}(t, f)$, $\mathbf{A}_z(\boldsymbol{\theta}, f_0)$ and the subscript x with z in Equations (36)–(38);

$$\mathbf{T}_{z\{f_i\}_{\text{MOP}}} = \mathbf{V}_{z, e_s} \mathbf{U}_{z, e\{f_i\}_s}^{\dagger} \quad (39)$$

$$\frac{1}{P} \sum_{i=1}^P \mathbf{T}_{z\{f_i\}_{\text{MOP}}} \mathbf{R}_{z\{f_i, f_i\}} \mathbf{T}_{z\{f_i\}_{\text{MOP}}}^H = \mathbf{Q}_z \mathbf{\Lambda}_z \mathbf{Q}_z^H + \mathbf{\Pi}_z, \quad (40)$$

$$\mathbf{Q}_z \mathbf{\Lambda}_z \mathbf{Q}_z^H \equiv \mathbf{A}_z(\boldsymbol{\theta}, f_0) \left(\frac{1}{P} \sum_{i=1}^P \mathbf{S}_{\{f_i, f_i\}} \right) \mathbf{A}_z^H(\boldsymbol{\theta}, f_0). \quad (41)$$

3.3.1. DOA Estimation Scheme via MUSIC

MUSIC estimates the DOA of the sources by locating the peaks of MUSIC spectrum along with exploiting the orthogonality of the signal and noise subspaces [12,48]. Let us define the complementary orthogonal space $(\mathbf{I}_M - \mathbf{Q}\mathbf{Q}^H)$ which is orthogonal to $\mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f_0)$;

$$\mathbf{a}^H(\phi_k, \theta_k, f_0) (\mathbf{I}_M - \mathbf{Q}\mathbf{Q}^H) \mathbf{a}(\phi_k, \theta_k, f_0) = 0, \quad (42)$$

for all $k \in \{1, 2, \dots, K\}$, where $\mathbf{a}(\phi_k, \theta_k, f_0) \in \mathbb{C}^M$ is a k^{th} column of $\mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f_0)$ as shown in Equation (3). Additionally, the following complementary orthogonal space is also valid;

$$\mathbf{a}^H(\phi_k, \theta_k, f_0) (\mathbf{I}_M - \mathbf{V}_{e_s} \mathbf{V}_{e_s}^H) \mathbf{a}(\phi_k, \theta_k, f_0) = 0, \quad (43)$$

by the fact that $\mathbf{Q}\mathbf{Q}^H = \mathbf{V}_{e_s} (\mathbf{L}\mathbf{L}^H) \mathbf{V}_{e_s}^H = \mathbf{V}_{e_s} \mathbf{V}_{e_s}^H$, which implies that it is possible to reduce a computational complexity of Equation (33) by using only \mathbf{V}_{e_s} instead of calculating \mathbf{Q} . The computationally efficient two-dimensional MUSIC (2D-MUSIC) spectrum is expressed as

$$p_{2\text{D-MUSIC}}(\phi, \theta) = \frac{1}{\mathbf{a}^H(\phi, \theta, f_0) (\mathbf{I}_M - \mathbf{V}_{e_s} \mathbf{V}_{e_s}^H) \mathbf{a}(\phi, \theta, f_0)}. \quad (44)$$

When the denominator in Equation (44) approaches zero for the true angles of the signals, the 2D-MUSIC spectrum will have peak spikes indicating this angles. In case of the L-shaped array structure, the x and z subarray angles are estimated separately by locating the spectral peaks of the following equations:

$$\begin{aligned} p_{x\text{MUSIC}}(\phi) &= \frac{1}{\mathbf{a}_x^H(\phi, f_0) (\mathbf{I}_N - \mathbf{V}_{x, e_s} \mathbf{V}_{x, e_s}^H) \mathbf{a}_x(\phi, f_0)}, \\ p_{z\text{MUSIC}}(\theta) &= \frac{1}{\mathbf{a}_z^H(\theta, f_0) (\mathbf{I}_N - \mathbf{V}_{z, e_s} \mathbf{V}_{z, e_s}^H) \mathbf{a}_z(\theta, f_0)} \end{aligned} \quad (45)$$

where $\mathbf{a}_x(\phi, f_0)$, $\mathbf{a}_z(\theta, f_0) \in \mathbb{C}^N$ are i^{th} column of $\mathbf{A}_x(\boldsymbol{\phi}, f_0)$, $\mathbf{A}_z(\boldsymbol{\theta}, f_0)$, respectively.

3.3.2. DOA Estimation Scheme via ESPRIT

We start by recalling the array manifold $\mathbf{A}_x(\boldsymbol{\phi}, f_0)$ and $\mathbf{A}_z(\boldsymbol{\theta}, f_0)$ in Equation (3). ESPRIT takes advantage of the rotational invariance property of ULA [13], as follows:

$$\begin{aligned} \mathbf{A}_{x_2}(\boldsymbol{\phi}, f_0) &= \mathbf{A}_{x_1}(\boldsymbol{\phi}, f_0) \boldsymbol{\Phi}_x, \\ \mathbf{A}_{z_2}(\boldsymbol{\theta}, f_0) &= \mathbf{A}_{z_1}(\boldsymbol{\theta}, f_0) \boldsymbol{\Theta}_z, \end{aligned} \quad (46)$$

where

$$\begin{aligned}\Phi_x &= \text{diag} \left(e^{\alpha_x(\phi_1, f_0)j}, e^{\alpha_x(\phi_2, f_0)j}, \dots, e^{\alpha_x(\phi_K, f_0)j} \right), \\ \Theta_z &= \text{diag} \left(e^{\alpha_z(\theta_1, f_0)j}, e^{\alpha_z(\theta_2, f_0)j}, \dots, e^{\alpha_z(\theta_K, f_0)j} \right),\end{aligned}\quad (47)$$

$A_{x_1}(\boldsymbol{\phi}, f_0), A_{z_1}(\boldsymbol{\theta}, f_0) \in \mathbb{C}^{N-1 \times K}$ and $A_{x_2}(\boldsymbol{\phi}, f_0), A_{z_2}(\boldsymbol{\theta}, f_0) \in \mathbb{C}^{N-1 \times K}$ stand for the first and last $(N-1)$ rows of $A_x(\boldsymbol{\phi}, f_0), A_z(\boldsymbol{\theta}, f_0)$, respectively. Similar to [20,21,26], the matrices $\mathbf{Q}_x, \mathbf{Q}_z$ can be simplified with Equations (3), (36)–(38) and (46), as follows:

$$\begin{aligned}\mathbf{Q}_{x_1} &= A_{x_1}(\boldsymbol{\phi}, f_0) \mathbf{C}_x^{-1}, & \mathbf{Q}_{z_1} &= A_{z_1}(\boldsymbol{\theta}, f_0) \mathbf{C}_z^{-1}, \\ \mathbf{Q}_{x_2} &= A_{x_2}(\boldsymbol{\phi}, f_0) \mathbf{C}_x^{-1}, & \mathbf{Q}_{z_2} &= A_{z_2}(\boldsymbol{\theta}, f_0) \mathbf{C}_z^{-1},\end{aligned}\quad (48)$$

where $\mathbf{C}_x, \mathbf{C}_z \in \mathbb{C}^{K \times K}$ are invertible matrices, $\mathbf{Q}_{x_1}, \mathbf{Q}_{z_1} \in \mathbb{C}^{N-1 \times K}$ and $\mathbf{Q}_{x_2}, \mathbf{Q}_{z_2} \in \mathbb{C}^{N-1 \times K}$ stand for the first and last $(N-1)$ rows of $\mathbf{Q}_x, \mathbf{Q}_z$, respectively. Considering Equation (48), we can construct new matrices Γ_x, Γ_z as follows:

$$\begin{aligned}\Gamma_x &= \mathbf{Q}_{x_1}^\dagger \mathbf{Q}_{x_2} & \Gamma_z &= \mathbf{Q}_{z_1}^\dagger \mathbf{Q}_{z_2} \\ &= \mathbf{C}_x \Phi_x \mathbf{C}_x^{-1}, & &= \mathbf{C}_z \Theta_z \mathbf{C}_z^{-1}.\end{aligned}\quad (49)$$

The angles ϕ_k, θ_k can thus be estimated by the eigenvalues of Γ_x, Γ_z , as follows:

$$\phi_k = \cos^{-1} \left(\text{angle}(\lambda_{x_k}) \frac{\lambda}{2\pi d} \right), \quad \theta_k = \cos^{-1} \left(\text{angle}(\lambda_{z_k}) \frac{\lambda}{2\pi d} \right), \quad (50)$$

where $\lambda_{x_k}, \lambda_{z_k} \in \mathbb{C}$ is the k^{th} eigenvalue of Γ_x, Γ_z , respectively. Furthermore, it is possible to reduce the computational complexity by using only \mathbf{V}_{e_s} as well as MUSIC;

$$\begin{aligned}\mathbf{V}_{x_1, e_s}^\dagger \mathbf{V}_{x_2, e_s} &= \mathbf{L}_x \Gamma_x \mathbf{L}_x^{-1} & \mathbf{V}_{z_1, e_s}^\dagger \mathbf{V}_{z_2, e_s} &= \mathbf{L}_z \Gamma_z \mathbf{L}_z^{-1} \\ &= (\mathbf{L}_x \mathbf{C}_x) \Phi_x (\mathbf{L}_x \mathbf{C}_x)^{-1}, & &= (\mathbf{L}_z \mathbf{C}_z) \Theta_z (\mathbf{L}_z \mathbf{C}_z)^{-1},\end{aligned}\quad (51)$$

where

$$\begin{aligned}\mathbf{V}_{x_1, e_s} &= A_{x_1}(\boldsymbol{\phi}, f_0) (\mathbf{L}_x \mathbf{C}_x)^{-1}, & \mathbf{V}_{z_1, e_s} &= A_{z_1}(\boldsymbol{\theta}, f_0) (\mathbf{L}_z \mathbf{C}_z)^{-1}, \\ \mathbf{V}_{x_2, e_s} &= A_{x_2}(\boldsymbol{\phi}, f_0) (\mathbf{L}_x \mathbf{C}_x)^{-1}, & \mathbf{V}_{z_2, e_s} &= A_{z_2}(\boldsymbol{\theta}, f_0) (\mathbf{L}_z \mathbf{C}_z)^{-1},\end{aligned}\quad (52)$$

$\mathbf{V}_{x_1, e_s}, \mathbf{V}_{z_1, e_s} \in \mathbb{C}^{N-1 \times K}$ and $\mathbf{V}_{x_2, e_s}, \mathbf{V}_{z_2, e_s} \in \mathbb{C}^{N-1 \times K}$ stand for the first and last $(N-1)$ rows of $\mathbf{V}_{x, e_s}, \mathbf{V}_{z, e_s}$, respectively.

