A study on underdetermined DOA estimation techniques based on higher-order correlation analysis utilizing non-linear mapping

September   2016

Yuya Sugimoto

A study on underdetermined DOA estimation techniques based on higher-order correlation analysis utilizing non-linear mapping

Graduate School of Systems and Information Engineering
University of Tsukuba

September / 2016

Yuya Sugimoto

Abstract

Direction Of Arrival (DOA) estimation is a traditional research subject in array signal processing. Because of the importance of DOA information in many applications, a wide range of studies have already been reported. However, representative high-resolution DOA estimatiors suffer from a notable problem that they are constrained by the signal dimensionality and cannot estimate DOAs in underdetermined cases, where the number of sources exceeds the number of sensors. Although several conventional methods, exploiting higher order cumulants of signals, have been proposed to overcome the problem, they also have another shortcoming for the short-time analysis. Therefore, we set our research goal to propose a novel approach that can estimate DOAs with high resolution even in underdetermined conditions and achieve superior estimation accuracy in short-time analysis.

The work for this study is mainly two-fold. First, we establish concept of our proposed method employing higher-dimensional nonlinear mapping. By increasing signal dimensionality with the mapping, we extend the analysis of MUltiple SIgnal Classification (MUSIC), the representative high-resolution DOA estimator, to the underdetermined DOA estimation. Furthermore, for the higher estimation accuracy in short-time analysis, we propose a class of mapping which enables us to analyze higher order moments of signals. Through the comparisons focusing on the bias-variance tradeoff between moment and cumulant, we demonstrate the superiority of the proposed method for short-time analysis in terms of both estimation accuracy and computational complexity.

Next, for the further improvement of estimation accuracy, we propose an extension of the proposed method. We describe an algorithm that utilizes multiple maps simultaneously for the joint analysis of moment of multiple high orders. Also, we discuss the advantage of this approach and conduct experiments to evaluate its effectiveness. Experimental results reveal that the extended proposed method achieves quite higher estimation accuracy than other conventional methods.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Study background and purpose

Array signal processing, taking an quite important sub-area of signal processing, has been widely used in several fields of science and engineering such as, radar, sonar, wireless communication, radio astronomy, seismology, acoustics and medical imaging [1, 2, 3, 4]. Early development of array signal processing has mainly been made in the fields of wireless communication and radar system among the first half of 20th century. From the second half of the 20th century, with the progress of digital signal processing hardware, those fundamental knowledge conveyed to other fields had been growing up and numerous new developments and contributions are being made nowadays.

Direction Of Arrival (DOA) estimation is a traditional research subject in array signal processing and have mainly been developed for applications in the radio and wireless communication field. Because of the importance of DOA information in many applications, a wide range of studies have already been reported, and these techniques are adopted as a processing step in many sensor systems [5]. In acoustic and speech signal processing field, these techniques have also been introduced. Not only as a straightforward usage for audio scene recognition or source localization, these methods are now essential as a preprocessing for various other applications, such as noise reduction, dereverberation, and source separation [6].

The beginning of DOA estimation techniques was beamformer approach utilizing the time differences between signal arrivals [7, 8], such as generalized cross-correlation (GCC) [9] based on the whitened cross-correlation function. These methods have a low computational cost but inadequate estimation accuracy. While extensions and generalizations of GCC, such as the steered response power (SRP) [10] have been shown to improve the estimation accuracy for cases of more than two microphones, the issue of low estimation resolution remains when the number of sources is large.

1

Over the past few decades, many DOA estimators with high resolution have been proposed. Early contributions were made by introducing constrained conditions of the algorithm for the adaptive array processing [11, 12, 13, 14, 15], and later, various methods based on the eigenvalue analysis have been proposed for higher resolution [16, 17, 18, 19]. Among these methods, MUltiple SIgnal Classification (MUSIC) [18] is the prevalent technique. However, MUSIC has a notable shortcoming that it is restricted by the dimensionality of the covariance matrix. This is because MUSIC requires the residual subspace of the observation to be reserved only for noise. To estimate DOAs of N sources, $M(> N)$ microphones are required and the DOA estimation performance degrades as N approaches M. Thus, a huge microphone array is required for the DOA estimation of a large number of sources. To overcome this issue, several extensions of MUSIC have been proposed. An extension of MUSIC utilizing 4th-order cumulants [20] was developed in the 1990s, and later this approach was generalized as $2q$-MUSIC [21], which exploits the cumulants of arbitrary even orders. Utilizing the increased dimensionality of the matrix composed of cumulants with additivity, these methods can estimate DOAs in underdetermined cases, where the number of sources exceeds the number of microphones, and also improve the resolution of DOA estimation by virtually increasing the signal expressiveness by expanding the nonlinear subspace [22]. However, since the cumulant have a large variance and a complex calculation, these methods still suffer from the problems of estimation accuracy in short-time analysis and slow processing speed, and cannot satisfy the request from many practical applications such as real time source tracking.

Meanwhile in recent years, in the context of acoustic signal processing, wide variety of researches based on the sparse representation of source signals have been proposed to improve performance and to achieve availability for the case which the conventional approaches cannot adopt well. For example, the sparse sound field decomposition [23, 24] to improve accuracy even when the source signals are highly correlated and the sources are in a highly noise environment, and source localization based on sparse representation [25, 26, 27] to achieve higher spatial resolution.

In this study, by utilizing such sparse property of signal referred to as W-disjoint orthogonality [28], that is a condition in which the time-frequency representations of the sources do not overlap, we aims to achieve a underdetermined DOA estimation with high resolution that also has higher reliability for short-time analysis and lower computational complexity than conventional $2q$-MUSIC. The work for this dissertation is mainly two-fold. First, we propose a new DOA estimator based on different approach from $2q$-MUSIC. For scenarios with more noise sources than microphones, a speech enhancement method called complementary beamforming [29] has been proposed, which has been applied to DOA estimation problems [30]. Complementary beamforming has also been extended and explained as the mapping of a signal onto a higher-dimensional space using a kernel function [31]. Since this is a direct way to increase the subspace dimensionality and can easily

change the tendency of the analysis with the class of mapping, the approach to increasing the signal dimensionality in these methods is expected to be particularly effective for MUSIC, in which the dimensionality is critical. Therefore, as the combination method of the higher-dimensional mapping and MUSIC, we propose *mapped MUSIC* [32, 33], an extension of MUSIC for underdetermined DOA estimation that is based on the correlation analysis of the mapping. We describe the estimation algorithm of mapped MUSIC and introduce a class of mapping functions with which mapped MUSIC analyzes the moments of arbitrary even orders. Moreover, we explain the availability of the moment analysis for the underdetermined case exploiting W-disjoint orthogonality of the source signals. After we establish bias-variance tradeoff between moment and cumulant through the discussion of statistical property between those two statistics, we hypothesize the proposed method performs well for short-time analysis, where the influence of the variance is greater. To evaluate the bias-variance tradeoff, we compare the performances of the proposed method and $2q$-MUSIC by using simulated and real data under the assumption that the true number of sources N is given as prior information. Moreover, we demonstrate the computational efficiency of the proposed method and introduce alternative mappings to further reduce the computational complexity for 4th- and 6th-order moment analysis. Experimental results reveal the superior performance of the proposed method for short-time analysis in terms of both estimation accuracy and computational complexity.

Next, we propose the extension of mapped MUSIC based on joint analysis of moments of multiple orders [34] to further improve estimation accuracy. Although use of higher order statistics yields a higher dimensional subspace for the DOA estimation, the result lose robustness due to the greater bias and variance in the higher-order statistics as shown in the first work. Thus the question of a strategy for trade-off between resolution and robustness arises. To explore the possibility of realizing both advantages of the high and low orders of statistics, we evaluate moments of multiple orders simultaneously. This goal was found to be possible and can be accomplished by properly combining low-and-high-ordered moments in a nonlinearly expanded space. Experimental results of speech DOA estimation demonstrate that the mapped MUSIC with the new map, for joint analysis of moments of multiple orders, achieves a higher accuracy than other MUSIC extensions.

The subsequent section describes our contributions in more detail.

1.2   Thesis overview and contributions

The following chapters will proceed as follows:

- Chapter 2 will present problem statement of the DOA estimation of sound sources using microphone array. We show the observation signal model assumed in this dissertation. Subsequently, we describe specific estimation procedure using steering vector.

- Chapter 3 will review several conventional methods. We describe estimation algorithm of delay-and-sum beamformer and its extension SRP as the DOA estimators with no constraint about the number of estimated DOAs. We also describe estimation algorithm of MUSIC, a representative DOA estimator with high resolution, and $2q$-MUSIC, MUSIC extension exploiting higher order cumulant for underdetemined DOA estimation.

- Chapter 4 will propose mapped MUSIC. We describe estimation algorithm of mapped MUSIC based on the correlation analysis of the nonlinear mapping of the signals. Also, we propose the suitable map allow us to estimate moments of arbitrary even order. Through the comparison with $2q$-MUSIC focusing of the difference between moment and cumulant, we confirm mapped MUSIC becomes a superior option for practical short-time analysis.

- Chapter 5 will propose extension of mapped MUSIC. To further improve the DOA estimation accuracy, we propose new mapping function for joint analysis of moments of multiple orders. Experimental results of speech DOA estimation show the efficiency of this method which combines both advantages of the high and low orders of statistics.

- Chapter 6 will conclude this dissertation. We summarize our study and discuss future direction.

Chapter 2

Problem statement

This chapter describes problem statement of DOA estimation scenario with microphone array and sound sources. Section 2.1 describes an aim of DOA estimation assuming sound sources. Section 2.2 shows observation signal model assumed throughout this dissertation. Section 2.3 describes specific procedure of DOA estimation.