4. Numerical Simulations

In this section, performances of fusion methods by using the proposed framework are demonstrated in four types of the following scenarios: (1) a performance of selected method and the proposed methods with respect to source types, (2) the performance with respect to the number of microphone elements, (3) the performance with considering automatic pairing of the x and z subarray angles, and (4) the performance under a reverberation environment. Scenarios 1, 2 and 4 have to find DOA of x and z subarray angles separately by using the data model in Equation (2). Whereas Scenario 3 has to find DOA of x and z subarray angles simultaneously with considering automatic pairing, by using the data model in Equation (1). We provided the simulation tests of the proposed methods in comparison to following methods: IMUSIC [31], TOFS [32], TOPS [33], Squared-TOPS [34], WS-TOPS [35]. Note that the CSS-based methods are excluded in these tests; this is because unintended biases, causing by a process of DOA preliminary estimation, should be taken into consideration to other candidate methods as discussed in the literature [31–33,41].

To measure the overall performance of estimating the x and z subarray angles for each scenario, the root-mean-square-error (RMSE) and standard division (SD) are defined as the following equations;

$$\text{RMSE} = \sqrt{\frac{1}{2JK} \sum_{j=1}^J \sum_{k=1}^K \left((\hat{\phi}_k^{(j)} - \phi_k)^2 + (\hat{\theta}_k^{(j)} - \theta_k)^2 \right)}, \quad (53)$$

$$\text{SD} = \sqrt{\frac{1}{2JK} \sum_{j=1}^J \sum_{k=1}^K \left((\hat{\phi}_k^{(j)} - \bar{\phi}_k)^2 + (\hat{\theta}_k^{(j)} - \bar{\theta}_k)^2 \right)}, \quad (54)$$

where K is the source number, J is the number of trials, $\hat{\phi}_k^{(j)}, \hat{\theta}_k^{(j)}$ represent the estimated x and z subarray angles each trial, $\bar{\phi}_k, \bar{\theta}_k$ represent an average of the estimated x and z subarray angles, and ϕ_k, θ_k represent true x and z subarray angles.

Computer simulations were carried out in Matlab[®] R2017a, using PC with Debian GNU/Linux 9.4 × 86_64, Intel[®] Core[™] i5-4590 CPU 3.30 GHz, 16G RAM, Intel[®] Math Kernel Library 11.3.1 on BLAS and LAPACK 3.5.0. Each scenario is repeated 100 times, and simulation parameters are chosen as follows: sampling frequency is 48 kHz, an output of each microphone is captured at 1 s, speed of sound c is 343 m/s, the spacing of microphone elements d is 5 cm, STFT focusing frequency range is from 0.1 to 16 kHz, the reference frequency f_o is 3.43 kHz. Note that we used perturbations of the true angles by adding Gaussian random noise.

4.1. Scenario 1: Performance with Respect to Source Types

Figures 2 and 3 showed performance comparisons of the selected methods and the proposed methods in term of RMSE and SD over a range of SNR. The proposed methods are the modified MUSIC in Equation (45) and ESPRIT in Equations (50)–(52). The number of microphone elements each subarray is six, and the three uncorrelated source angles (ϕ_k, θ_k) are placed at $(41.41^\circ, 60^\circ)$, $(60^\circ, 45^\circ)$ and $(75.52^\circ, 30^\circ)$. In Figures 2a and 3a, sources are human speeches. Sources in Figures 2b and 3b are recorded sound on a piano comprising various monochromatic notes and containing sampling frequency range up to 48 kHz. Note that all sources are not stationary signals. The results in Figures 2 and 3 showed that the proposed method with ESPRIT can efficiently handle both source types compared to other candidate methods with acceptable SNR ranges. Subsequently, it is interesting to take a close look at 40 dB SNR in Figures 2 and 3 where IMUSIC, TOFS, the proposed method with MUSIC and ESPRIT showed very low RMSE, which could attest to good DOA estimation. When decreasing the SNR to 25 dB, IMUSIC and TOFS begin to demonstrate worse RMSE quality, which is much higher than the proposed methods, and it is clearly seen when decreasing the SNR to 10 dB that all tested methods are significantly dominated, but the proposed method with ESPRIT is still associated with more satisfactory results compared to using other methods. It should be mentioned that IMUSIC and TOFS require the number of sensor elements to be much higher than the number of sources to achieve fairly good results [31–33,41]. Hence, the simulation results in Figures 2 and 3 are able to provide evidence that the proposed methods perform better in estimation than other candidate methods when the incident sources are wideband and non-stationary signals. Although the performances of the proposed method with MUSIC is also dominated by the noises, the overall performances is still more effective than other methods.

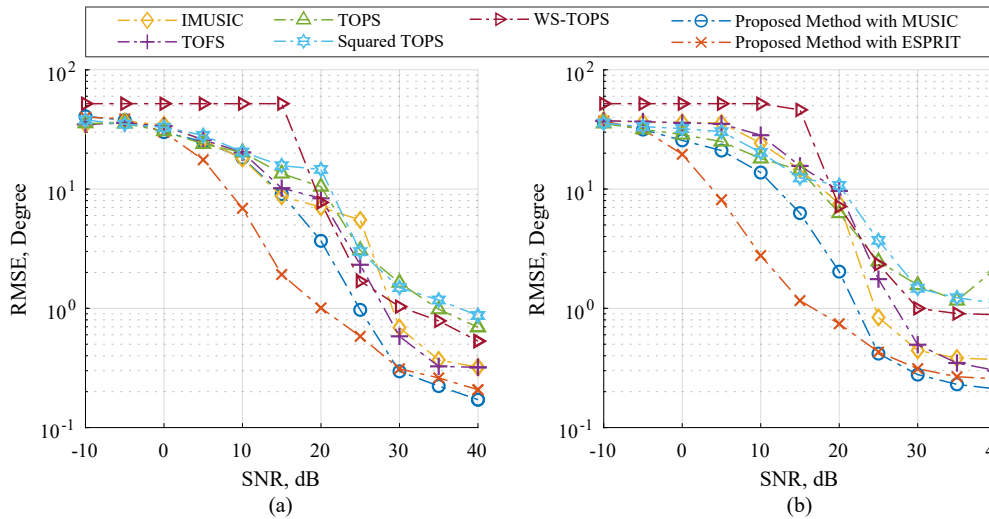


Figure 2. RMSE estimation performance versus SNR on Scenario 1; (a) three different human speeches, and (b) three uncorrelated musical sounds where six microphones are employed each subarray.

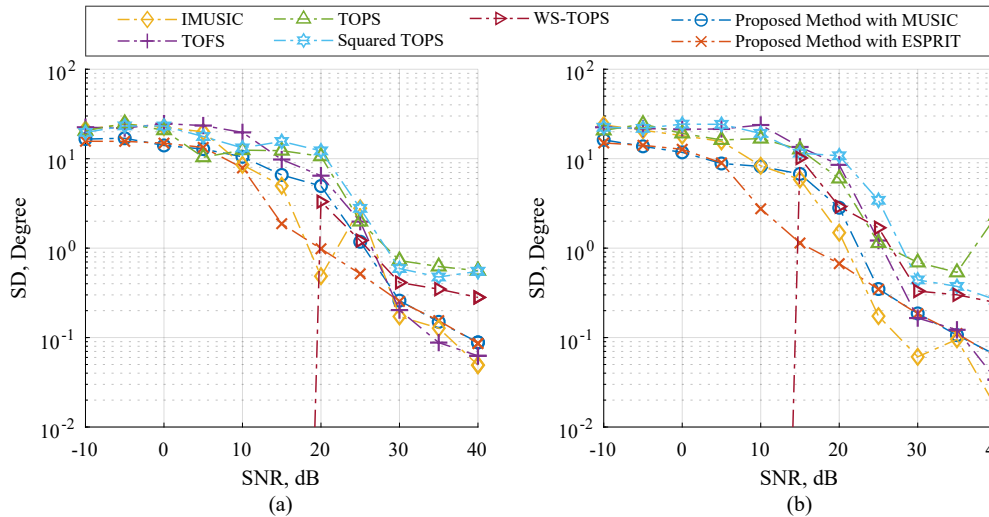


Figure 3. SD estimation performance versus SNR on Scenario 1; (a) three different human speeches, and (b) three uncorrelated musical sounds where six microphones is employed each subarray.

4.2. Scenario 2: Performance with Respect to the Number of Microphone Elements

Figures 4 and 5 illustrates performance comparisons of the selected methods and the proposed methods in term of RMSE and SD over a range of SNR. The three uncorrelated source angles are human speeches, and are placed as previously used. Firstly, let us start by looking at the case of twelve microphones in Figures 4c and 5c. IMUSIC, TOFS and WS-TOPS exhibited remarkably low levels of RMSE in the SNR range from 15 to 30 dB; this is because their performances dramatically depend on the number of sensor elements more than the number of sources [31–33,41]. Likewise, the proposed method with MUSIC and ESPRIT also demonstrated very low RMSE, which may imply that the performance of the proposed methods, IMUSIC, TOFS and WS-TOPS are especially effective for a wideband DOA estimation. However, the low number of microphone elements should be considered for providing more practical applications. In the case of eight microphones in each subarray, the performances of the selected methods are dominated by the number of microphone elements as illustrated in Figures 4b and 5b. Furthermore, the performances of selected methods are dramatically degraded when employing four microphones as illustrated in Figures 4a and 5a. The relevant reason is that an undesirable false peak in the spatial spectrum of the selected methods occurred, caused by the perturbation of noise; when power of the noise at some frequency is high or grater than source power,

the orthogonality between the noise subspace and search space at that frequency may be not sufficient to prevent the false-alarm peaks [41]. On the contrary, RMSE performance of the proposed methods are also dominated, but less than the other methods, by exhibiting the subspace for all frequency bins simultaneously as shown in Section 3. Therefore, the proposed methods provide substantially better RMSE performance than the other methods, which implies that dependency between the number of microphone elements and sources can be relaxed. This substantial ability is more meaningful for many practical applications.

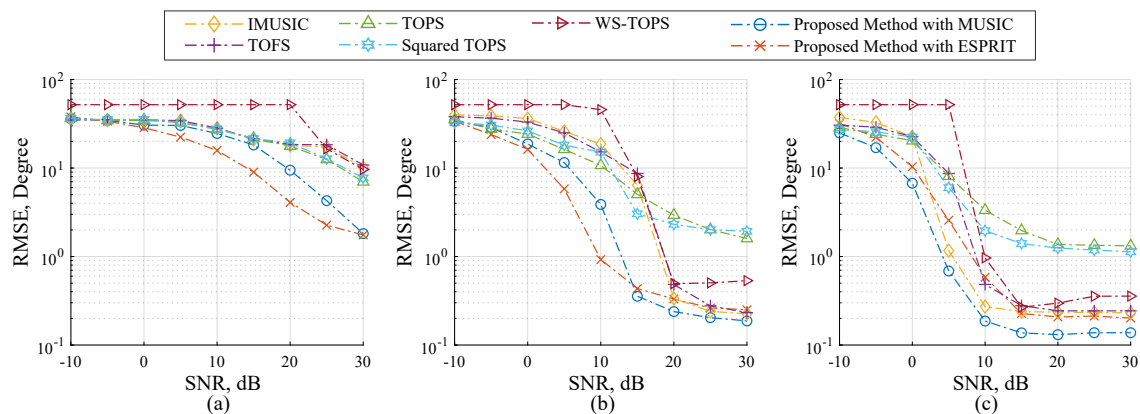


Figure 4. RMSE estimation performance versus SNR on Scenario 2; three human speeches are employed and the number of microphone elements each subarray on (a) $N = 4$, (b) $N = 8$, and (c) $N = 12$.

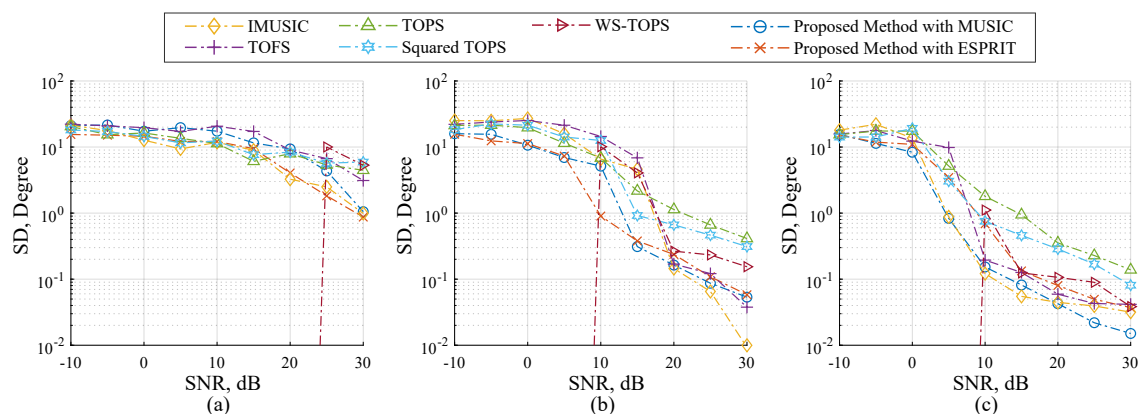


Figure 5. SD estimation performance versus SNR on Scenario 2; three human speeches are employed and the number of microphone elements each subarray on (a) $N = 4$, (b) $N = 8$, and (c) $N = 12$.

4.3. Scenario 3: Performance with Considering Automatic Pairing

This scenario estimated the DOA of x and z subarray angles simultaneously with considering automatic pairing and following the data model in Equation (1). As the L-shaped array structure consisting of two ULAs as illustrated in Figure 1, some research works estimate the DOA of x and z subarray angles separately by implementing 1D DOA estimation for each ULA [17–26]. When utilizing more than one source, these algorithms require an additional angle pair matching procedure to map the relationship between the two independent subarray angles. For instance, finding the corresponding angle pairs by rearranging the alignment of $\mathbf{a}_x(\phi_k, f)$ with a fixed right-hand side of the array manifolds of the z -subarray in the sample cross-covariance matrix [52]. It should be noted that a pair-matching procedure may result in a performance degradation caused by pair-matching error. In order to achieve the automatic pairing without the pair-matching procedure, we selected the modified 2D-MUSIC in Equation (44) as the proposed method in this scenario. Furthermore, TONS, Squared-TONS, WS-TONS are excluded in these tests by the fact that the methods have only supported the ULA model. Note that the 2D peak finding algorithm was employed on 2D-IMUSIC, 2D-TOFS and

the proposed method. Figures 6 and 7 showed performance comparisons of 2D-IMUSIC, 2D-TOFS and the proposed method in term of RMSE and SD over a range of SNR, where the number of microphone elements including all subarray is eight, the three uncorrelated source angles are human speeches, and are placed as previously used. Figure 6 indicates that the proposed method with 2D-MUSIC exhibits extremely similar overall performances to 2D-IMUSIC and 2D-TOFS when the SNR increases to more than 10 dB; however, computational burden of the proposed method can be significantly lower than those of the other methods, which Section 4.5 will reveal further insight.

4.4. Scenario 4: Performance under Reverberation Environment

In this scenario, we compared RMSE and SD performances of the proposed methods to other methods with respect to reverberation time. This scenario estimated DOA of x and z subarray angles separately by using the data model in Equation (2) without considering automatic pairing. The proposed methods in this scenario are the modified MUSIC in Equation (45) and ESPRIT on Equations (50)–(52). The reverberations were simulated by the following procedure [53], and its simulated wall absorption coefficients are shown in Table 3, where the dimensions of enclosure room is $15 \times 15 \times 5$ m, a measurement protocol of reverberation time is RT60, and the reverberation time is from 200 to 1000 ms. The three uncorrelated source angles are employed in the same way as previously used, and the number of microphone elements in each subarray is twelve. Figure 8 illustrated performance comparisons of the selected methods and the proposed methods, where a color of the graph on Figure 8a denotes RMSE, whereas a color of the graph on Figure 8b denotes SD estimation performance. The vertical axis is represented as the reverberation time and horizontal axis is represented as a range of SNR. Simulation results in Figure 8 indicated that reverberation has strong effects on RMSE and SD performances in both of the selected methods and the proposed methods, and the performances decreased more significantly at the high noise levels and the long reverberation times. Since the reverberation time is decreasing, all selected methods begin to demonstrate low RMSE. It means that the trade-off between the robustness of reverberation and SNR should be considered deeply in actual applications, for instance, applying a reverberation cancellation technique or a noise cancellation technique to provide much more reliable estimation performances of both RMSE and SD. The proposed methods, however, largely outperform the other methods with respect to the reverberation time index and SNR level range between 10 and 40 dB without considering the trade-off. This can support that the performance of the proposed methods can be especially effective for a wideband DOA estimation under a reverberant environment.

Table 3. Wall absorption coefficients at various reverberation time in Scenario 4 [53].

Reverberation Time based on RT60 (Millisecond)	Axial Wall Plane					
	Positive Direction			Negative Direction		
	$x - z$	$x - z$	$x - y$	$x - z$	$x - z$	$x - y$
200	0.7236	0.2021	0.6844	0.0792	0.2436	0.5586
300	0.7142	0.1687	0.7666	0.2650	0.2387	0.7043
400	0.7306	0.0555	0.7731	0.4091	0.8493	0.8587
500	0.5064	0.4974	0.8248	0.4189	0.8069	0.7572
600	0.6074	0.6299	0.8028	0.7599	0.6373	0.8209
700	0.7442	0.7624	0.8734	0.6922	0.6480	0.7893
800	0.6779	0.6827	0.7865	0.8045	0.8386	0.8430
900	0.6992	0.7111	0.7741	0.8752	0.8233	0.9081
1000	0.7622	0.7707	0.9394	0.8248	0.8192	0.8398

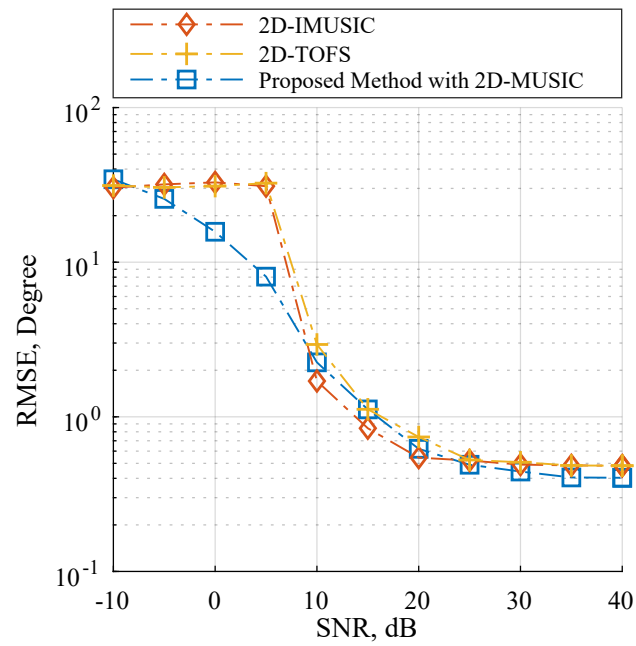


Figure 6. RMSE estimation performance versus SNR on Scenario 3 where $M = 8$.

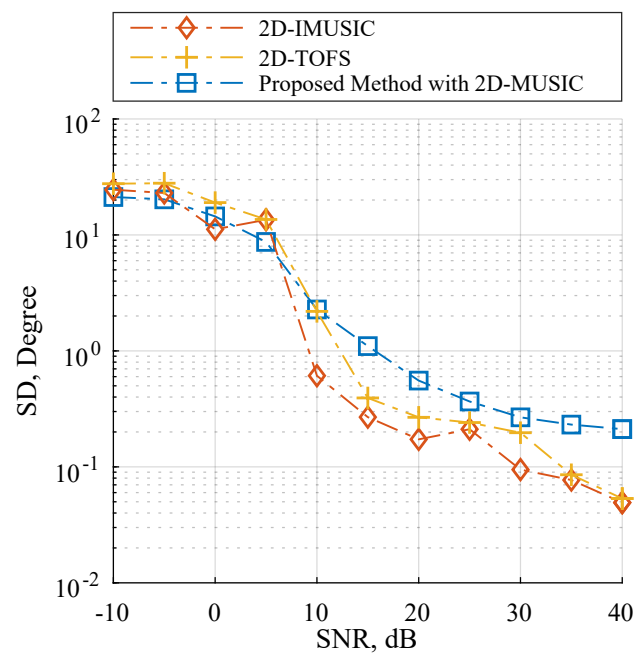


Figure 7. SD estimation performance versus SNR on Scenario 3 where $M = 8$.