2.1 Aim of DOA estimation of sound sources

We assume a recording environment shown in Fig. 2.1. An aim of DOA estimation of sound sources is to identify correct direction of direct sound emitted from target sound source by analyzing the recorded mixture containing direct sound, reflected sound, and sudden noise.
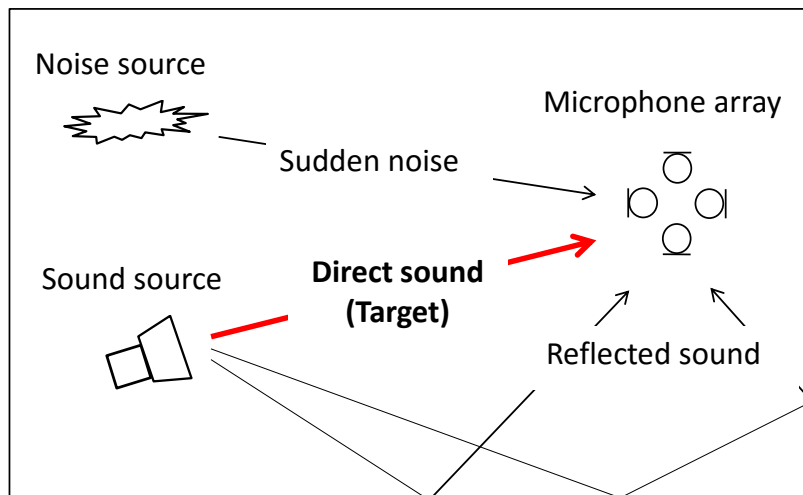
Figure 2.1: An example of room recording.
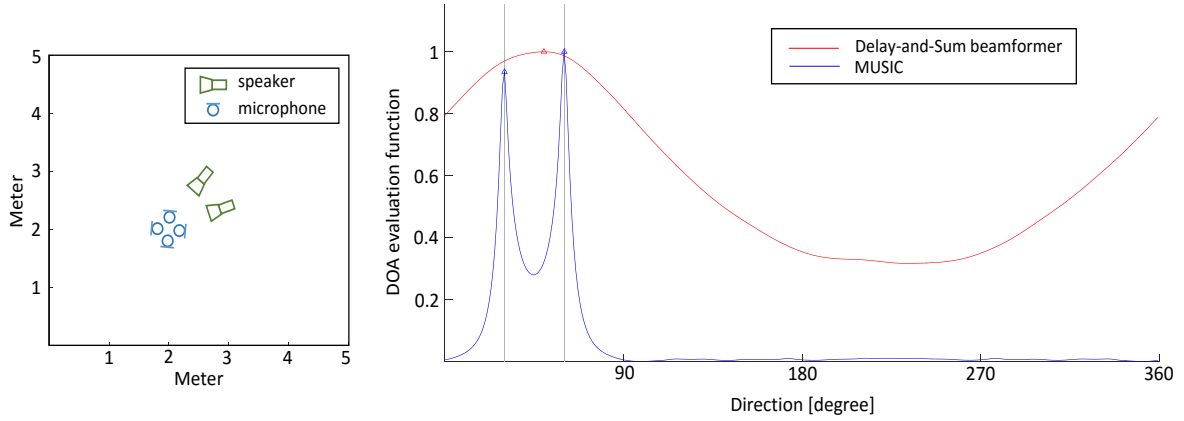
Figure 2.2: A DOA estimation of two sound sources located at the directions of 30 and 60 from the circular microphone array. Left figure shows the locations of the microphones and the sources in 5 m square room. Right figure shows the DOA evaluation function of Delay-and-Sum beamformer and MUSIC defined later. DOAs are estimated as the peak of these plots, triangles denote estimated DOAs, and the true source directions are denoted as vertical lines. MUSIC with high resolution can estimate respective sources correctly, but Delay-and-Sum beamformer estimates single DOAs at wrong direction because of its low resolution.

Figure 2.2 shows the DOA estimation examples. As this figure suggests, the resolution of estimation is essential for the superior estimation performance as well as the the robustness against disturbance, such as reflection and sudden noise. In particular, the resolution has greater importance under the condition when the number of sources becomes large and each source inevitably locate nearby. Although high resolution methods like MUSIC are good option for the case, there still remain the problem how we provide sufficient number of microphones. Overcoming this issue is main objective of this dissertation.

2.2 Observation signal model

Throughout this dissertation, signals are expressed as complex amplitudes. We assume N source signals propagating to an array consisting of M microphones, and the true number of sources N is given as prior information. The signals are statistically independent and the single signal emitted by the nth source, $s_n(\omega, t)$, is a zero-mean super-Gaussian complex random variable. Also, the signals emitted at different times are assumed to be statistically independent. Thus, a noise-free
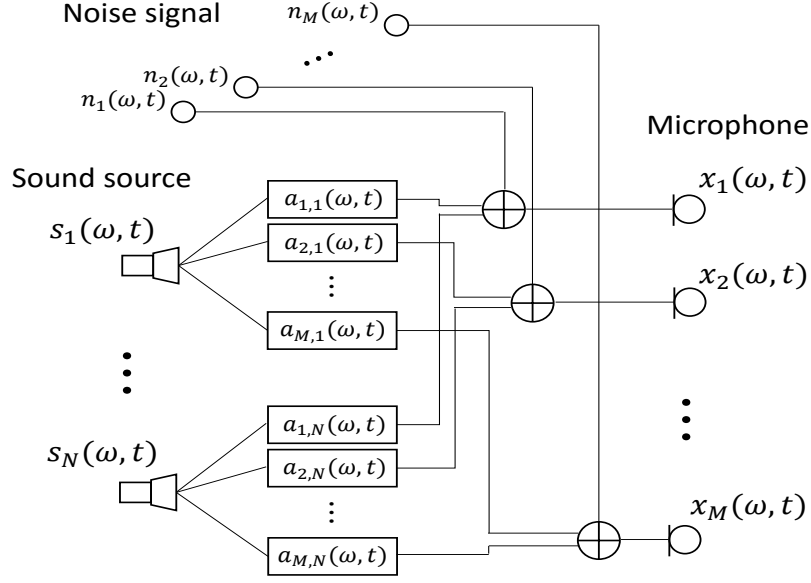
Figure 2.3: Observation signal model assuming N sound sources and M microphones.

vector $\mathbf{z}(\omega, t)$ of the observed signals is assumed to be given by

$$\mathbf{z}(\omega, t) = [z_1(\omega, t), \cdots, z_M(\omega, t)]^T$$
$$= \mathbf{A}(\omega)\mathbf{s}(\omega, t), \tag{2.1}$$
$$\mathbf{A}(\omega) = [\mathbf{a}_1(\omega), \cdots, \mathbf{a}_N(\omega)], \tag{2.2}$$
$$\mathbf{s}(\omega, t) = [s_1(\omega, t), \cdots, s_N(\omega, t)]^T, \tag{2.3}$$

where $\mathbf{a}_i(\omega) = [a_{1,i}(\omega), \cdots, a_{M,i}(\omega)]^T$ is an array manifold vector from the direction of the ith source, which consists of the transfer function from the ith source to the jth microphone $a_{j,i}(\omega, t)$, ω is the angular frequency, $t = 1, \cdots, L$ is the time frame index, and $[\cdot]^T$ denotes the transpose. Furthermore, the noise-corrupted observed vector $\mathbf{x}(\omega, t)$ is given by

$$\mathbf{x}(\omega, t) = [x_1(\omega, t), \cdots, x_M(\omega, t)]^T$$
$$= \mathbf{z}(\omega, t) + \mathbf{n}(\omega, t), \tag{2.4}$$
$$\mathbf{n}(\omega, t) = [n_1(\omega, t), \cdots, n_M(\omega, t)]^T, \tag{2.5}$$

where $n_j(\omega, t)$ is the zero-mean noise signal superimposed on the jth microphone. Figure 2.3 is a diagram of this observing system.

2.3 Problem statement of DOA estimation

In the signal model expressed by Eq. (2.4), each microphone observes a mixture of source signals and noise signals. The problem in this dissertation is the estimation of the DOAs $\theta_1, \cdots, \theta_N$ of the sources $s_1(\omega, t), \cdots, s_N(\omega, t)$ by searching for directions θ with the steering vector $\mathbf{b}(\omega; \theta)$ close to each of the array manifold vectors $\mathbf{a}_1(\omega), ..., \mathbf{a}_N(\omega)$. The steering vector $\mathbf{b}(\omega; \theta)$ is given by

$$\mathbf{b}(\omega; \theta) = \frac{1}{\sqrt{M}} \left[1, \exp(-\jmath \omega \Delta \tau_{12}), \cdots, \exp(-\jmath \omega \Delta \tau_{1M}) \right]^T, \qquad (2.6)$$

where $\Delta \tau_{ij}$ $(i, j = 1, \cdots, M)$ are time differences between ith microphone and jth microphone, given by a signal from direction θ, and $\Delta \tau_{ii}$ is obviously 0.

Note that an estimation of the number of sources is an important aspect of DOA estimation [35, 36]. For the purpose, besides manual thresholding of the eigenvalue, several approaches have been proposed to estimate the number of sources, based on Akaike information criteria [37] or minimum descriptive length criteria [38], and applying the support vector machine [39]. However this dissertation does not contain this discussion. This is because both our proposed mapped MUSIC and conventional $2q$-MUSIC, which is mainly compared in this dissertation, have a side effect to expand signal dimensionality and lose applicability for those approaches in the realistic conditions. Therefore, this dissertation assumes the true number of sources N is given as prior information and just focus on the comparison of DOA estimation using N.

Chapter 3

Conventional methods

This chapter reviews conventional DOA estimators, based on traditional inter-channel correlation and subspace analysis. Note that this chapter partially contains the explanations in time domain. Section 3.1 reviews delay-and-sum beamformer. Section 3.2 reviews generalized cross correlation. Section 3.3 reviews steered response power. Section 3.4 reviews MUSIC. Section 3.5 reviews $2q$-MUSIC.

3.1 Delay-and-sum beamformer

Delay-and-Sum beamformer is the most straightforward approach for the Time Difference Of Arraival (TDOA) estimation based on the analysis of the peak of the inter-channel correlation. Also, since DOA can be uniquely determined from TDOA if we know the array arrangement, many methods based on the concept of delay-and-sum beamformer are also applied for the estimation of DOA.

We assume following Fig. 3.1 where a plane wave is imping on M microphones and observed signals are aligned with delay operators D_i $(i = 1, \cdots, M)$. When all D_i equals to real TDOAs between each channel pair, observed signals $x_1, \cdots, x_M$ become in-phase, and output is maximized. Therefore, delay-and-sum beamformer estimate TDOAs by finding D_i which give maximal output. Although delay-and-sum beamformer is directly represented by cross-correlation function in the continuous time domain, this operation require to estimate respective TDOAs for each channel pair if the array consists of multiple microphones such as Fig. 3.1. Hence, the practical processing is usually conducted in the discrete time-frequency domain in which we can compute correlations of all channel pairs at once and simplify convolution integral into multiplication.

The estimation of delay-and-sum beamformer is formulated as finding the vectors $\boldsymbol{\tau}$ whose entries
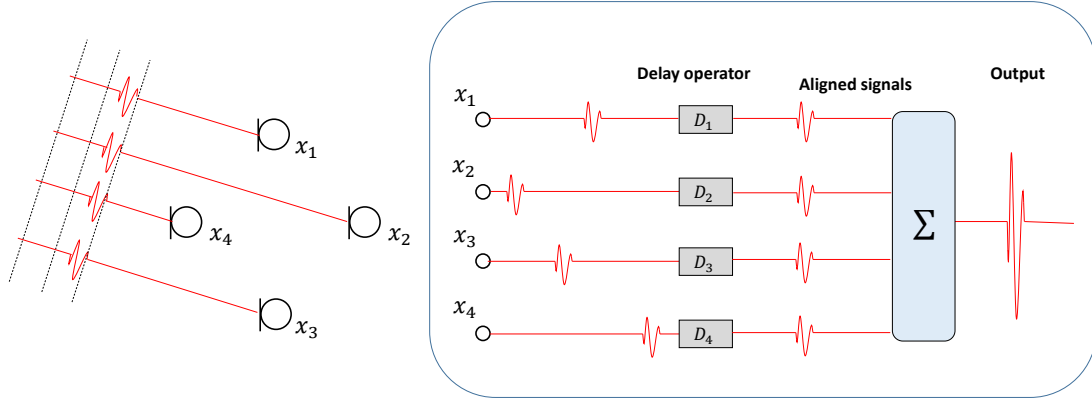
Figure 3.1: A diagram of delay-and-sum beamformer. The left hand figure shows microphone array of $M = 4$. The right hand figure shows output is maximized when all delay operators D_i ($i = 1, \cdots, 4$) equal to real TDOAs $\Delta\tau_{11}, \cdots, \Delta\tau_{14}$ respectively.

are true TDOAs $\Delta\tau_{11}, \cdots, \Delta\tau_{1M}$.

$$\boldsymbol{\tau} = [\Delta\tau_{11}, \cdots, \Delta\tau_{1M}]^T. \tag{3.1}$$

Then, the correct estimation of the vector $\boldsymbol{\tau}$ ought to maximize all cross-correlation functions between each channel pair. Since the delay of the length $\Delta\tau_{ij}$ is expressed by the phase rotation $\exp(-\jmath\omega\Delta\tau_{ij})$ in the time-frequency domain, the formulation can be also expressed as finding filters $\exp(-\jmath\omega\Delta_{1j})$ ($i = 1, \cdots, M$) as follows:

$$\boldsymbol{\tau} \leftarrow \arg\max_{\boldsymbol{\tau}'} f_{\text{DaS}}(\boldsymbol{\tau}'), \tag{3.2}$$

$$f_{\text{DaS}}(\boldsymbol{\tau}') = \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{\omega} G_{x_i x_j}(\omega)\exp(-\jmath\omega\Delta_{ij}'),$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{\omega} G_{x_i x_j}(\omega)\exp\left(-\jmath\omega(-\Delta_{1i}' + \Delta_{1j}')\right), \tag{3.3}$$

$$G_{x_i x_j}(\omega) = E[x_i^*(\omega, t)x_j(\omega, t)], \tag{3.4}$$

$$\boldsymbol{\tau}' = \left[\Delta\tau_{11}', \cdots, \Delta\tau_{1M}'\right]^T \quad \text{subject to } \Delta\tau_{11}' = 0, \tag{3.5}$$

where $E[\cdot]$ denotes the expectation of the argument, $[\cdot]^*$ denotes the complex conjugate, and $G_{x_i x_j}(\omega)$ is the cross spectrum given as the discrete Fourier transform (DFT) of the cross correlation function between ith channel and jth channel.