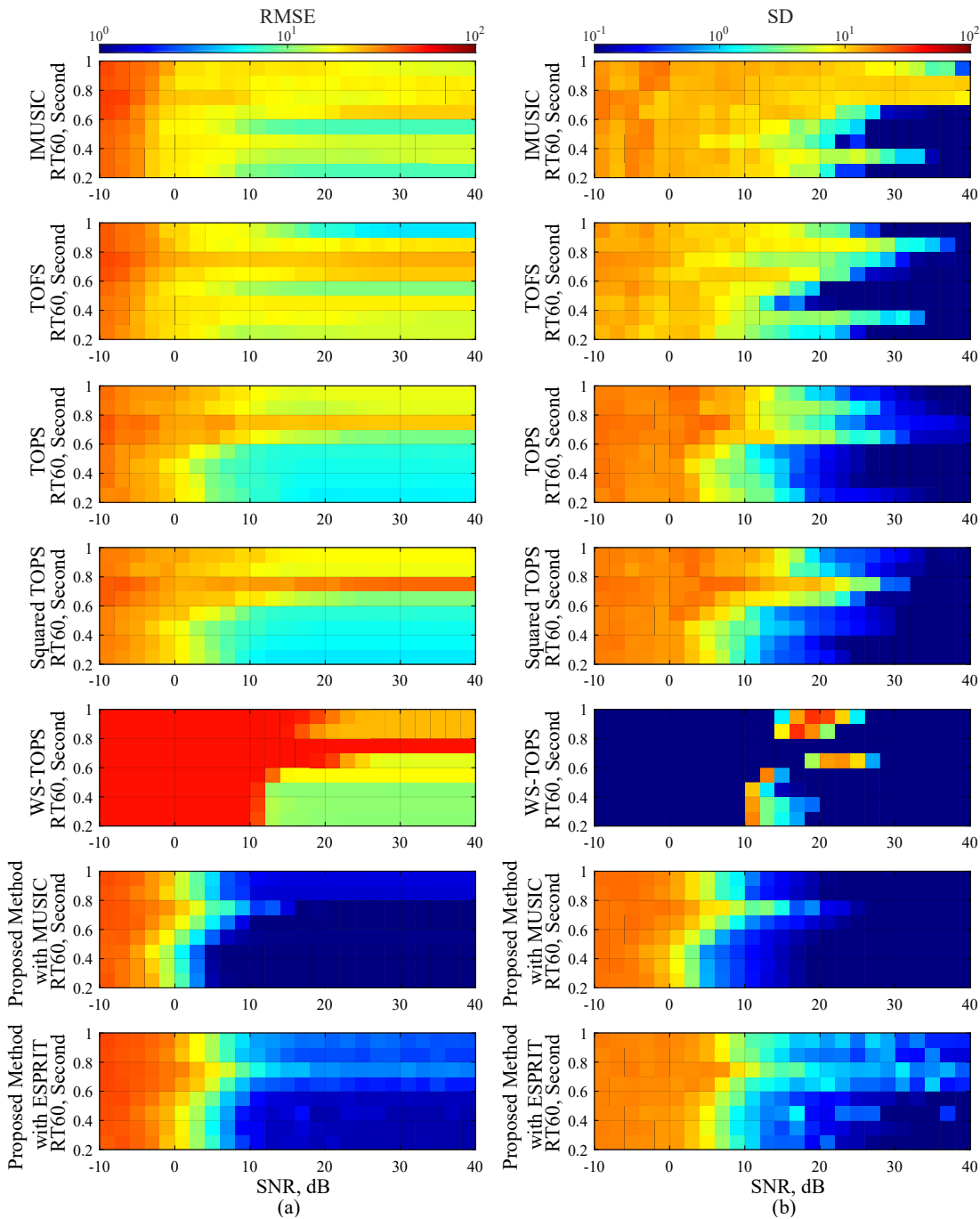


Figure 8. Performance evaluations of Scenario 4; (a) RMSE estimation performance versus SNR, and (b) SD estimation performance versus SNR, where three uncorrelated human speeches are employed along with a reverberant environment. The reverberations were simulated by the following procedure [53], where dimensions of enclosure room is $15 \times 15 \times 5$ m, a measurement protocol of reverberation time is RT60, and wall absorption coefficients are followed on Table 3.

4.5. Computational Complexity

Computational complexity of the proposed methods was evaluated using execution time measurement under a stable environment. We provided a computational complexity in comparison with the following cases: (1) calculating DOAs of x and z subarray angles separately as shown in Figure 9a, and (2) calculating the DOAs of both subarray angles simultaneously as shown in Figure 9b. Note that computational burdens of a peak searching algorithm are relevant in this study, where the number of searching angle in each subarray is 180. It is apparently seen in Figure 9 that

computation time of the other methods presented higher growth rates than the proposed methods. This is because the peak searching algorithm execution time is potentially high, and almost all selected methods require intensive computations by testing the orthogonality of subspace and search space of narrowband sample cross-correlation matrices for all frequency bins, which results in high computation costs. On the contrary, the proposed methods transform all narrowband sample cross-correlation matrices across all frequency bins into a single matrix as shown in Equations (33)–(35), and this matrix contains useful information of source cross-correlation matrices across all frequency bins as $\frac{1}{P} \sum_{i=1}^P \mathbf{S}_{\{f_i, f_i\}}$; in other words, the orthogonality testing of subspace and search space can be done by using the wideband cross-correlation matrix in Equations (33)–(35) instead of narrowband sample cross-correlation matrices for all frequency bins. Therefore, the computational complexity of the proposed methods remarkably less than the other methods, which is confirmed by the test results in Figure 9.

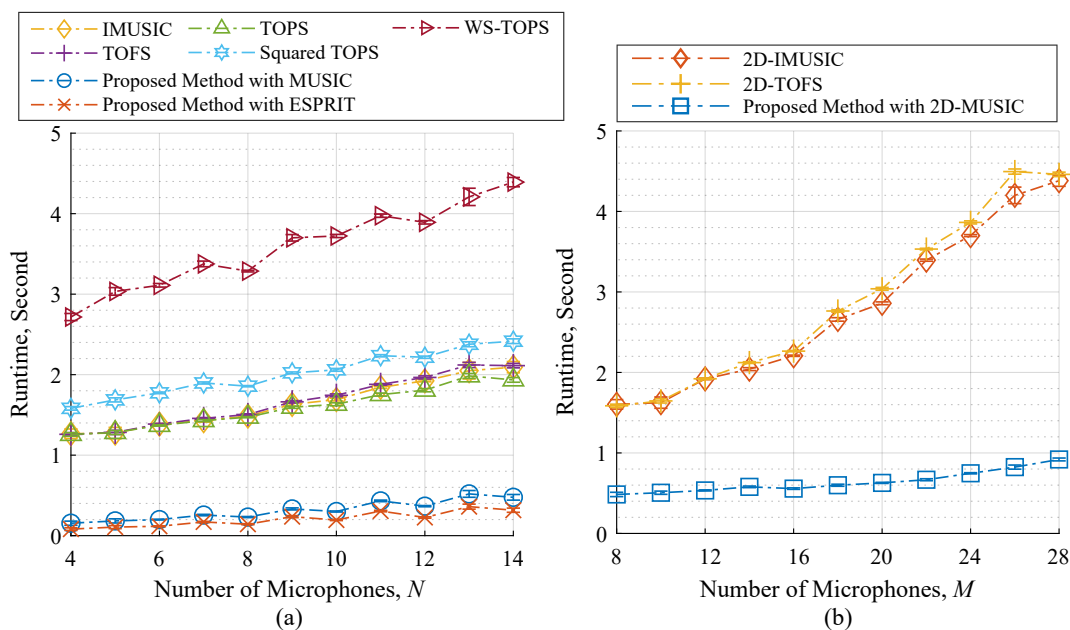


Figure 9. Computational complexities; (a) changing the number of microphone elements each subarray N , and (b) the number of microphone elements including all subarray M where the number of incident sources $K = 3$.

5. Experimental Results

In this section, experiments were carried out to examine the performance of the proposed methods. Experimental parameters were chosen as the previous simulations, except as follows: We used human speakers as sources of the original speech with random sentences. Their speeches were recorded for 20 runs continuously, and each record signal, approximating 1 min long, was cut into 3 s epochs. Structure of the microphone was followed by Figures 1 and 10, and the specifications of the microphone and its recording device were followed on Table 4. The experiment was performed in an indoor meeting room, and its dimensions are shown in Figure 11, where sound pressure level in the meeting room in a normal situation is 46.6 dBA, and the estimated reverberation time is based on RT60 is 219 ms.

Two scenarios are considered: (1) estimating DOA of x and z subarray angles separately, and (2) estimating DOA of x and z subarray angles simultaneously while considering automatic pairing. In case of Experiment 1, the proposed methods are the modified MUSIC in Equation (45) and ESPRIT in Equations (50)–(52), comparing with the following methods: IMUSIC [31], TOFS [32], TOPS [33], Squared-TOPS [34], WS-TOPS [35]. In case of Experiment 2, the proposed method is the modified 2D-MUSIC in Equation (44), comparing with 2D-IMUSIC [31], and 2D-TOFS [32].

Table 4. System specification

Hardware Type/Parameter	Specification/Value
Audio Interface	Roland® Octa-capture (UA-1010)
Sampling Frequency	48,000 Hz
Microphone Name	Behringer® C-2 studio condenser microphone
Number of Microphones	8
Pickup Patterns	Cardioid (8.9 mV/Pa; 20–20,000 Hz)
Diaphragm Diameter	16 mm
Equivalent Noise Level	19.0 dBA (IEC 651)
SNR Ratio	75 dB
Microphone Structure	L-shaped Array
Spacing of Microphone	9 cm

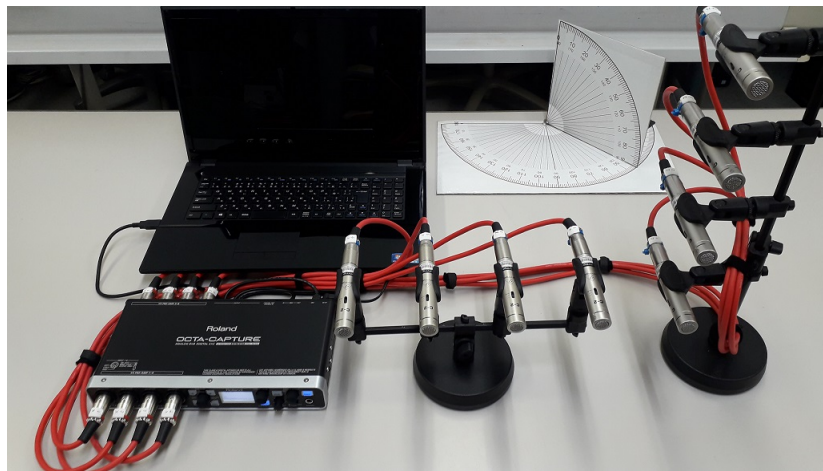


Figure 10. Photograph of the microphone array system.