Furthermore, Eqs. (3.2)–(3.5) to maximize a $f_{\text{DaS}}(\boldsymbol{\tau}')$ can be replaced by the quadratic form using covariance matrix $\mathbf{C}_2(\omega)$ and the steering vector $\mathbf{b}(\omega; \theta)$ defined for DOA estimation in Eq.

10

(2.6), because DFT of possible vector $\boldsymbol{\tau}'$ corresponds to the steering vector of any direction and desired filters $\exp(-\jmath\omega\Delta_{1j})$ $(i = 1, \cdots, M)$ are equivalent to the steering vector $\mathbf{b}(\omega; \theta)$ when direction θ accords with the DOA of target signal. We define the replaced function $\overline{f_{\text{DaS}}(\theta)}$ as the DOA evaluation function of delay-and-sum beamformer which is maximal in a direction θ close to true source DOA.

$$\overline{f_{\text{DaS}}(\theta)} \triangleq \sum_{\omega} f_{\text{DaS}}(\omega; \theta), \tag{3.6}$$

$$f_{\text{DaS}}(\omega; \theta) \triangleq \mathbf{b}^H(\omega; \theta)\mathbf{C}_2(\omega)\mathbf{b}(\omega; \theta), \tag{3.7}$$

$$\mathbf{C}_2(\omega) = E\left[\mathbf{x}(\omega, t)\mathbf{x}^H(\omega, t)\right],$$
$$= \left[G_{x_i x_j}(\omega)\right]_{ij}, \tag{3.8}$$

where $[\cdot]^H$ denotes the complex conjugate transpose, and $[\cdot]_{ij}$ denotes a matrix consisting of the argument in parentheses as its (i, j) entry. Finally, we can estimate DOAs by finding peaks from the function $\overline{f_{\text{DaS}}(\theta)}$.

3.2 Generalized cross correlation

Delay-and-sum beamformer is weak against errors and the peak of the cross correlation $C_{x_1 x_2}$ between two observations, x_1 and x_2, is easily made unclear by the reverberant, noisy recording environments and the colored signals. Generalized Cross Correlation (GCC) is the most representative extension of delay-and-sum beamformer which utilizes whitening filter to ensure robustness against such disturbance. By substituting the cross correlation function $C_{x_1 x_2}$ with its whitened version $\psi_{12} \circledast C_{x_1 x_2}$ ($\circledast$ is the convolution operator) applying the whitening filter ψ_{12}, GCC clarifies the peak and achieves the greatly improved estimation accuracy even in the noisy or reverberant environment.. Although it is not the scope of our study, GCC particularly shows high estimation performance in the scenario of stereo channel [40] for which the latter mentioned methods based on subspace analysis cannot perform well.

The estimation of GCC is an addition of whitening filter into delay-and-sum beamformer and it is formulated according to the following equation in time domain,

$$\Delta\tau_{12} \leftarrow \underset{\Delta\tau'_{12}}{\arg\max} \, \psi_{12}(\Delta\tau'_{12}) \circledast C_{x_1 x_2}(\Delta\tau'_{12}). \tag{3.9}$$

As with delay-and-sum beamformer, the formulation of GCC shown as Eq. (3.9) is expressed in frequency domain according to following equation,

$$\Delta\tau_{12} \leftarrow \underset{\Delta\tau'_{12}}{\arg\max} \sum_{\omega} \Psi_{12}(\omega)G_{x_1 x_2}(\omega)\exp(-\jmath\omega\Delta\tau'_{12}), \tag{3.10}$$

11

where Ψ_{12} is the DFT of ψ_{12}. For the improvement of estimation accuracy, appropriately design of whitening filter is necessary and the filter design is conducted in the frequency domain for Ψ_{12}. The following are examples of specific filters among several filter design methods on the basis of different criteria.

- GCC-SCOT filter

$$\mathbf{\Psi}_{12}(\omega) = \frac{1}{\sqrt{G_{x_1 x_1}(\omega) G_{x_2 x_2}(\omega)}}$$

This filter is robust against noise.

- GCC-PHAT filter

$$\mathbf{\Psi}_{12}(\omega) = \frac{1}{|G_{x_1 x_2}(\omega)|}$$

This filter is robust against the reverberation.

3.3 Steered response power

Steered Response Power (SRP) is a multichannel extension of GCC. Because of its moderate resolution without constraint about the number of target sources, SRP is frequently applied as the reliable and common DOA estimator.

SRP is the whitened version of delay-and-sum beamformer with whitening filter of GCC when $M \geq 3$. Therefore, the estimation of SRP is easily expressed as the combination of Eqs. (3.2)–(3.5) and GCC filter $\Psi_{ij}(\omega)$ for the ith and jth channel pair.

$$\boldsymbol{\tau} \leftarrow \arg \max_{\boldsymbol{\tau}'} f_{\mathrm{SRP}}(\boldsymbol{\tau}'), \tag{3.11}$$

$$f_{\mathrm{SRP}}(\boldsymbol{\tau}') = \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{\omega} \Psi_{ij}(\omega) G_{x_i x_j}(\omega) \exp\left(-\jmath\omega(-\Delta'_{1i} + \Delta'_{1j})\right). \tag{3.12}$$

$$\boldsymbol{\tau}' = \left[\Delta\tau'_{11}, \cdots, \Delta\tau'_{1M}\right]^{T} \quad \text{subject to } \Delta\tau'_{11} = 0. \tag{3.13}$$

Also, it can be expressed by the following quadratic form of the generalized correlation matrix
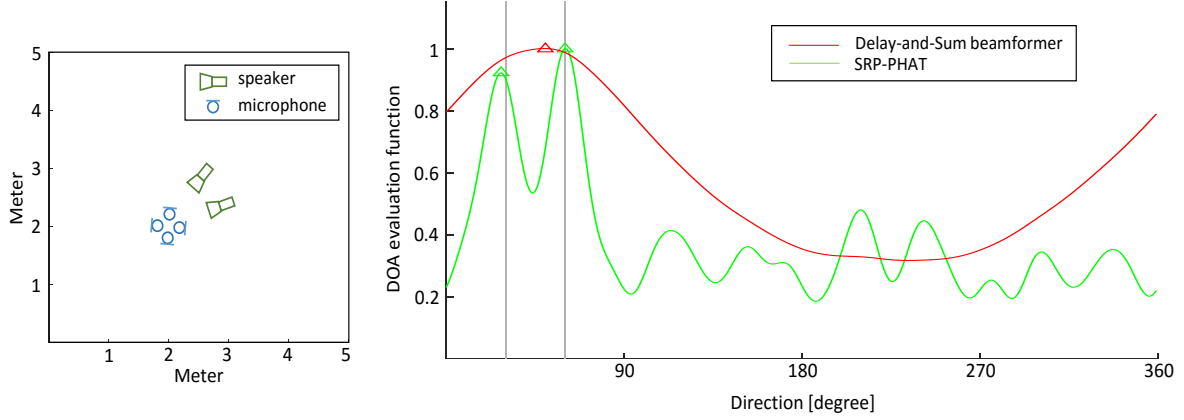
Figure 3.2: Comparison of DOA estimation between delay-and-sum beamformer and SRP-PHAT. Condition of the estimation is same as that of Fig. 2.2. SRP-PHAT achieves moderate resolution and improve estimation accuracy of delay-and-sum beamformer. However, there is a small error at the direction of $30°$ and the resolution is still lower than MUSIC.

$\mathbf{\Psi}(\omega) \odot \mathbf{C}_2(\omega)$ and the steering vector $\mathbf{b}(\omega; \theta)$.

$$\overline{f_{\mathrm{SRP}}(\theta)} \triangleq \sum_\omega f_{\mathrm{SRP}}(\omega; \theta), \tag{3.14}$$

$$f_{\mathrm{SRP}}(\omega; \theta) \triangleq \mathbf{b}^H(\omega; \theta) \left(\mathbf{\Psi}(\omega) \odot \mathbf{C}_2(\omega)\right) \mathbf{b}(\omega; \theta), \tag{3.15}$$

$$\mathbf{\Psi}(\omega) = [\Psi_{ij}]_{ij}, \tag{3.16}$$

where $\odot$ denotes Hadamard product, and $\mathbf{\Psi}(\omega)$ is the filter matrix whose (i, j) entry is Ψ_{ij}. We define $\overline{f_{\mathrm{SRP}}(\theta)}$ as the DOA evaluation function of SRP. Note that the estimation using SRP when $M = 2$ is equivalent to that using GCC because

$$f_{\mathrm{GCC}}(\Delta\tau'_{12}) = \sum_\omega \Psi_{12}(\omega) G_{x_1 x_2}(\omega) \exp(-\jmath\omega\Delta\tau'_{12}), \tag{3.17}$$

$$\begin{aligned}
f_{\mathrm{SRP}}(\boldsymbol{\tau}') = f_{\mathrm{SRP}}\left([\Delta\tau'_{11}, \Delta\tau'_{12}]^T\right) &= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_\omega \Psi_{ij}(\omega) G_{x_i x_j}(\omega) \exp\left(-\jmath\omega(-\Delta'_{1i} + \Delta'_{1j})\right), \\
&= \sum_\omega \Psi_{12}(\omega) G_{x_1 x_2}(\omega) \exp\left(-\jmath\omega\Delta'_{12}\right) \\
&\quad + \sum_\omega \Psi_{21}(\omega) G_{x_2 x_1}(\omega) \exp\left(\jmath\omega\Delta'_{12}\right) \\
&= 2 f_{\mathrm{GCC}}(\Delta\tau'_{12}) \\
&\propto f_{\mathrm{GCC}}(\Delta\tau'_{12}). \tag{3.18}
\end{aligned}$$

Figure 3.2 shows the improvement of DOA estimation performance between delay-and-sum beamformer and SRP-PHAT.

13

3.4 MUSIC

MUSIC is a DOA estimator based on subspace analysis exploiting the covariance of the signals. According to an analogy of the steering vectors to the eigenvectors of the covariance matrix associated with the non-zero eigenvalues, MUSIC estimates DOAs with high resolution.