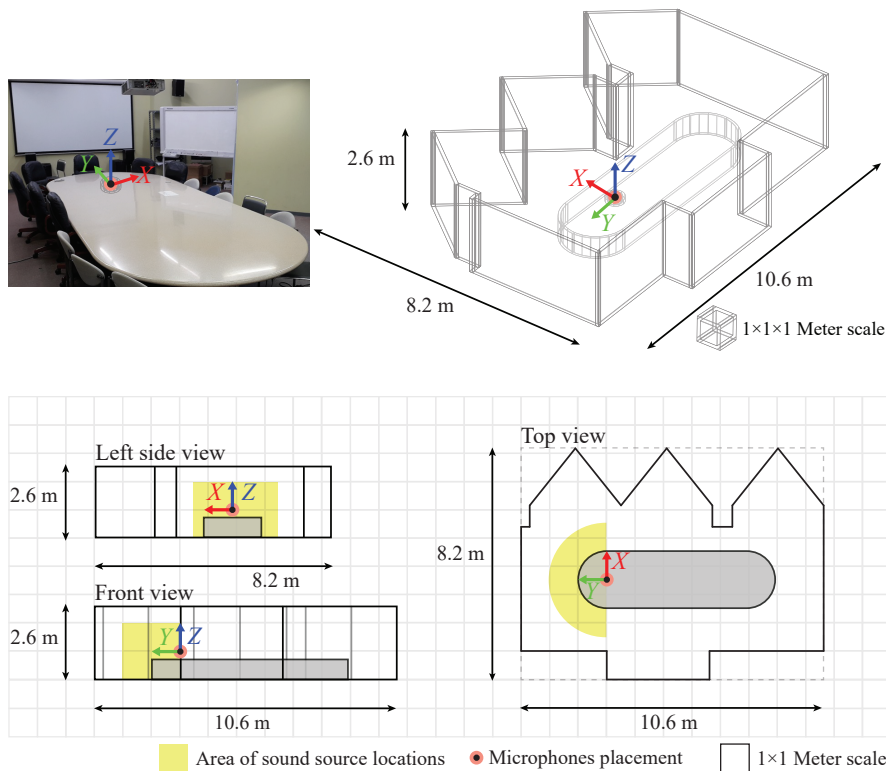


Figure 11. Photograph of the experimental environment, floor plan and the room dimensions.

Tables 5 and 6 showed performance comparisons of the selected methods and the proposed method in term of RMSE over the range of source number, where Table 5 is for Experiment 1, and Table 6 is for Experiment 2. The boldfaced results highlight the optimal minimum RMSE in each problem. As highlighted in Table 5, the performance of IMUSIC exhibited the lowest RMSE when a single source was used, but the performance of the other methods including the proposed methods also exhibited similarly low RMSE in an acceptable error range. When the two sources are performed, the performance of TOPS, Squared-TOPS and WS-TOPS are directly dominated, whereas IMUSIC, TOFS and the proposed methods are slightly dominated, but still maintained sufficiently good performance. When the incident sources are increasing to three, we clearly see that the performance of IMUSIC, TOFS, TOPS, Squared-TOPS and WS-TOPS are significantly dominated by the number of incident sources, because those methods require the number of sensor elements to be much more higher than the number of sources to achieve reasonably good results, which can be verified by referring to the simulation results in Section 4 and Figures 4 and 5. The proposed methods, however, are able estimate the DOA of three sources effectively and better than the selected methods. The reason is that the proposed methods focus on the subspace across all frequency bins simultaneously instead of focusing each frequency band individually, which is stated in Section 3.2. In case of Experiment 2 in Table 6, the experiment results indicate that the proposed method with 2D-MUSIC exhibit extremely similar overall performances to 2D-IMUSIC and 2D-TOFS. As already stated in Section 4.5, the computational complexity of the proposed method is definitely lower than 2D-IMUSIC and 2D-TOFS by the fact that those methods check the orthogonality of subspace and search space of narrowband sample cross-correlation matrices for all frequency bins, resulting in very high computation requirement. The proposed method tests the orthogonality of subspace and search space by using the wideband sample cross-correlation matrix in Equation (33) instead of using the subspace of narrowband sample cross-correlation matrices for all frequency bins, but it is sufficient to exhibit significant effects as well as using the subspace of narrowband sample cross-correlation matrices for all frequency bins. In the end, the experimental results from Tables 5 and 6 are able to provide evidence that the proposed methods have better estimating performance than other methods with respect to the number of incident sources.

Table 5. Performance evaluation on Experiment 1. The boldfaced results highlight the optimal minimum RMSE.

Incident Sources			RMSE of DOAs (Degree)						
Number	Position	Angle (Degree)	IMUSIC	TOFS	TOPS	Squared TOPS	WS-TOPS	Proposed Method with MUSIC	Proposed Method with ESPRIT
1	ϕ_1	96	0.3050	0.2050	1.0950	1.3350	0.5600	0.7750	0.7074
	θ_1	86	0.5400	1.2600	1.2750	2.0150	0.6850	0.5700	0.6915
	Average		0.4225	0.7325	1.1850	1.6750	0.6225	0.6725	0.6995
2	ϕ_1	65	1.1857	1.7286	20.0143	28.5857	37.8714	1.5000	2.0284
	θ_1	150	9.6000	6.6857	26.3571	39.7857	88.2000	8.8143	8.6800
	ϕ_2	55	1.0714	1.6857	22.2571	19.4000	32.2429	2.9714	3.8695
	θ_2	100	8.3714	8.3857	5.0143	6.7857	60.2286	6.6714	3.1630
Average		5.0571	4.6214	18.4107	23.6393	54.6357	4.9893	4.4353	
3	ϕ_1	58	2.1400	2.3900	46.5500	52.8100	40.9500	3.6600	4.0334
	θ_1	55	55.0000	55.0000	55.0000	55.0000	55.0000	9.4300	4.1057
	ϕ_2	100	1.8400	2.0000	41.5700	62.4000	70.9100	1.8700	2.4554
	θ_2	95	95.0000	83.4200	52.4500	71.4800	95.0000	9.7700	5.8638
	ϕ_3	130	10.9300	11.8900	28.8300	32.2800	95.2400	8.2500	6.9071
θ_3	120	26.9800	25.8400	16.1200	18.0100	91.2800	5.9400	7.3165	
Average		31.9817	30.0900	40.0867	48.6633	74.7300	6.4867	5.1137	

Since the sound source directions are static in Tables 5 and 6, it is necessary to consider moving sound sources for more practical use. In future work, we will extend the proposed method for moving sound sources, and further develop the prototype to support more realistic tasks.

Table 6. Performance evaluation on Experiment 2. The boldfaced results highlight the optimal minimum RMSE.

Incident Sources			RMSE of DOAs (Degree)		
Number	Position	Angle (Degree)	2D-IMUSIC	2D-TOFS	Proposed Method with 2D-MUSIC
1	ϕ_1	96	0.9000	0.9000	0.9000
	θ_1	86	0.4000	1.0500	0.7500
	Average		0.6500	0.9750	0.8250
2	ϕ_1	57	0.9500	1.1500	1.1000
	θ_1	91	1.0500	1.8000	1.7000
	ϕ_2	139	4.9500	5.2000	5.4500
	θ_2	96	3.1500	3.3000	2.0500
	Average		2.5250	2.8625	2.5750
3	ϕ_1	48	0.9500	1.5500	1.9500
	θ_1	86	1.4500	0.8000	2.4500
	ϕ_2	98	0.9000	1.8000	1.1500
	θ_2	95	1.4500	2.1500	2.6000
	ϕ_3	152	2.7000	2.4000	5.9000
	θ_3	95	4.5000	3.9000	1.4500
	Average		1.9917	2.1000	2.5833
4	ϕ_1	100	5.8095	6.5238	3.2857
	θ_1	94	2.4286	2.6190	1.6667
	ϕ_2	51	1.2381	1.0952	2.5714
	θ_2	95	0.5714	0.6667	1.3333
	ϕ_3	134	1.9524	1.8571	3.9524
	θ_3	103	10.0952	10.2857	9.2857
	ϕ_4	153	7.4762	7.8095	7.8571
	θ_4	89	4.7143	4.7143	5.3810
Average		4.2857	4.4464	4.4167	

6. Conclusions

An efficient framework for estimating DOA of wideband sound sources was presented. The issue of transforming multiple narrowband cross-correlation matrices for all frequency bins into a wideband cross-correlation matrix has been addressed successfully by focusing on signal subspace for all frequency bins simultaneously instead of the pairing of temporal and reference frequency as done by the CSS-based methods. A new solution to this problem has been given by performing HOGSVD of the array of novel cross-correlation matrices, where elements in the row and column positions are a sample cross-correlation matrix between received signal and itself on two distinct frequencies. It was shown in the theoretical analysis that the proposed transformation procedure provided the best solution under appropriate constraints, and no longer required any process of DOA preliminary estimation. Subsequently, we provided an alternative to construct the wideband cross-correlation matrix via the proposed transformation procedure, and wideband DOAs were estimated easily using this wideband matrix along with a single scheme of estimating DOAs in any narrowband subspace methods. A major contribution of this paper is that the proposed framework enables cutting-edge studies in the recent narrowband subspace methods to estimate DOA of the wideband sources directly, which results in reducing computational complexity and facilitating the estimation algorithm. We also have performed several examples of using the proposed framework, such as 2D-MUSIC, MUSIC, and ESPRIT method integration with the L-shaped microphone arrays. Furthermore, the simulation and experimental results showed that the fusion methods by using the proposed framework exhibited

especially effective performance compared to other wideband DOA estimation methods over a range of SNR with much fewer sensors, high noise and reverberation conditions. We believe that the proposed method represents an efficient way for wideband DOA estimation and would be able to improve wideband DOA estimates not only for acoustic signal processing but also other possible related fields.

Author Contributions: B.S. conceived of the hypothesis, provided the mathematical proof, designed and performed the experiments and wrote the manuscript as part of a PhD project. M.F. supervises the project and contributed to the development of the ideas.

Funding: This work was supported by JSPS KAKENHI Grant Number JP18K12111 and MEXT Grant Number 91506000972.

Acknowledgments: The authors are grateful to Kochi University of Technology for Monthly Support via a grant of Special Scholarship Program over a period of three years.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theorem 1

This appendix provides a detailed derivation of Theorem 1. We begin by considering the cross-correlation matrices in Equation (7). $\mathbf{R}_{\{f_o, f_o\}}$ can be constructed into the EVD form, which is given by

$$\mathbf{R}_{\{f_o, f_o\}} = \mathbf{Q}_{\{f_o\}_s} \mathbf{\Lambda}_{\{f_o\}_s} \mathbf{Q}_{\{f_o\}_s}^H + \mathbf{Q}_{\{f_o\}_n} \mathbf{\Lambda}_{\{f_o\}_n} \mathbf{Q}_{\{f_o\}_n}^H, \quad (\text{A1})$$

where $\mathbf{Q}_{\{f_o\}_s} \in \mathbb{C}^{M \times K}$, $\mathbf{\Lambda}_{\{f_o\}_s} \in \mathbb{R}^{K \times K}$ are the matrix of eigenvectors and diagonal matrix of eigenvalues in signal subspace, and likewise, $\mathbf{Q}_{\{f_o\}_n} \in \mathbb{C}^{M \times M-K}$, $\mathbf{\Lambda}_{\{f_o\}_n} \in \mathbb{R}^{M-K \times M-K}$ are with noise subspace. In case of $\mathbf{R}_{\{f, f_o\}}$, it can be derived by performing SVD, which directly follows from Equation (10). Since $\mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f)$ and $\mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f_o)$ are full rank matrices [36], its remaining components are expressed as follows [48]:

$$\mathbf{U}_{\{f, f_o\}_s} = \mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f) \mathbf{F}_{\{f, f_o\}}^{-1}, \quad \mathbf{V}_{\{f, f_o\}_s} = \mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f_o) \mathbf{G}_{\{f, f_o\}}^{-1}, \quad \mathbf{Q}_{\{f_o\}_s} = \mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f_o) \mathbf{H}_{\{f_o\}}^{-1}, \quad (\text{A2})$$

where

$$\boldsymbol{\Sigma}_{\{f, f_o\}_s} = \mathbf{F}_{\{f, f_o\}} \mathbf{S}_{\{f, f_o\}} \mathbf{G}_{\{f, f_o\}}^H, \quad \mathbf{\Lambda}_{\{f_o\}_s} = \mathbf{H}_{\{f_o\}} \mathbf{S}_{\{f_o, f_o\}} \mathbf{H}_{\{f_o\}}^H, \quad (\text{A3})$$

$\mathbf{F}_{\{f, f_o\}}, \mathbf{G}_{\{f, f_o\}}, \mathbf{H}_{\{f_o\}} \in \mathbb{C}^{K \times K}$ are also full rank and invertible. Note again that $\mathbf{U}_{\{f, f_o\}_s}, \mathbf{V}_{\{f, f_o\}_s}, \mathbf{Q}_{\{f_o\}_s}$ have orthonormal columns [46], hence, it is obvious to see that

$$\mathbf{Q}_{\{f_o\}_s} = \mathbf{V}_{\{f, f_o\}_s} \mathbf{G}_{\{f, f_o\}} \mathbf{H}_{\{f_o\}}^{-1}, \quad (\text{A4})$$

$$\begin{aligned} \mathbf{H}_{\{f_o\}}^H \mathbf{H}_{\{f_o\}} &= \mathbf{A}^H(\boldsymbol{\phi}, \boldsymbol{\theta}, f_o) \mathbf{A}(\boldsymbol{\phi}, \boldsymbol{\theta}, f_o) \\ &= \mathbf{G}_{\{f, f_o\}}^H \mathbf{G}_{\{f, f_o\}}. \end{aligned} \quad (\text{A5})$$

From Equation (A4), we may expect that $\mathbf{G}_{\{f, f_o\}}, \mathbf{H}_{\{f_o\}}$ have unitary property, but it is incorrect when considering Equation (A5). Therefore, a proposition of $\mathbf{G}_{\{f, f_o\}} \mathbf{H}_{\{f_o\}}^{-1}$ have to be identity;

$$\begin{aligned} \mathbf{G}_{\{f, f_o\}} \mathbf{H}_{\{f_o\}}^{-1} &= \mathbf{V}_{\{f, f_o\}_s}^H \mathbf{Q}_{\{f_o\}_s} \\ &= \mathbf{I}_K. \end{aligned} \quad (\text{A6})$$

When considering only the signal subspace, it can be seen from Equations (A4)–(A6) that the right singular vectors of $\mathbf{R}_{\{f, f_o\}}$ and the eigenvectors of $\mathbf{R}_{\{f_o, f_o\}}$ are identical.

Next, we continue to generalize the objective function in Equation (15) by utilizing *Orthogonal Procrustes* (OP) [54], but with some modification (MOP). The objective function in Equation (15) is rederived by

$$\begin{aligned} & \left\| \mathbf{R}_{\{f_o, f_o\}} - \mathbf{T}_{\{f\}} \mathbf{R}_{\{f, f_o\}} \right\|_{\mathbb{F}}^2 \\ &= \text{tr} \left(\mathbf{R}_{\{f_o, f_o\}} \mathbf{R}_{\{f_o, f_o\}}^H \right) + \text{tr} \left(\mathbf{T}_{\{f\}} \mathbf{R}_{\{f, f_o\}} \mathbf{R}_{\{f, f_o\}}^H \mathbf{T}_{\{f\}}^H \right) - 2 \cdot \Re \left(\text{tr} \left(\mathbf{T}_{\{f\}} \mathbf{R}_{\{f, f_o\}} \mathbf{R}_{\{f_o, f_o\}}^H \right) \right), \end{aligned} \quad (\text{A7})$$

where $\Re(a)$ returns the real part of the variable a , $\text{tr}(\mathbf{A})$ trace of the square matrix \mathbf{A} . Considering each expression in Equation (A7), the trace of a product of two square matrices is independent of the orders;

$$\text{tr} \left(\mathbf{R}_{\{f_o, f_o\}} \mathbf{R}_{\{f_o, f_o\}}^H \right) = \text{tr} \left(\mathbf{\Lambda}_{\{f_o\}_s} \mathbf{\Lambda}_{\{f_o\}_s} \right) + \text{tr} \left(\mathbf{\Lambda}_{\{f_o\}_n} \mathbf{\Lambda}_{\{f_o\}_n} \right). \quad (\text{A8})$$

Employing Lemma 1, next, we have

$$\begin{aligned} & \text{tr} \left(\mathbf{T}_{\{f\}} \mathbf{R}_{\{f, f_o\}} \mathbf{R}_{\{f, f_o\}}^H \mathbf{T}_{\{f\}}^H \right) \\ &= \text{tr} \left(\mathbf{\Sigma}_{\{f, f_o\}_s} \mathbf{\Sigma}_{\{f, f_o\}_s} \right) + \text{tr} \left(\mathbf{\Sigma}_{\{f, f_o\}_n} \mathbf{\Sigma}_{\{f, f_o\}_n} \mathbf{U}_{\{f, f_o\}_n}^H \mathbf{T}_{\{f\}}^H \mathbf{T}_{\{f\}} \mathbf{U}_{\{f, f_o\}_n} \right), \end{aligned} \quad (\text{A9})$$

From Equation (A6), we finally have

$$\begin{aligned} & \text{tr} \left(\mathbf{T}_{\{f\}} \mathbf{R}_{\{f, f_o\}} \mathbf{R}_{\{f_o, f_o\}}^H \right) \\ &= \text{tr} \left(\mathbf{\Sigma}_{\{f, f_o\}_s} \mathbf{\Lambda}_{\{f_o\}_s} \mathbf{Q}_{\{f_o\}_s}^H \mathbf{T}_{\{f\}} \mathbf{U}_{\{f, f_o\}_s} \right) + \text{tr} \left(\mathbf{\Sigma}_{\{f, f_o\}_n} \mathbf{\Lambda}_{\{f_o\}_n} \mathbf{Q}_{\{f_o\}_n}^H \mathbf{T}_{\{f\}} \mathbf{U}_{\{f, f_o\}_n} \right) \end{aligned} \quad (\text{A10})$$

Substituting Equations (A8)–(A10) into Equation (A7), the objective function is simplified as

$$\begin{aligned} \left\| \mathbf{R}_{\{f_o, f_o\}} - \mathbf{T}_{\{f\}} \mathbf{R}_{\{f, f_o\}} \right\|_{\mathbb{F}}^2 &= \text{tr} \left(\mathbf{\Lambda}_{\{f_o\}_s} \mathbf{\Lambda}_{\{f_o\}_s} \right) + \text{tr} \left(\mathbf{\Lambda}_{\{f_o\}_n} \mathbf{\Lambda}_{\{f_o\}_n} \right) + \text{tr} \left(\mathbf{\Sigma}_{\{f, f_o\}_s} \mathbf{\Sigma}_{\{f, f_o\}_s} \right) \\ &+ \text{tr} \left(\mathbf{\Sigma}_{\{f, f_o\}_n} \mathbf{\Sigma}_{\{f, f_o\}_n} \mathbf{U}_{\{f, f_o\}_n}^H \mathbf{T}_{\{f\}}^H \mathbf{T}_{\{f\}} \mathbf{U}_{\{f, f_o\}_n} \right) \\ &- 2 \cdot \Re \left(\text{tr} \left(\mathbf{\Sigma}_{\{f, f_o\}_s} \mathbf{\Lambda}_{\{f_o\}_s} \mathbf{Q}_{\{f_o\}_s}^H \mathbf{T}_{\{f\}} \mathbf{U}_{\{f, f_o\}_s} \right) \right) \\ &- 2 \cdot \Re \left(\text{tr} \left(\mathbf{\Sigma}_{\{f, f_o\}_n} \mathbf{\Lambda}_{\{f_o\}_n} \mathbf{Q}_{\{f_o\}_n}^H \mathbf{T}_{\{f\}} \mathbf{U}_{\{f, f_o\}_n} \right) \right). \end{aligned} \quad (\text{A11})$$

Three expressions of $\text{tr} \left(\mathbf{\Lambda}_{\{f_o\}_s} \mathbf{\Lambda}_{\{f_o\}_s} \right)$, $\text{tr} \left(\mathbf{\Lambda}_{\{f_o\}_n} \mathbf{\Lambda}_{\{f_o\}_n} \right)$, $\text{tr} \left(\mathbf{\Sigma}_{\{f, f_o\}_s} \mathbf{\Sigma}_{\{f, f_o\}_s} \right)$ are completely isolated from $\mathbf{T}_{\{f\}}$. Therefore, the optimization problem is redefined as

$$\begin{aligned} & \text{minimize}_{\mathbf{T}_{\{f\}}} \quad \text{tr} \left(\mathbf{\Sigma}_{\{f, f_o\}_n} \mathbf{\Sigma}_{\{f, f_o\}_n} \mathbf{U}_{\{f, f_o\}_n}^H \mathbf{T}_{\{f\}}^H \mathbf{T}_{\{f\}} \mathbf{U}_{\{f, f_o\}_n} \right) \\ & \quad - 2 \cdot \Re \left(\text{tr} \left(\mathbf{\Sigma}_{\{f, f_o\}_s} \mathbf{\Lambda}_{\{f_o\}_s} \mathbf{Q}_{\{f_o\}_s}^H \mathbf{T}_{\{f\}} \mathbf{U}_{\{f, f_o\}_s} \right) \right) \\ & \quad - 2 \cdot \Re \left(\text{tr} \left(\mathbf{\Sigma}_{\{f, f_o\}_n} \mathbf{\Lambda}_{\{f_o\}_n} \mathbf{Q}_{\{f_o\}_n}^H \mathbf{T}_{\{f\}} \mathbf{U}_{\{f, f_o\}_n} \right) \right) \\ & \text{subject to} \quad \sum_{k=1}^K \sigma_k^2 \left(\mathbf{T}_{\{f\}} \mathbf{R}_{\{f, f_o\}} \right) = \sum_{k=1}^K \sigma_k^2 \left(\mathbf{R}_{\{f, f_o\}} \right). \end{aligned} \quad (\text{A12})$$

Now there are two possible cases which we need to consider. The first case is when the $M - K$ smallest singular values of $\mathbf{R}_{\{f, f_o\}}$ are close to zeros; the other is when some of the $M - K$ smallest singular values of $\mathbf{R}_{\{f, f_o\}}$ are more than zeros.