MUSIC analyzes the following $M \times M$ covariance matrix $\mathbf{C}_2(\omega)$, whose entries are the covariance of the observation $\mathbf{x}(\omega, t)$:

$$\mathbf{C}_2(\omega) = E[\mathbf{x}(\omega, t)\mathbf{x}^H(\omega, t)], \tag{3.19}$$

Under the assumption that noise signals observed with each microphone have sufficiently small correlations and noise covariances almost do not affect nondiagonal entries, the following equations are obtained by the eigendecomposition of the covariance matrix $\mathbf{C}_2(\omega)$:

$$\mathbf{C}_2(\omega) = \mathbf{U}(\omega)\mathbf{G}(\omega)\mathbf{U}^H(\omega), \tag{3.20}$$

$$\mathbf{U}(\omega) = [\mathbf{u}_1(\omega), \cdots, \mathbf{u}_M(\omega)],$$

$$\mathbf{U}^H(\omega)\mathbf{U}(\omega) = \mathbf{I}_M, \tag{3.21}$$

$$\mathbf{G}(\omega) = \mathrm{diag}[g_1(\omega), \cdots, g_M(\omega)],$$

$$g_1(\omega) \geq \cdots \geq g_N(\omega),$$

$$g_{N+1}(\omega) \approx \cdots \approx g_M(\omega) \approx 0, \tag{3.22}$$

where $\mathbf{u}_1(\omega), \cdots, \mathbf{u}_M(\omega)$ are the eigenvectors associated with the respective eigenvalues $g_1(\omega), \cdots, g_M(\omega)$, $\mathbf{I}_i$ denotes the i-dimensional identity matrix, $\mathrm{diag}[\cdot]$ is a diagonal matrix with the arguments in the diagonal entries. The first term of Eq. (3.22) can be classified into N large eigenvalues and $(M - N)$ nearly zero eigenvalues, and N eigenvectors associated with the former N eigenvalues span the almost identical space to signal subspace $\mathcal{S}$ defined as the span of a set of array manifold vectors $\mathbf{a}_1(\omega), \cdots, \mathbf{a}_N(\omega)$.

$$\mathrm{span}\,[\mathbf{u}_1(\omega), \cdots, \mathbf{u}_N(\omega)] \approx \mathcal{S} \triangleq \mathrm{span}\,[\mathbf{a}_1(\omega), \cdots, \mathbf{a}_N(\omega)], \tag{3.23}$$

where $\mathrm{span}[\cdot]$ denotes the subspace spanned by the argument vectors. Moreover, the orthogonal complement of $\mathcal{S}$ in $\mathrm{span}\,[\mathbf{u}_1(\omega), \cdots, \mathbf{u}_M(\omega)]$ is defined as the noise subspace. According to the orthogonality between eigenvectors $\mathbf{u}_1(\omega), \cdots, \mathbf{u}_M(\omega)$ and the Eq. (3.23), the following relation is satisfied between the array manifold vectors $\mathbf{a}_1(\omega), \cdots, \mathbf{a}_N(\omega)$ and the vectors $\mathbf{u}_{N+1}(\omega), \cdots, \mathbf{u}_M(\omega)$, which span the noise subspace orthogonal to the signal subspace:

$$\mathbf{a}_i^H(\omega)\mathbf{u}_j(\omega) \approx 0$$

$$\text{for } i = 1, \cdots, N \ \ j = N+1, \cdots, M. \tag{3.24}$$

14

With the steering vectors $\mathbf{b}(\omega; \theta_i) \simeq \mathbf{a}_i(\omega)$ $(i = 1, \cdots, N)$, we define the following DOA evaluation function $f(\omega; \theta)$, which utilizes the orthogonality in Eq. (3.24):SIC

$$f_{\mathrm{MUSIC}}(\omega; \theta) \triangleq \frac{1}{\sum_{j=N+1}^{M} |\mathbf{b}^H(\omega; \theta)\mathbf{u}_j(\omega)|^2}, \tag{3.25}$$

which is maximal in a direction θ close to source directions.

Note that the evaluation function $f_{\mathrm{MUSIC}}(\omega; \theta)$ is defined for each narrowband, and it is necessary to integrate the information from all frequency bins in order to obtain a single wideband DOA estimation $\overline{f_{\mathrm{MUSIC}}(\theta)}$. A common approach is to average Eq. (3.25) over frequencies and several average operations have been discussed [41, 42, 43] thus far.

$$\overline{f_{\mathrm{MUSIC}}(\theta)} = \begin{cases} \dfrac{1}{J} \displaystyle\sum_{\omega} f_{\mathrm{MUSIC}}(\omega; \theta) & \text{(arithmetic mean)} \\[3mm] \left[\displaystyle\prod_{\omega} f_{\mathrm{MUSIC}}(\omega; \theta) \right]^{\frac{1}{J}} & \text{(geometric mean)} \\[3mm] \dfrac{J}{\displaystyle\sum_{\omega} \dfrac{1}{f_{\mathrm{MUSIC}}(\omega; \theta)}} & \text{(harmonic mean)} \end{cases}, \tag{3.26}$$

where J denotes the number of averaged frequency bins. As we reviewed above, delay-and-sum beamformer and its extensions conduct simple addition of narrowband estimation along frequency direction, which corresponds to arithmetic mean. On the other hand, ideal MUSIC estimation has infinite peaks on Eq. (3.25), and the evaluation function for directions other than true DOAs of sound sources approaches nearly zero. Therefore, we use the following geometric mean for MUSIC-based methods among these operations:

$$\overline{f_{\mathrm{MUSIC}}(\theta)} \triangleq \left[\prod_{\omega} f_{\mathrm{MUSIC}}(\omega; \theta) \right]^{\frac{1}{J}}. \tag{3.27}$$

Finally, DOAs are estimated by finding the peaks of $\overline{f_{\mathrm{MUSIC}}(\theta)}$. Note that MUSIC cannot correctly estimate DOAs in underdetermined conditions, because it requires at least one-dimensional noise subspace to compute Eq. (3.25). Also, for the more reliable estimation, M must become sufficiently greater than N to accurately evaluate the linear dependency between steering vectors and noise subspace.

3.5 2q-MUSIC

2q-MUSIC is an extension of MUSIC exploiting 2qth-order cumulants with an arbitrary positive integer q. Also, 2q-MUSIC is equivalent to standard MUSIC when $q = 1$ because the second-order cumulant is equivalent to the covariance.

$2q$-MUSIC analyzes the following $M^q \times M^q$ cumulant matrix $\mathbf{C}_{2q}(\omega)$, whose entries are the temporal mean of the $2q$th-order cross-cumulants of the observation $\mathbf{x}(\omega, t)$:

$$\mathbf{C}_{2q}(\omega) = \left[c_{ij}(\omega)\right]_{ij}, \tag{3.28}$$

$$c_{ij}(\omega) \triangleq \mathrm{Cum}[x^{\oplus}_{o_{i,1}}(\omega, t), \cdots, x^{\oplus}_{o_{i,q}}(\omega, t),$$
$$x^{\ominus}_{o_{j,q+1}}(\omega, t), \cdots, x^{\ominus}_{o_{j,2q}}(\omega, t)], \tag{3.29}$$

$$x^{\oplus}_{o_{i,l}}(\omega, t) \triangleq \begin{cases} x_{o_{i,l}}(\omega, t) & (\text{if } l \text{ is odd}) \\ x^{*}_{o_{i,l}}(\omega, t) & (\text{if } l \text{ is even}) \end{cases}, \tag{3.30}$$

$$x^{\ominus}_{o_{j,l}}(\omega, t) \triangleq \begin{cases} x^{*}_{o_{j,l}}(\omega, t) & (\text{if } l \text{ is odd}) \\ x_{o_{j,l}}(\omega, t) & (\text{if } l \text{ is even}) \end{cases}, \tag{3.31}$$

where $\mathrm{Cum}\left[\cdots\right]$ denotes the $2q$th-order cumulant given by its $2q$ arguments. The tuples of the q indices $\{o_{i,1}, \cdots, o_{i,q}\}$ and $\{o_{j,q+1}, \cdots, o_{j,2q}\}$ are composed of repeated permutations of the values $1, \cdots, M$:

$$o_{i,l} = 1 + \left\lfloor \frac{i-1}{M^{q-l}} \right\rfloor \bmod M$$
$$\text{for } i = 1, \cdots, M^d \ l = 1, \cdots, q, \tag{3.32}$$

$$o_{j,l} = 1 + \left\lfloor \frac{j-1}{M^{2q-l}} \right\rfloor \bmod M$$
$$\text{for } j = 1, \cdots, M^d \ l = q+1, \cdots, 2q, \tag{3.33}$$

where $\lfloor \cdot \rfloor$ denotes the floor function and mod denotes the modulus. As described in [22], the entries in $\mathbf{C}_{2q}$ are interpreted as the correlations between M^q-dimensional virtual observations, whose spatial arrangements are given by the higher-dimensional steering vectors $\mathbf{b}_q(\omega; \theta)$ stated below. The cumulant matrix $\mathbf{C}_{2q}(\omega)$ can be diagonalized as

$$\mathbf{C}_{2q}(\omega) = \begin{bmatrix} \mathbf{U}_s(\omega) \ \mathbf{U}_n(\omega) \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta}_s(\omega) & \\ & \boldsymbol{\Delta}_n(\omega) \end{bmatrix} \begin{bmatrix} \mathbf{U}_s^H(\omega) \\ \mathbf{U}_n^H(\omega) \end{bmatrix}, \tag{3.34}$$

K is the nonnegative integer parameter indicating the dimensionality of the *signal subspace*, $\boldsymbol{\Delta}_s(\omega)$ is the diagonal matrix of the largest K eigenvalues of the cumulant matrix $\mathbf{C}_{2q}(\omega)$, $\boldsymbol{\Delta}_n(\omega)$ is the diagonal matrix of the other $M^q - K$ eigenvalues of $\mathbf{C}_{2q}(\omega)$, and $\mathbf{U}_s(\omega)$ and $\mathbf{U}_n(\omega)$ are the unitary eigenmatrices composed of the corresponding eigenvectors. Here, the column eigenvectors of $\mathbf{U}_s(\omega)$ and $\mathbf{U}_n(\omega)$ are regarded as the bases of the signal and noise subspaces, respectively. Because of the additivity of cumulants, the N signal sources ideally make only the N eigenvalues large and they span the N-dimensional signal subspace. Thus, the appropriate setting of the signal

subspace dimensionality parameter is $K \leftarrow N$. Although the automatic estimation of parameter K is theoretically possible by finding the boundary between the large-eigenvalue group and the small-eigenvalue group, the boundary between the signal and noise subspaces is not clear in practical analysis under the existence of variance and noise. Therefore, in this dissertation we manually set the signal subspace dimensionality parameter $K \leftarrow N$ given the number of sources N. Making use of the orthogonality between the higher-dimensional steering vectors $\mathbf{b}_q(\omega; \theta)$ and the noise subspace $\mathbf{U}_n(\omega)$, $2q$-MUSIC constructs the DOA evaluation function $f_{2q}(\omega; \theta)$ as follows:

$$
\mathbf{b}_q(\omega; \theta) \triangleq \begin{cases} \mathbf{b}(\omega; \theta) & (q = 1) \\ \mathbf{b}(\omega; \theta) \otimes \mathbf{b}_{q-1}^*(\omega; \theta) & (q \geq 2) \end{cases}, \tag{3.35}
$$

$$
f_{2q}(\omega; \theta) \triangleq \frac{1}{\left| \mathbf{b}_q^H(\omega; \theta) \mathbf{U}_n(\omega) \right|^2}, \tag{3.36}
$$

where $\otimes$ denotes the Kronecker product operator. Finally, we merge the narrowband evaluation functions into a wideband evaluation function involving the geometric mean:

$$
\overline{f_{2q}(\theta)} \triangleq \left[\prod_\omega f_{2q}(\omega; \theta) \right]^{\frac{1}{J}}. \tag{3.37}
$$

DOAs are estimated by finding the peaks of Eq. (3.37). When $q \geq 2$, $2q$-MUSIC achieves underdetermined DOA estimation capability with high resolution owing to its richer expressiveness of the subspaces based on dimensional expansion from M to M^q.