Case 1: Assume that all the $M - K$ smallest singular values of $\mathbf{R}_{\{f,fo\}}$ are close to zeros, we have

$$\begin{aligned} & \underset{\mathbf{T}_{\{f\}}}{\text{maximize}} \quad 2 \cdot \Re \left(\text{tr} \left(\boldsymbol{\Sigma}_{\{f,fo\}_s} \boldsymbol{\Lambda}_{\{fo\}_s} \mathbf{Q}_{\{fo\}_s}^H \mathbf{T}_{\{f\}} \mathbf{U}_{\{f,fo\}_s} \right) \right) \\ & \text{subject to} \quad \sum_{k=1}^K \sigma_k^2 \left(\mathbf{T}_{\{f\}} \mathbf{R}_{\{f,fo\}} \right) = \sum_{k=1}^K \sigma_k^2 \left(\mathbf{R}_{\{f,fo\}} \right), \quad \sum_{m=K+1}^M \sigma_m^2 \left(\mathbf{R}_{\{f,fo\}} \right) = 0. \end{aligned} \quad (\text{A13})$$

Using the proposition of Equation (A6) and employing Lemma 1, two possible solutions to reach the maximum point of Equation (A13) can be found. The first solution is given by

$$\mathbf{T}_{\{f\}_{\text{MOP}}} = \mathbf{Q}_{\{fo\}_s} \boldsymbol{\Omega}_{\{f\}_{\text{MOP}}} \mathbf{U}_{\{f,fo\}_s}^\dagger, \quad (\text{A14})$$

where orthonormal columns has not yet been defined on $\mathbf{U}_{\{f,fo\}_s}$, and the second solution is given by

$$\mathbf{T}_{\{f\}_{\text{OP}}} = \mathbf{Q}_{\{fo\}_s} \boldsymbol{\Omega}_{\{f\}_{\text{OP}}} \mathbf{U}_{\{f,fo\}_s}^H, \quad (\text{A15})$$

where $\mathbf{U}_{\{f,fo\}_s}$ has orthonormal columns. Note that the subscript \dagger denotes the pseudo-inverse. When the constraints in Equation (A13) are imposed into Equations (A15) and (A14), we can have that $\boldsymbol{\Omega}_{\{f\}_{\text{MOP}}} = \mathbf{I}_K$, $\boldsymbol{\Omega}_{\{f\}_{\text{OP}}} = \mathbf{I}_K$, and the maximum is achieved;

$$\bar{q}_{\{f\}_{\text{case 1}}} = 2 \cdot \text{tr} \left(\boldsymbol{\Sigma}_{\{f,fo\}_s} \boldsymbol{\Lambda}_{\{fo\}_s} \right). \quad (\text{A16})$$

Case 2: Assume that some of the $M - K$ smallest singular values of $\mathbf{R}_{\{f,fo\}}$ are more than zeros, the best solution of Equation (A12) can be given the same as Equation (A15), and its minimum is equal to Equation (A16);

$$\underline{q}_{\{f\}_{\text{case 2OP}}} = -2 \cdot \text{tr} \left(\boldsymbol{\Sigma}_{\{f,fo\}_s} \boldsymbol{\Lambda}_{\{fo\}_s} \right). \quad (\text{A17})$$

On the contrary, when using Equation (A14) in Equation (A12), the minimum of cost function is remained by

$$\underline{q}_{\{f\}_{\text{case 2MOP}}} = \text{tr} \left(\boldsymbol{\Sigma}_{\{f,fo\}_n} \boldsymbol{\Sigma}_{\{f,fo\}_n} \left(\mathbf{U}_{\{f,fo\}_s}^\dagger \mathbf{U}_{\{f,fo\}_n} \right)^H \mathbf{U}_{\{f,fo\}_s}^\dagger \mathbf{U}_{\{f,fo\}_n} \right) - 2 \cdot \text{tr} \left(\boldsymbol{\Sigma}_{\{f,fo\}_s} \boldsymbol{\Lambda}_{\{fo\}_s} \right). \quad (\text{A18})$$

Using the solution of Equation (A14) rather than Equation (A15) allows us to relax the error constraint in the hope of arriving at a reduction in the computation of HOGSVD (For details, see Section 3.2), but this is still sufficient for estimating $\mathbf{T}_{\{f\}}$ without loss of generality; the squares of $M - K$ smallest singular values of $\mathbf{R}_{\{f,fo\}}$ are very close to zeros, so we can assume that $\boldsymbol{\Sigma}_{\{f,fo\}_n}^2 \approx \mathbf{O}_{M-K \times M-K}$. Remark that error of the transformation remains consistent with the following equation;

$$\begin{aligned} \varepsilon_{\{f\}_{\text{MOP}}} &= \left| 2 \cdot \Re \left(\text{tr} \left(\boldsymbol{\Sigma}_{\{f,fo\}_s} \boldsymbol{\Lambda}_{\{fo\}_s} \left(\boldsymbol{\Omega}_{\{f\}} - \mathbf{I}_K \right) \right) \right) \right. \\ & \quad \left. + \text{tr} \left(\boldsymbol{\Sigma}_{\{f,fo\}_n}^2 \left(\mathbf{U}_{\{f,fo\}_s}^\dagger \mathbf{U}_{\{f,fo\}_n} \right)^H \mathbf{U}_{\{f,fo\}_s}^\dagger \mathbf{U}_{\{f,fo\}_n} \right) \right|. \end{aligned} \quad (\text{A19})$$

To further reduce a computational burden caused by performing SVD of $\mathbf{R}_{\{f,fo\}}$ and EVD of $\mathbf{R}_{\{fo,fo\}}$, we reinitialize the cross-correlation matrix as

$$\begin{aligned} & \mathbf{R}_{\{f,fo\}} \mathbf{R}_{\{fo,fo\}}^H \\ &= \left(\mathbf{U}_{\{f,fo\}_s} \boldsymbol{\Sigma}_{\{f,fo\}_s} \mathbf{V}_{\{f,fo\}_s}^H + \mathbf{U}_{\{f,fo\}_n} \boldsymbol{\Sigma}_{\{f,fo\}_n} \mathbf{V}_{\{f,fo\}_n}^H \right) \left(\mathbf{Q}_{\{fo\}_s} \boldsymbol{\Lambda}_{\{fo\}_s} \mathbf{Q}_{\{fo\}_s}^H + \mathbf{Q}_{\{fo\}_n} \boldsymbol{\Lambda}_{\{fo\}_n} \mathbf{Q}_{\{fo\}_n}^H \right)^H \\ &= \mathbf{U}_{\{f,fo\}_s} \boldsymbol{\Sigma}_{\{f,fo\}_s} \boldsymbol{\Lambda}_{\{fo\}_s} \mathbf{Q}_{\{fo\}_s}^H + \mathbf{U}_{\{f,fo\}_n} \boldsymbol{\Sigma}_{\{f,fo\}_n} \boldsymbol{\Lambda}_{\{fo\}_n} \mathbf{Q}_{\{fo\}_n}^H \end{aligned} \quad (\text{A20})$$

which is possible to reduce the computation by performing single SVD operation on $\mathbf{R}_{\{f, f_0\}} \mathbf{R}_{\{f_0, f_0\}}^H$.