Because of the additivity of the cumulants, $2q$-MUSIC has little bias in its model if it can utilize an infinite number of snapshots without noise to obtain the temporal mean. However, $2q$-MUSIC suffers from performance degradation when the number of snapshots is small because the variance of the estimated cumulants is large. There is also a problem of computational complexity: $2q$-MUSIC must compute the complicated Leonov–Shiryaev formula [44] for every entry $c_{ij}(\omega)$ of $\mathbf{C}_{2q}(\omega)$.

Chapter 4

Proposed method: mapped MUSIC

This chapter proposes mapped MUSIC, an estimation algorithm that nonlinealy maps the observed signal onto a space with expanded dimensionality and conducts MUSIC-based correlation analysis in the expanded space. By way of increasing the dimensionality of the noise subspace for sparse signal via the higher-dimensional mapping, the proposed method enables the estimation of DOAs in the case of underdetermined conditions.We describes the algorithm of mapped MUSIC and compares with the $2q$-MUSIC, a similar MUSIC extension based on high order cumulants. Finally, we present the efficiency of mapped MUSIC for short time analysis through the evaluation experiments. Section 4.1 explains DOA estimation algorithm of mapped MUSIC. Section 4.2 proposes a class of map ϕ_d with which mapped MUSIC analyzes the moments of arbitrary even orders. Moreover, we explain how the map achieve underdetermined DOA estimation for the sparse data. Section 4.3 describes the alternative maps that gives the equivalent DOA estimation to ϕ_d and have lower computational complexity. Furthermore, we present the alternative mappings in the case of $d = 2, 3$. Section 4.4 discusses the comparison between mapped MUSIC and $2q$-MUSIC from the viewpoint of statistical properties and computational complexity. Section 4.5 compares the practical estimation performances of the proposed methods and the conventional methods about DOA estimation accuracy and execution time.. Section 4.6 concludes this chapter.

4.1   DOA estimation algorithm of mapped MUSIC

Mapped MUSIC maps the M-dimensional observed signal $\mathbf{x}(\omega, t)$ onto an M'-dimensional Euclidean space $(M' \geq M)$ with a nonlinear function $\phi : \mathbb{C}^M \to \mathbb{C}^{M'}$ and conducts a similar analysis to MUSIC with the mapped observation vector $\phi(\mathbf{x}(\omega, t))$. Mapped MUSIC is also a generalization of standard MUSIC, which corresponds to the special case that $\phi(\mathbf{x}(\omega, t)) = \mathbf{x}(\omega, t)$. To estimate

DOAs accurately with mapped MUSIC, the information on the correlations between observation vectors must be retained after mapping. To maintain the spatial properties, we impose the following three conditions in the choice of the mapping function $\phi(\mathbf{x}(\omega, t))$:

1. The magnitude relation of the norm is retained.

$$||\phi(\mathbf{x})||_2 \geq ||\phi(\mathbf{y})||_2 \text{ if } ||\mathbf{x}||_2 \geq ||\mathbf{y}||_2. \tag{4.1}$$

2. The origin remains intact.

$$\phi(\mathbf{x}) = 0 \text{ if } \mathbf{x} = 0. \tag{4.2}$$

3. The orthogonality between vectors is preserved.

$$\phi^H(\mathbf{x})\phi(\mathbf{y}) = 0 \text{ if } \mathbf{x}^H\mathbf{y} = 0. \tag{4.3}$$

By using mapping functions satisfying these three conditions, mapped MUSIC appropriately achieves underdetermined DOA estimation capability with high resolution without any major adverse effects.

We describe the DOA estimation algorithm of mapped MUSIC using the mapping ϕ satisfying Eqs. (4.1)–(4.3). The covariance matrix of $\phi(\mathbf{x}(\omega, t))$ is expressed as

$$\mathbf{R}(\omega) = E[\phi(\mathbf{x}(\omega, t))\phi^H(\mathbf{x}(\omega, t))], \tag{4.4}$$

The following equations are obtained by the eigendecomposition of the covariance matrix $\mathbf{R}(\omega)$:

$$\mathbf{R}(\omega) = \mathbf{V}(\omega)\mathbf{E}(\omega)\mathbf{V}^H(\omega), \tag{4.5}$$

$$\mathbf{V}(\omega) = [\mathbf{v}_1(\omega), \cdots, \mathbf{v}_{M'}(\omega)],$$

$$\mathbf{V}^H(\omega)\mathbf{V}(\omega) = \mathbf{I}_{M'}, \tag{4.6}$$

$$\mathbf{E}(\omega) = \text{diag}[e_1(\omega), \cdots, e_{M'}(\omega)],$$

$$e_1(\omega) \geq \cdots \geq e_{M'}(\omega), \tag{4.7}$$

$$M' = \dim[\phi(\mathbf{x}(\omega, t))], \tag{4.8}$$

where $\mathbf{v}_1(\omega), \cdots, \mathbf{v}_{M'}(\omega)$ are the eigenvectors associated with the respective eigenvalues $e_1(\omega), \cdots, e_{M'}(\omega)$, and $\dim[\cdot]$ is the dimensionality of the argument vector. By manually setting the signal subspace dimensionality parameter K similarly to in $2q$-MUSIC, we define the subspace spanned by $\mathbf{v}_1(\omega), \cdots, \mathbf{v}_K(\omega)$ as the signal subspace $\mathcal{S}(\omega)$ in the mapped space,

$$\mathcal{S}(\omega) \triangleq \text{span}[\mathbf{v}_1(\omega), \cdots, \mathbf{v}_K(\omega)]. \tag{4.9}$$

Moreover, the orthogonal complement of $\mathcal{S}(\omega)$ in $\mathrm{span}[\mathbf{v}_1(\omega), \cdots, \mathbf{v}_{M'}(\omega)]$ is defined as the noise subspace. The following relation is satisfied between the maps of array manifold vectors $\mathbf{a}_1(\omega), \cdots, \mathbf{a}_N(\omega)$ and the vectors $\mathbf{v}_{K+1}(\omega), \cdots, \mathbf{v}_{M'}(\omega)$ on the noise subspace:

$$\phi^H(\mathbf{a}_i(\omega))\mathbf{v}_j(\omega) \approx 0$$
$$\text{for } i = 1, \cdots, N \;\; j = K + 1, \cdots, M'. \tag{4.10}$$

Under the condition $\mathbf{b}(\omega; \theta_i) \simeq \mathbf{a}_i(\omega)$ $(i = 1, \cdots, N)$, we can find the true sound source directions by searching for the orthogonal projection onto the mapped noise subspace from the mapped steering vectors $\phi(\mathbf{b}(\omega; \theta))$. Similarly to in MUSIC and $2q$-MUSIC, we define the following DOA evaluation function $f_{\mathrm{map}}(\omega; \theta)$, which utilizes the orthogonality in Eq. (4.10), for mapped MUSIC:

$$f_{\mathrm{map}}(\omega; \theta) \triangleq \frac{1}{\sum_{j=K+1}^{M'} |\phi^H(\mathbf{b}(\omega; \theta))\mathbf{v}_j(\omega)|^2}, \tag{4.11}$$

which is maximal in a direction θ close to θ_i. Finally, we merge the narrowband evaluation functions into a wideband evaluation function involving the geometric mean:

$$\overline{f_{\mathrm{map}}(\theta)} \triangleq \left[\prod_\omega f_{\mathrm{map}}(\omega; \theta) \right]^{\frac{1}{J}}. \tag{4.12}$$

4.2   Mapping for analysis of $2d$th-order moments

As shown in Sect. 4.1, mapped MUSIC can use any mapping function satisfying Eqs. (4.1)–(4.3), but its properties change with the choice of mapping. In this dissertation, to quantitatively evaluate the properties of the mapping, we focus on the mapping $\phi_d : \mathbb{C}^M \to \mathbb{C}^{M^d}$, which gives a $2d$th-order cross-moment matrix as its covariance matrix. The mapping function $\phi_d(\mathbf{x}(\omega, t))$ is defined recursively as

$$\phi_d(\mathbf{x}(\omega, t)) \triangleq \begin{cases} \mathbf{x}(\omega, t) & (d = 1) \\ \mathbf{x}(\omega, t) \otimes \phi_{d-1}^*(\mathbf{x}(\omega, t)) & (d \geq 2) \end{cases}. \tag{4.13}$$

Figure 4.1 shows specific examples of ϕ_d when $M = 2$ and d up to 3. According to this definition, each entry of the map $\phi_d(\mathbf{x}(\omega, t))$ is given as the product of d observed signals corresponding to its entry index,

$$\phi_d(\mathbf{x}(\omega, t)) = \left[\prod_{l=1}^d x_{o_{1,l}}^\oplus(\omega, t), \cdots, \prod_{l=1}^d x_{o_{M^d,l}}^\oplus(\omega, t) \right]^T, \tag{4.14}$$

$$x_{o_{i,l}}^\oplus(\omega, t) \triangleq \begin{cases} x_{o_{i,l}}(\omega, t) & (\text{if } l \text{ is odd}) \\ x_{o_{i,l}}^*(\omega, t) & (\text{if } l \text{ is even}) \end{cases}. \tag{4.15}$$
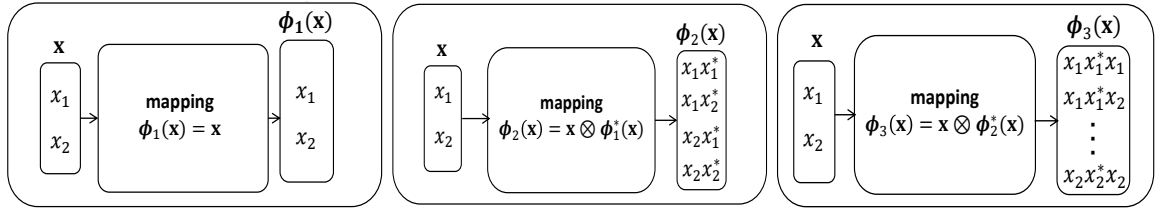
Figure 4.1: Specific examples of mapping function ϕ_d with $M = 2$ and d up to 3.