References

- Haykin, S.; Liu, K.R. *Handbook on Array Processing and Sensor Networks*; Wiley: Hoboken, NJ, USA, 2010. [[CrossRef](#)]
- Zekavat, R.; Buehrer, R.M. *Handbook of Position Location: Theory, Practice and Advances*, 1st ed.; Wiley: Hoboken, NJ, USA, 2011. [[CrossRef](#)]
- Song, K.; Liu, Q.; Wang, Q. Olfaction and Hearing Based Mobile Robot Navigation for Odor/Sound Source Search. *Sensors* **2011**, *11*, 2129–2154. [[CrossRef](#)] [[PubMed](#)]
- Velasco, J.; Pizarro, D.; Macias-Guarasa, J. Source Localization with Acoustic Sensor Arrays Using Generative Model Based Fitting with Sparse Constraints. *Sensors* **2012**, *12*, 13781–13812. [[CrossRef](#)] [[PubMed](#)]
- Tiete, J.; Domínguez, F.; Silva, B.D.; Segers, L.; Steenhaut, K.; Touhafi, A. SoundCompass: A Distributed MEMS Microphone Array-Based Sensor for Sound Source Localization. *Sensors* **2014**, *14*, 1918–1949. [[CrossRef](#)] [[PubMed](#)]
- Clark, B.; Flint, J.A. Acoustical Direction Finding with Time-Modulated Arrays. *Sensors* **2016**, *16*, 2107. [[CrossRef](#)] [[PubMed](#)]
- Hoshiba, K.; Washizaki, K.; Wakabayashi, M.; Ishiki, T.; Kumon, M.; Bando, Y.; Gabriel, D.; Nakadai, K.; Okuno, H.G. Design of UAV-Embedded Microphone Array System for Sound Source Localization in Outdoor Environments. *Sensors* **2017**, *17*, 2535. [[CrossRef](#)] [[PubMed](#)]
- Liu, H.; Li, B.; Yuan, X.; Zhou, Q.; Huang, J. A Robust Real Time Direction-of-Arrival Estimation Method for Sequential Movement Events of Vehicles. *Sensors* **2018**, *18*, 992. [[CrossRef](#)] [[PubMed](#)]
- Knapp, C.; Carter, G. The generalized correlation method for estimation of time delay. *IEEE Trans. Acoust. Speech Signal Process.* **1976**, *24*, 320–327. [[CrossRef](#)]
- Sawada, H.; Mukai, R.; Araki, S.; Makino, S. A robust and precise method for solving the permutation problem of frequency-domain blind source separation. *IEEE Trans. Speech Audio Process.* **2004**, *12*, 530–538. [[CrossRef](#)]
- Yokoi, K.; Hamada, N. ICA-Based Separation and DOA Estimation of Analog Modulated Signals in Multipath Environment. *IEICE Trans. Commun.* **2005**, *88-B*, 4246–4249. [[CrossRef](#)]
- Schmidt, R. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.* **1986**, *34*, 276–280. [[CrossRef](#)]
- Roy, R.; Kailath, T. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 984–995. [[CrossRef](#)]
- Marcos, S.; Marsal, A.; Benidir, M. Performances analysis of the propagator method for source bearing estimation. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-94), Adelaide, SA, Australia, 19–22 April 1994; pp. IV/237–IV/240. [[CrossRef](#)]
- Marcos, S.; Marsal, A.; Benidir, M. The propagator method for source bearing estimation. *Signal Process.* **1995**, *42*, 121–138. [[CrossRef](#)]
- Hua, Y.; Sarkar, T.K.; Weiner, D.D. An L-shaped array for estimating 2-D directions of wave arrival. *IEEE Trans. Antennas Propag.* **1991**, *39*, 143–146. [[CrossRef](#)]
- Porozantidou, M.G.; Chryssomallis, M.T. Azimuth and elevation angles estimation using 2-D MUSIC algorithm with an L-shape antenna. In Proceedings of the 2010 IEEE Antennas and Propagation Society International Symposium, Toronto, ON, Canada, 11–17 July 2010; pp. 1–4. [[CrossRef](#)]
- Wang, G.; Xin, J.; Zheng, N.; Sano, A. Computationally Efficient Subspace-Based Method for Two-Dimensional Direction Estimation With L-Shaped Array. *IEEE Trans. Signal Process.* **2011**, *59*, 3197–3212. [[CrossRef](#)]
- Nie, X.; Wei, P. Array Aperture Extension Algorithm for 2-D DOA Estimation with L-Shaped Array. *Progress Electromagn. Res. Lett.* **2015**, *52*, 63–69. [[CrossRef](#)]
- Tayem, N. Azimuth/Elevation Directional Finding with Automatic Pair Matching. *Int. J. Antennas Propag.* **2016**, *2016*, 5063450. [[CrossRef](#)]
- Wang, Q.; Yang, H.; Chen, H.; Dong, Y.; Wang, L. A Low-Complexity Method for Two-Dimensional Direction-of-Arrival Estimation Using an L-Shaped Array. *Sensors* **2017**, *17*, 190. [[CrossRef](#)]

22. Li, J.; Jiang, D. Joint Elevation and Azimuth Angles Estimation for L-Shaped Array. *IEEE Antennas Wirel. Propag. Lett.* **2017**, *16*, 453–456. [[CrossRef](#)]
23. Dong, Y.Y.; Chang, X. Computationally Efficient 2D DOA Estimation for L-Shaped Array with Unknown Mutual Coupling. *Math. Probl. Eng.* **2018**, *2018*, 1–9. [[CrossRef](#)]
24. Hsu, K.C.; Kiang, J.F. Joint Estimation of DOA and Frequency of Multiple Sources with Orthogonal Coprime Arrays. *Sensors* **2019**, *19*, 335. [[CrossRef](#)]
25. Wu, T.; Deng, Z.; Li, Y.; Li, Z.; Huang, Y. Estimation of Two-Dimensional Non-Symmetric Incoherently Distributed Source with L-Shape Arrays. *Sensors* **2019**, *19*, 1226. [[CrossRef](#)] [[PubMed](#)]
26. Gao, X.; Hao, X.; Li, P.; Li, G. An Improved Two-Dimensional Direction-of-Arrival Estimation Algorithm for L-Shaped Nested Arrays with Small Sample Sizes. *Sensors* **2019**, *19*, 2176. [[CrossRef](#)] [[PubMed](#)]
27. Omer, M.; Quadeer, A.A.; Al-Naffouri, T.Y.; Sharawi, M.S. An L-shaped microphone array configuration for impulsive acoustic source localization in 2-D using orthogonal clustering based time delay estimation. In Proceedings of the 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA), Sharjah, UAE, 12–14 February 2013; pp. 1–6. [[CrossRef](#)]
28. Wajid, M.; Kumar, A.; Bahl, R. Direction-of-arrival estimation algorithms using single acoustic vector-sensor. In Proceedings of the International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT), Aligarh, India, 24–26 November 2017; pp. 84–88. [[CrossRef](#)]
29. Sugimoyo, Y.; Miyabe, S.; Yamada, T.; Makino, S.; Juang, B.H. An Extension of MUSIC Exploiting Higher-Order Moments via Nonlinear Mapping. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **2016**, *E99.A*, 1152–1162. [[CrossRef](#)]
30. Suksiri, B.; Fukumoto, M. Multiple Frequency and Source Angle Estimation by Gaussian Mixture Model with Modified Microphone Array Data Model. *J. Signal Process.* **2017**, *21*, 163–166. [[CrossRef](#)]
31. Su, G.; Morf, M. The signal subspace approach for multiple wide-band emitter location. *IEEE Trans. Acoust. Speech Signal Process.* **1983**, *31*, 1502–1522. [[CrossRef](#)]
32. Yu, H.; Liu, J.; Huang, Z.; Zhou, Y.; Xu, X. A New Method for Wideband DOA Estimation. In Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing, Shanghai, China, 21–25 September 2007; pp. 598–601. [[CrossRef](#)]
33. Yoon, Y.S.; Kaplan, L.M.; McClellan, J.H. TOPS: New DOA estimator for wideband signals. *IEEE Trans. Signal Process.* **2006**, *54*, 1977–1989. [[CrossRef](#)]
34. Okane, K.; Ohtsuki, T. Resolution Improvement of Wideband Direction-of-Arrival Estimation “Squared-TOPS”. In Proceedings of the IEEE International Conference on Communications, Cape Town, South Africa, 23–27 May 2010; pp. 1–5. [[CrossRef](#)]
35. Hirota, H.; Tomoaki, O. DOA estimation for wideband signals based on weighted Squared TOPS. *EURASIP J. Wirel. Commun. Netw.* **2016**, *2016*, 243. [[CrossRef](#)]
36. Wang, H.; Kaveh, M. Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources. *IEEE Trans. Acoust. Speech Signal Process.* **1985**, *33*, 823–831. [[CrossRef](#)]
37. Hung, H.; Kaveh, M. Focussing matrices for coherent signal-subspace processing. *IEEE Trans. Acoust. Speech Signal Process.* **1988**, *36*, 1272–1281. [[CrossRef](#)]
38. Valaee, S.; Kabal, P. Wideband array processing using a two-sided correlation transformation. *IEEE Trans. Signal Process.* **1995**, *43*, 160–172. [[CrossRef](#)]
39. Valaee, S.; Champagne, B.; Kabal, P. Localization of wideband signals using least-squares and total least-squares approaches. *IEEE Trans. Signal Process.* **1999**, *47*, 1213–1222. [[CrossRef](#)]
40. Suksiri, B.; Fukumoto, M. A Computationally Efficient Wideband Direction-of-Arrival Estimation Method for L-Shaped Microphone Arrays. In Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 27–30 May 2018; pp. 1–5. [[CrossRef](#)]
41. Abdelbari, A. Direction of Arrival Estimation of Wideband RF Sources. Ph.D. Thesis, Near East University, Nicosia, Cyprus, 2018. [[CrossRef](#)]
42. Ponnappalli, S.P.; Saunders, M.A.; Van Loan, C.F.; Alter, O. A Higher-Order Generalized Singular Value Decomposition for Comparison of Global mRNA Expression from Multiple Organisms. *PLoS ONE* **2011**, *6*, 1–11. [[CrossRef](#)] [[PubMed](#)]
43. Xin, J.; Zheng, N.; Sano, A. Simple and Efficient Nonparametric Method for Estimating the Number of Signals Without Eigendecomposition. *IEEE Trans. Signal Process.* **2007**, *55*, 1405–1420. [[CrossRef](#)]

44. Nadler, B. Nonparametric Detection of Signals by Information Theoretic Criteria: Performance Analysis and an Improved Estimator. *IEEE Trans. Signal Process.* **2010**, *58*, 2746–2756. [[CrossRef](#)]
45. Diehl, R. Acoustic and auditory phonetics: The adaptive design of speech sound systems. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **2008**, *363*, 965–978. [[CrossRef](#)] [[PubMed](#)]
46. Van Der Veen, A.; Deprettere, E.F.; Swindlehurst, A.L. Subspace-based signal analysis using singular value decomposition. *Proc. IEEE* **1993**, *81*, 1277–1308. [[CrossRef](#)]
47. Hogben, L. Discrete Mathematics and Its Applications. In *Handbook of Linear Algebra*; CRC Press: Boca Raton, FL, USA, 2006.
48. Naidu, P. *Sensor Array Signal Processing*; Taylor & Francis: Abingdon-on-Thames, UK, 2000.
49. Meyer, C.D. (Ed.) *Matrix Analysis and Applied Linear Algebra*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2000.
50. Horn, R.A.; Johnson, C.R. *Matrix Analysis*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2012. [[CrossRef](#)]
51. Van Loan, C.F. Structured Matrix Computations from Structured Tensors: Lecture 6. In *The Higher-Order Generalized Singular Value Decomposition*; Cornell University: Ithaca, NY, USA, 2015.
52. Wei, Y.; Guo, X. Pair-Matching Method by Signal Covariance Matrices for 2D-DOA Estimation. *IEEE Antennas Wirel. Propag. Lett.* **2014**, *13*, 1199–1202. [[CrossRef](#)]
53. Lehmann, E.A.; Johansson, A.M. Prediction of energy decay in room impulse responses simulated with an image-source model. *J. Acoust. Soc. Am.* **2008**, *124*, 269–277. [[CrossRef](#)] [[PubMed](#)]
54. Gower, J.C.; Dijksterhuis, G.B. *Procrustes Problems*, 1st ed.; Oxford University Press: Oxford, UK, 2004. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).