The entries of $\phi_d(\mathbf{x}(\omega, t))$ are similar to the first half of those in Eq. (3.29); thus, the tuple of d indices $\{o_{k,1}, \cdots, o_{k,d}\}$ is given in the same manner as Eq. (3.32):

$$o_{k,l} = 1 + \left\lfloor \frac{k-1}{M^{d-l}} \right\rfloor \mod M$$

$$\text{for } k = 1, \cdots, M^d \ \ l = 1, \cdots, d. \tag{4.16}$$

Then, the covariance matrix of the mapping of the observed signal $\phi_d(\mathbf{x}(\omega, t))$ is explicitly expressed as an $M^d \times M^d$ moment matrix whose entries are $2d$th-order cross-moments of the observations,

$$\mathbf{R}_{2d}(\omega) = E\left[\phi_d(\mathbf{x}(\omega, t))\phi_d^H(\mathbf{x}(\omega, t))\right]$$

$$= [r_{ij}(\omega)]_{ij}, \tag{4.17}$$

$$r_{ij}(\omega) \triangleq E\left[\prod_{l=1}^{d} x_{o_{i,l}}^{\oplus}(\omega, t) \prod_{l=1}^{d} x_{o_{j,l}}^{\ominus}(\omega, t)\right], \tag{4.18}$$

$$x_{o_{j,l}}^{\ominus}(\omega, t) \triangleq \begin{cases} x_{o_{j,l}}^*(\omega, t) & (\text{if } l \text{ is odd}) \\ x_{o_{j,l}}(\omega, t) & (\text{if } l \text{ is even}) \end{cases}. \tag{4.19}$$

In the following we discuss the determination of the signal subspace dimensionality K in Eqs. (4.9)–(4.11) for the correct estimation using the mapping $\phi_d : \mathbb{C}^M \to \mathbb{C}^{M^d}$ for the higher-order moment analysis. In contrast to the cumulants exploited in $2q$-MUSIC, moments do not maintain additivity, and it should be noted that underdetermined DOA estimation is not obtained by the vector dimensionality extending itself as we employ the mapping ϕ_d. When the noise-free observation

model is expressed according to Eqs. (2.1)–(2.3),

$$\mathbf{x}(\omega, t) = [x_1(\omega, t), \cdots, x_M(\omega, t)]^T$$
$$= \mathbf{A}(\omega)\mathbf{s}(\omega, t), \tag{4.20}$$
$$\mathbf{A}(\omega) = [\mathbf{a}_1(\omega), \cdots, \mathbf{a}_N(\omega)], \tag{4.21}$$
$$\mathbf{a}_i(\omega) = [a_{1,i}(\omega), \cdots, a_{M,i}(\omega)]^T, \tag{4.22}$$
$$\mathbf{s}(\omega, t) = [s_1(\omega, t), \cdots, s_N(\omega, t)]^T, \tag{4.23}$$

its mapping $\phi_d(\mathbf{x})$ is expressed by the mapping of source vector $\phi_d(\mathbf{s})$ and array manifold vectors $\psi_d(\cdots)$ in the higher-dimensional space:

$$\phi_d(\mathbf{x}) = [\psi_d(\mathbf{a}_1, \mathbf{a}_1, \cdots, \mathbf{a}_1), \psi_d(\mathbf{a}_1, \mathbf{a}_1, \cdots, \mathbf{a}_2), \cdots, \psi_d(\mathbf{a}_N, \mathbf{a}_N, \cdots, \mathbf{a}_N)]\,\phi_d(\mathbf{s}) \tag{4.24}$$

$$\psi_d(\mathbf{a}_{i_1}, \mathbf{a}_{i_2}, \cdots, \mathbf{a}_{i_d}) \triangleq \left[\prod_{l=1}^{d} a^{\oplus}_{o_{1,l},i_l}(\omega, t), \cdots, \prod_{l=1}^{d} a^{\oplus}_{o_{M^d,l},i_l}(\omega, t) \right]^T$$

$$\text{for } i_1, \cdots, i_N = 1, \cdots, N, \tag{4.25}$$

$$\psi_d(\mathbf{a}_{i_1}, \mathbf{a}_{i_2}, \cdots, \mathbf{a}_{i_d}) = \phi_d(\mathbf{a}_n)$$

$$\text{if } i_1 = i_2 = \cdots = i_N = n. \tag{4.26}$$

From the above equations, we found the mapped observations $\phi_d(\mathbf{x})$ are given by the directional mapped sources $\phi_d(\mathbf{s})$ whose dimensionality is N^d, and the signal subspace obtained from the covariance matrix of mapped observation $\mathbf{R}_{2d}(\omega) = E\left[\phi_d(\mathbf{x}(\omega, t))\phi_d^H(\mathbf{x}(\omega, t))\right]$ also becomes N^d dimensional. This relation means that proposed mapping ϕ_d increases the dimensionality of the analysis from M to M^d, but does not directly clear the problem of standard MUSIC when we assume underdetermined condition $(N > M)$ for the original observation, because directional sound source also increase from N to N^d by the mapping. However, $N^d - N$ entries in mapped sources $\phi_d(\mathbf{s})$ are given as the cross product of different original sources, and these entries are reduced by assuming that the sources having high sparseness over time-frequency direction, which is referred to as W-disjoint orthogonality [28]. Therefore, for the highly sparse data such as speech

22

signal, that is the main target of this study, Eqs. (4.24) – (4.26) are approximated as follows.

$$
\phi_d(\mathbf{x}) \approx
\begin{bmatrix}
[\boldsymbol{\phi}_d(\mathbf{a}_1), \boldsymbol{\phi}_d(\mathbf{a}_2), \cdots, \boldsymbol{\phi}_d(\mathbf{a}_N)]
\begin{bmatrix}
|s_1|^{d-1}s_1 \\
|s_2|^{d-1}s_2 \\
\vdots \\
|s_N|^{d-1}s_N
\end{bmatrix} & \text{(if } d \text{ is odd)} \\[4em]
[\boldsymbol{\phi}_d(\mathbf{a}_1), \boldsymbol{\phi}_d(\mathbf{a}_2), \cdots, \boldsymbol{\phi}_d(\mathbf{a}_N)]
\begin{bmatrix}
|s_1|^{d} \\
|s_2|^{d} \\
\vdots \\
|s_N|^{d}
\end{bmatrix} & \text{(if } d \text{ is even)}
\end{bmatrix}. \tag{4.27}
$$

For the data with complete W-disjoint orthogonality, with which the time-frequency representations of the sources do not overlap at all, mapped sources span only N-dimensional subspacle and signal subspace dimensionality parameter K is set as $K \leftarrow N$. Although we cannot determine signal subspace dimensionality should N in actual analysis because of the source signals with not the perfect W-disjoint orthogonality, when the data are highly sparse, the signal subspace dimensionality becomes less than N^d and close to N. In the case, parameter setting $K \leftarrow N$ is enough suboptimal for the practical estimation and the increase in dimensionality of the covariance matrix $\mathbf{R}_{2d}(\omega)$ from M to M^d with the mapping ϕ_d enhances the expressiveness of the noise subspace to enables us to estimate the DOAs in underdetermined cases.

Also, note that the actual dimensionality after mapping becomes less than M^d. This is because the set of M^d mapped vectors $\phi_d(\mathbf{x}_i)$, $i = 1, \cdots, M^d$ with $\mathbf{x}_i \in \mathbb{C}^M$, contains several linearly dependent vectors and cannot span the whole M^d-dimensional space when $d > 2$. However, it is guaranteed that the dimensionality of the subspace spanned by the mapped vectors $\phi_d(\mathbf{x}_i)$ is an increasing function of d.

Here we discuss a similarity between the proposed method with the map ϕ_d and $2q$-MUSIC based on cumulant analysis. From Eqs. (3.35) and (4.13), the mapped steering vector $\phi_d(\mathbf{b}(\omega; \theta))$ is equal to the higher-dimensional steering vector $\mathbf{b}_q(\omega; \theta)$ in $2q$-MUSIC. This equality suggests that $2q$-MUSIC can also be interpreted as a mapping of $\mathbf{b}(\omega; \theta)$ onto a higher-dimensional Euclidean space identical to that in the proposed method. Regardless of the identity of the space, the matrices to be analyzed are composed of different statistics, resulting in a difference in their behaviors. In Sect. 4.4.1, we discuss their behaviors from the perspective of the bias-variance tradeoff.

4.3 Alternative mapping for efficient computation

As discussed above, DOA estimation by mapped MUSIC is based on the analysis of the linear dependence among the mapped vectors. Thus, two different maps ϕ and ϕ' give exactly the same DOA estimation when their inner products are identical,

$$\phi^H(\mathbf{x})\phi(\mathbf{y}) = \phi'^H(\mathbf{x})\phi'(\mathbf{y}), \ \forall \mathbf{x}, \mathbf{y} \in \mathbb{C}^M, \tag{4.28}$$

and the analyses using these two mapping functions give identical results. Therefore, the mapping function ϕ_d given by Eq. (4.13) also has alternative mappings that give the equivalent DOA estimation. In this section, we give a compact expression for the computational efficiency of mapped MUSIC in the cases of $d = 2, 3$.

The inner product between the mappings ϕ_d of two arbitrary complex vectors $\mathbf{x}$ and $\mathbf{y}$ is given by

$$\phi_d^H(\mathbf{x})\phi_d(\mathbf{y}) = \begin{cases} \left|\mathbf{x}^H\mathbf{y}\right|^{d-1}\mathbf{x}^H\mathbf{y} & (d \text{ is odd}) \\ \left|\mathbf{x}^H\mathbf{y}\right|^{d} & (d \text{ is even}) \end{cases}. \tag{4.29}$$

Note that the inner product becomes real when the degree of the mapping d is even. In this instance, a real-valued mapping $\phi'_d : \mathbb{C}^M \to \mathbb{R}^{M^d}$ satisfying Eq. (4.28) always exists, and it simplifies the construction of the covariance matrix and the eigenvalue problem because their calculations only include operations on real values.

For example, when $d = 2$, the following mapping $\phi'_2 : \mathbb{C}^M \to \mathbb{R}^{M^2}$ can be employed:

$$\phi'_2(\mathbf{x}) \triangleq \left[\phi'^T_{\text{abs}}(\mathbf{x}), \phi'^T_{\text{re}}(\mathbf{x}), \phi'^T_{\text{im}}(\mathbf{x})\right]^T, \tag{4.30}$$

$$\phi'_{\text{abs}}(\mathbf{x}) \triangleq [\forall |x_i|^2 | 1 \le i \le M]^T, \tag{4.31}$$

$$\phi'_{\text{re}}(\mathbf{x}) \triangleq \sqrt{2}[\forall \text{Re}[x_i x_j^*] | 2 \le i \le M, 1 \le j \le i - 1]^T, \tag{4.32}$$

$$\phi'_{\text{im}}(\mathbf{x}) \triangleq \sqrt{2}[\forall \text{Im}[x_i x_j^*] | 2 \le i \le M, 1 \le j \le i - 1]^T. \tag{4.33}$$

Moreover when $d > 2$, there are several redundant entries of equal value in ϕ_d, meaning that the rank of its covariance matrix becomes less than M^d. Hence, by designing an alternative mapping that satisfies Eq. (4.28) and omitting the redundant entries in the original mapping ϕ_d, we can remove the redundant computations. The mapping in the case of $d = 3$ can be simplified in this

manner. We define the mapping $\phi_3' : \mathbb{C}^M \to \mathbb{C}^{\frac{M^3+M^2}{2}}$ as

$$\phi_3'(\mathbf{x}) \triangleq \left[\phi_{3a}'^T(\mathbf{x}), \phi_{3b}'^T(\mathbf{x}), \phi_{3c}'^T(\mathbf{x}), \phi_{3d}'^T(\mathbf{x})\right]^T, \tag{4.34}$$

$$\phi_{3a}'(\mathbf{x}) \triangleq [\forall |x_i|^2 x_i^* | 1 \le i \le M]^T, \tag{4.35}$$

$$\phi_{3b}'(\mathbf{x}) \triangleq [\forall x_i x_j^{*2} | 1 \le i, j \le M, j \ne i]^T, \tag{4.36}$$

$$\phi_{3c}'(\mathbf{x}) \triangleq \sqrt{2}[\forall |x_i|^2 x_j^* | 1 \le i, j \le M, j \ne i]^T, \tag{4.37}$$

$$\phi_{3d}'(\mathbf{x}) \triangleq \sqrt{2}[\forall x_i x_j^* x_k^* | 1 \le i, j, k \le M]^T,$$

$$(i, j, \text{ and } k \text{ are different}). \tag{4.38}$$

4.4 Comparison of mapped MUSIC with $2q$-MUSIC

As discussed above, our proposed mapped MUSIC corresponds to the substitution of the moment matrix for the cumulant matrix in $2q$-MUSIC. We discuss the effectiveness of this substitution from the viewpoints of statistical properties and computational complexity.

4.4.1 Statistical properties

We compared the statistical properties of mapped MUSIC and $2q$-MUSIC in terms of the bias-variance tradeoff. Since the cumulants utilized in $2q$-MUSIC maintain additivity, the dimensionality of the signal subspace of the cumulant matrix coincides with the number of sources N if there is no noise and infinite snapshots [45, 46]. Thus, the signal and noise subspaces are identified correctly under such a condition where the cumulant matrix is appropriately estimated. In contrast to cumulants, moments do not have additivity, the signal subspace dimensionality of the moment matrix is generally greater than N, and this model bias degrades the accuracy of MUSIC analysis even if an accurate moment matrix is estimated. Thus, the proposed mapped MUSIC suffers from bias, in contrast to $2q$-MUSIC. However, the effect of the variance is more serious in $2q$-MUSIC. Since the cumulants are composed of multiple moments [22], the variance of the cumulants is larger than that of the moments of the same order. Thus, if a sufficient number of snapshots are unavailable, the accuracy of identification of the signal and noise subspaces is easily degraded by the variance.

In the following, we conducted a simulation of DOA estimation using pseudorandom numbers as the source to accurately evaluate the effects of bias and variance. In this simulation, we assumed noise-free observation and independent signals with different frequencies. Figure 4.2 shows the experimental environment. The number of microphones is four, the number of sources is five, and observations are created as mixtures of simulation-generated source signals $s(\omega, t)$ whose directivities are given by steering vectors $\mathbf{b}(\omega; \theta)$ with θ randomly chosen from $\theta = \{0°, 30°, \cdots, 330°\}$.
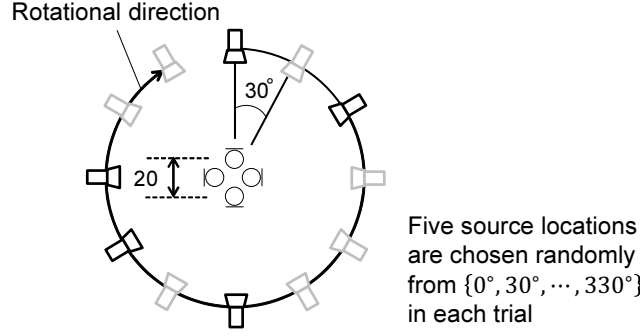
Figure 4.2: Experimental environment for statistical comparison.

Source signals $s(\omega, t)$ are generated from a circularly symmetric complex generalized Gaussian distribution (CGGD) [47] whose probability density function is given as

$$p\left(s(\omega, t)\right) = \frac{\beta \exp\left(-\left(\frac{|s(\omega, t)|}{\alpha}\right)^{\beta}\right)}{2\pi\alpha^2 \Gamma\left(\frac{2}{\beta}\right)}, \tag{4.39}$$

where $\exp\left(\cdot\right)$ denotes the natural exponential function, β is a shape parameter, and α is a scale parameter. In this experiment, we adopted the parameter settings $\beta = 0.315$ and $\alpha = 1$, which we found through a statistical investigation to have a similar property to that in the short-time Fourier analysis of speech [48], because speech processing is the main target of the proposed method. We conducted estimations with mapped MUSIC and $2q$-MUSIC with 4th- and 6th-order statistics to investigate the difference in the same statistics. Although standard MUSIC is not applicable to underdetermined case, for reference, we also conduct estimation with standard MUSIC by regarding the one-dimensional subspace concurrent with the minimum eigenvalue as the noise subspace in every frequency bin. We varied the number of snapshots L $(t = 1, \cdots, L)$ from 1 to 50,000 and conducted DOA estimation for each L. Moreover, we performed 100 trials with different combinations of the five source locations as a quantitative evaluation. Table 4.1 summarizes the experimental conditions.

As the evaluation criterion, we utilized the root-mean-squared error (RMSE):

$$\text{RMSE} \triangleq \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left|\hat{\theta}_i - \theta_i\right|^2}, \tag{4.40}$$

where N denotes the number of sources and $\hat{\theta}_i$ and θ_i are estimated DOA and true DOA of the ith sound source, respectively. Although there are several combinations for the correspondence be-

Table 4.1: Experimental conditions for statistical comparison

Microphone array	Circular array with radius of 0.1 m
# of microphones	4
Sound sources	Simulation-generated signals
# of sources	5
Source signal	Circularly symmetric complex generalized Gaussian distribution Set 1: $\beta = 0.315,\ \alpha = 1$ Set 2: $\beta = 0.05\text{--}3,\ \alpha = 1$
Frequency bandwidth	0.2–5 [kHz]
Frequency resolution	1 [kHz]
Array manifold vector	Steering vector $\mathbf{b}(\omega;\theta)$ $\theta = \{0°, 30°, \cdots, 330°\}$
Snapshot range	Set 1: 1–50,000 [samples] Set 2: 10–1,000 [samples]

tween estimated DOAs and true DOAs, we chose the combination that minimizes the error. The results are shown in Fig. 4.3. First, both $2q$-MUSIC and the proposed mapped MUSIC improve estimation accuracy with the increasing snapshots. Thus, the efficacy of analysis with the increased dimensionality based on higher-order statistics is ascertained. In the comparison between the same type of 4th- and 6th-order statistics, mapped MUSIC based on 6th-order moments shows better performance than that based on 4th-order moments with increasing number of snapshots. This is because higher-order statistics increase the dimensionality but are more affected by the variance. Although $2q$-MUSIC with 6th-order cumulants does not outperform that with 4th-order cumulants in the range in Fig. 4.3, the superiority of 6th-order cumulants is also expected upon further increasing the number of snapshots. For the range of 1–100 snapshots, mapped MUSIC performs better than $2q$-MUSIC, whereas $2q$-MUSIC performs better for a larger number of snapshots. Thus, the bias-variance tradeoff between $2q$-MUSIC and mapped MUSIC is confirmed. Furthermore, the difference in accuracy between $2q$-MUSIC and mapped MUSIC is still moderate even for 50,000 snapshots. This is because the bias in mapped MUSIC is canceled through the frequency-averaging operation as shown in Fig. 4.4. We can observe several pseudopeaks that do not correspond to true DOAs in the narrowband estimation as a result of the bias. However, these pseudopeaks disappear as a result of frequency averaging.
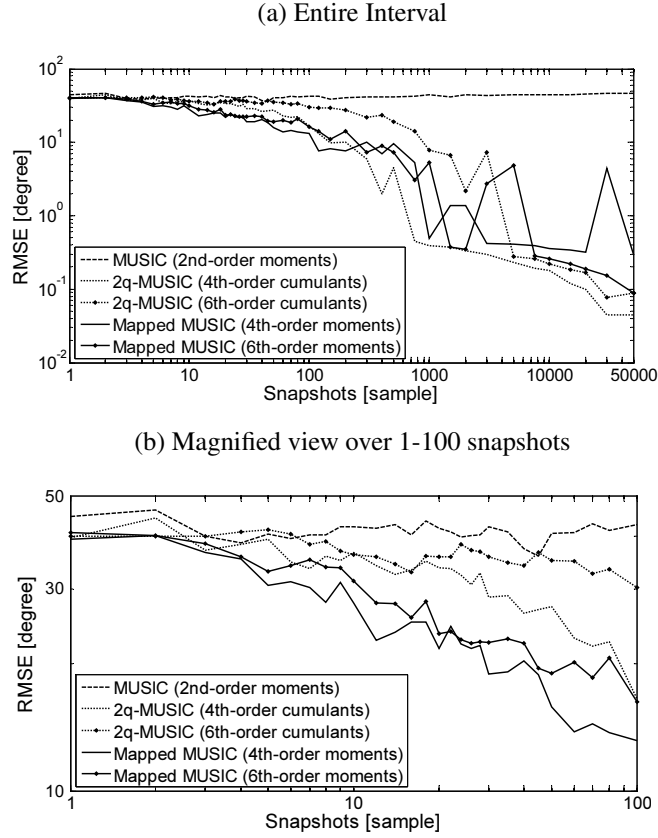
(a) Entire Interval

(b) Magnified view over 1-100 snapshots

Figure 4.3: Error transition with increasing number of snapshots ($\beta = 0.315$). Mapped MUSIC and $2q$-MUSIC show efficacy for underdetermined DOA estimation. The RMSE is evaluated over 100 trials. In the range of 1–100 snapshots, mapped MUSIC shows better performance and $2q$-MUSIC outperforms mapped MUSIC with the larger number of snapshots.

Furthermore, we evaluated the behavior of the bias-variance tradeoff for various signal characters by changing the shape parameter β of the CGGD. We varied β in the range from 0.05 to 3 with $\alpha = 1$, and evaluated each method in the same way as in the former experiment using the RMSE obtained from 100 trials. A lower β produces longer tails in the CGGD, and the CGGD is super-Gaussian when $\beta < 2$. As examples of practical numbers of snapshots, we used 10, 100, and 1000 snapshots for the evaluation. The results are shown in Fig. 4.5. We can see that the superiority of mapped MUSIC is maintained when β is small and that the distribution is sparse with a long tail, similarly to human speech. This is because the drawback of bias when using moments without additivity becomes less problematic with highly sparse data [49]. Thus, mapped MUSIC performs well with sparse signals such as speech.

(a) 3 kHz   (b) 4 kHz
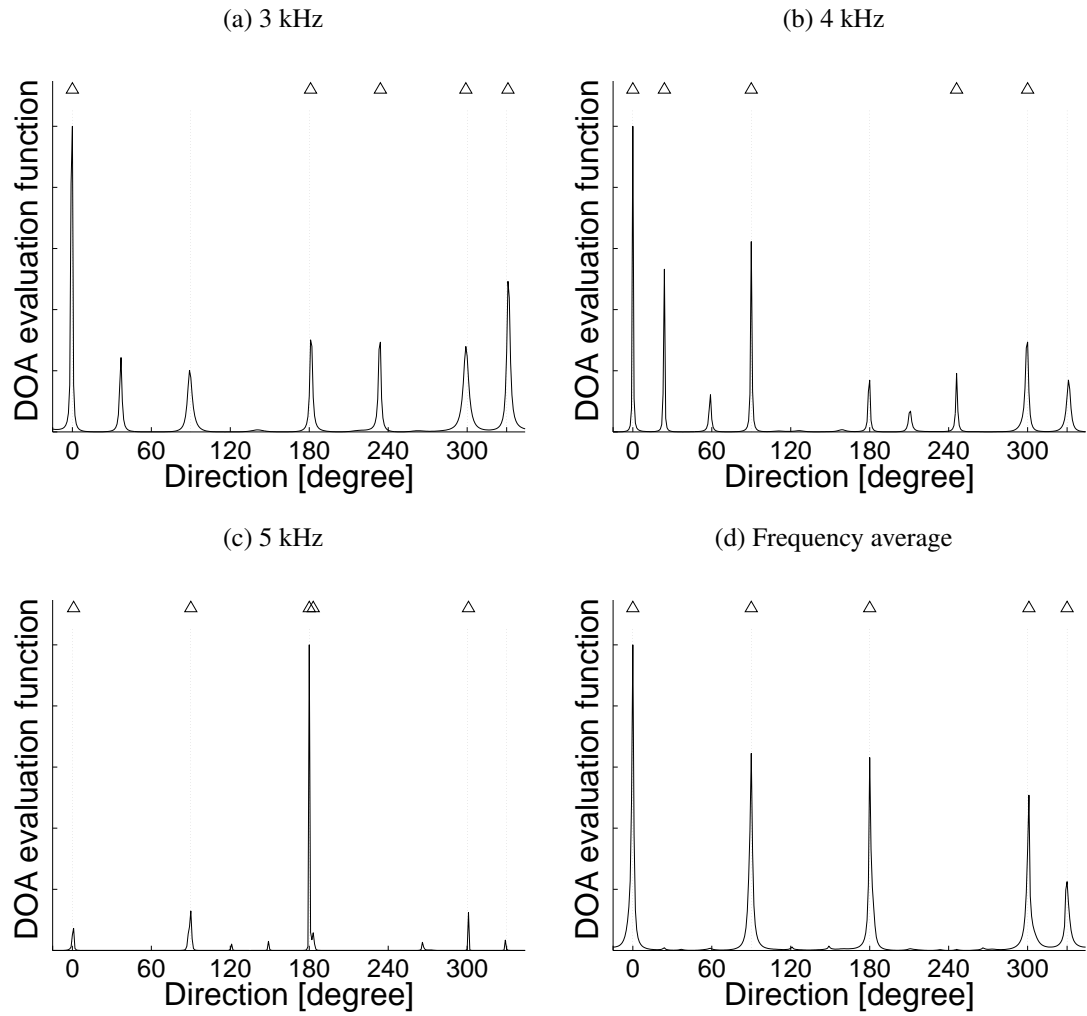
(c) 5 kHz   (d) Frequency average

Figure 4.4: DOA evaluation function of mapped MUSIC in the narrow and wide frequency bands. The estimation uses 4th-order moments and the sources are given by the parameters in Sect. 3, $\beta = 0.315$ and $\alpha = 1$. The number of snapshots is 50,000, triangles denote estimated DOAs, and the true source directions are denoted as vertical lines at $0°, 90°, 180°, 300°$, and $330°$. There are many pseudopeaks in the narrowband estimation, but these peaks are suppressed in the wideband estimation because they are canceled by the frequency-averaging operation.
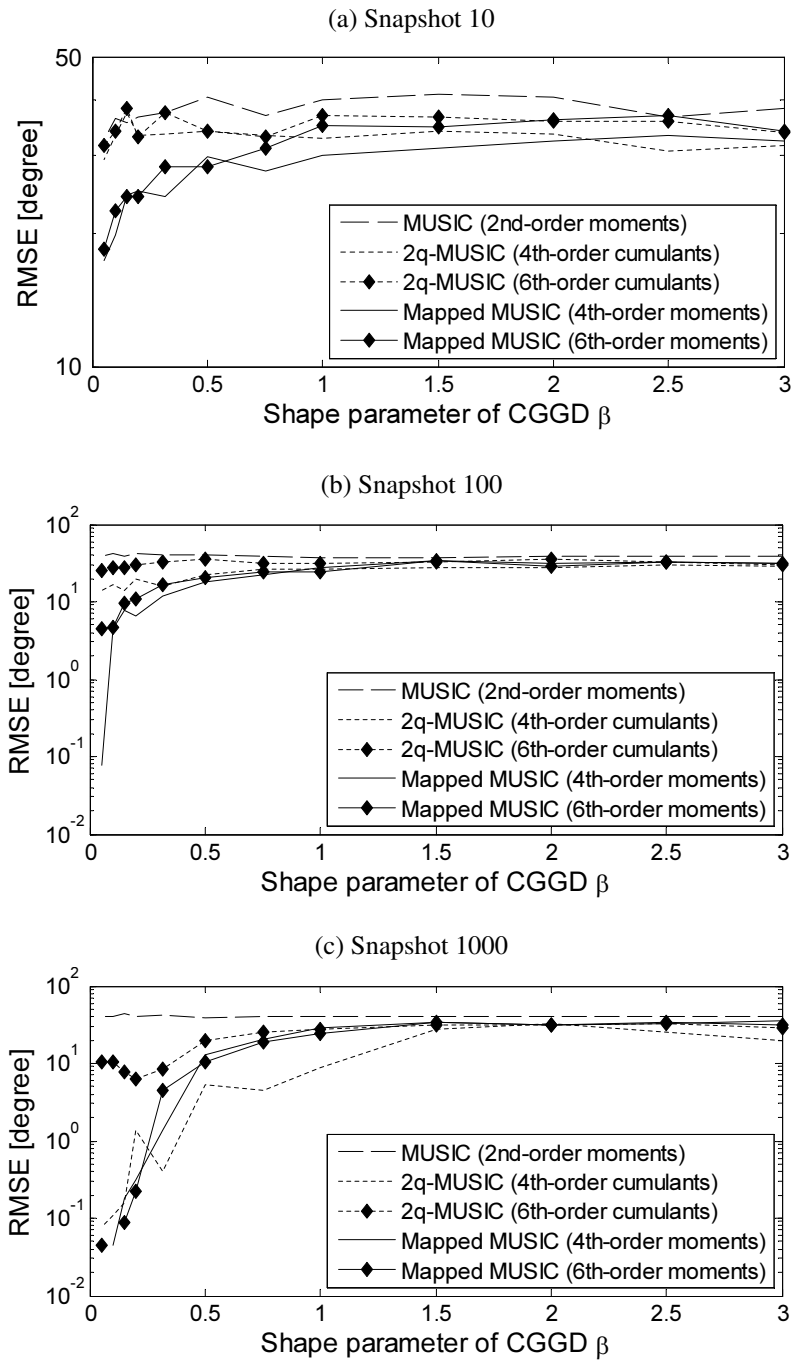
29

(a) Snapshot 10

(b) Snapshot 100

(c) Snapshot 1000

Figure 4.5: Dependence of error on shape parameter β. The RMSE is evaluated over 100 trials. Mapped MUSIC shows better performance when the shape parameter β is small because a lack of additivity in moment analysis becomes less problematic with sparse data.

4.4.2 Comparison of computational complexity

In this section, we discuss the computational complexity of the proposed method. As described in Sect. 3.5, the computation of the cumulants is complicated and the complexity increases rapidly when the statistical order becomes larger. $2q$-MUSIC must perform such computations for all entries of the cumulant matrix $\mathbf{C}_{2q}(\omega)$, and this procedure includes redundancy since many similar computations are required.

Conversely, mapped MUSIC achieves more rapid computation because a single moment has a simpler form than the cumulant, and the procedure in mapped MUSIC avoids redundancy in the construction of the moment matrix $\mathbf{R}_{2d}(\omega)$ by repeatedly using the same map $\phi_{2d}(\omega)$ given once. Table 4.2 shows a comparison of the numbers of multiplications required to construct the cumulant matrix $\mathbf{C}_{2q}(\omega)$ ($q = 2, 3$) in $2q$-MUSIC and the moment matrix $\mathbf{R}_{2d}(\omega)$ ($d = 2, 3$) in the proposed method. For reference, the numbers of multiplications with the alternative map ϕ'_{2d} ($d = 2, 3$), introduced in Sect. 4.3 for rapid computation, and for computing all entries individually (entrywise), such as in $2q$-MUSIC, are also shown. Moreover, Fig. 4.6 shows the transitions of the numbers of multiplications required to construct the cumulant matrix $\mathbf{C}_{2q}(\omega)$ and moment matrix $\mathbf{R}_{2d}(\omega)$ with the maps ϕ_{2d} and ϕ'_{2d}, respectively, with increasing number of snapshots from 0 to 10000 when the number of microphones is four.

4.5 Experiments

In this section, we compare the performances of the proposed method and the conventional methods via experiments assuming practical conditions for DOA estimation. The comparison is based on the DOA estimation accuracy and execution time, and the accuracy is evaluated using both simulation results and the real-room impulse responses.

4.5.1 Evaluation of DOA estimation accuracy

First, to investigate the estimation accuracy under various combinations of disturbances, we conducted a simulation experiment. We compared the accuracy of the proposed mapped MUSIC with those of SRP-PHAT [10], standard MUSIC, and $2q$-MUSIC. For this experiment, we created speech mixtures by convoluting Japanese speech samples [50] and room impulse responses assuming point sources in the environment shown in Fig. 4.7 using the image method [51]. We prepared seven different reverberant conditions and diffuse pink noises [52] with three different SNRs to evaluate the robustness of each method to disturbances. Also, we conducted the evaluation with six different signal lengths to examine the effect of the number of snapshots. Similarly to in Sect. 4.4.1, we use

Table 4.2: Computational complexity

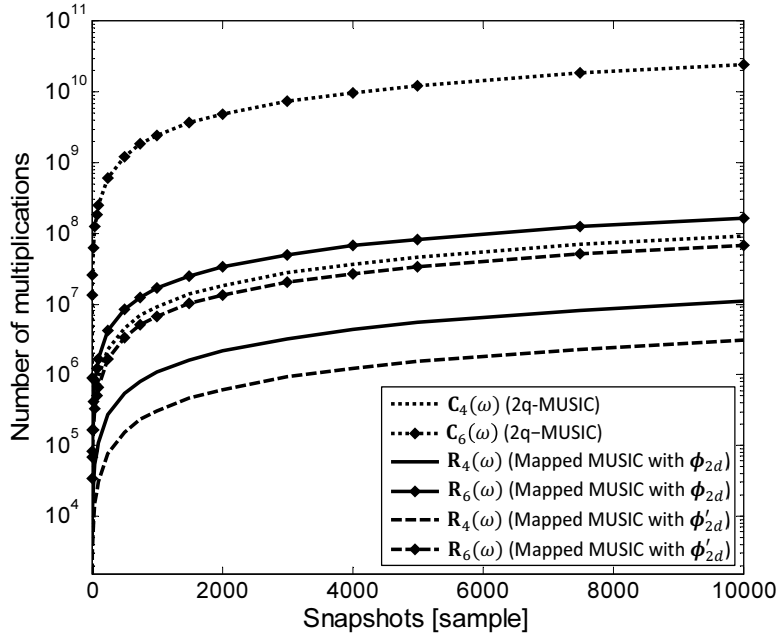| Matrix	Multiplication (times)
$\mathbf{C}_4(\omega)$	$12(3L+1)M^4$
$\mathbf{C}_6(\omega)$	$20(30L+11)M^6$
$\mathbf{R}_4(\omega)$ [with ϕ_{2d}]	$4L(M^2+M^4)$
$\mathbf{R}_6(\omega)$ [with ϕ_{2d}]	$4L(M^2+M^3+M^6)$
$\mathbf{R}_4(\omega)$ [with ϕ'_{2d}]	$L(M+3M^2+M^4)$
$\mathbf{R}_6(\omega)$ [with ϕ'_{2d}]	$L(3M^2+5M^3+M^4+2M^5+M^6)$
$\mathbf{R}_4(\omega)$ [entry-wise]	$12LM^4$
$\mathbf{R}_6(\omega)$ [entry-wise]	$20LM^6$

Figure 4.6: Transitions of the numbers of multiplications required to construct the statistical matrix with increasing number of snapshots. The number of microphones is four and the number of snapshots ranges from 0 to 10000. The proposed method has lower computational complexity, and using the computationally efficient map ϕ'_{2d} markedly reduces the computational complexity.

Table 4.4: Conditions for experiment using impulse response measured in real room

Microphone array	Circular array with radius of 0.1 m
# of microphones	4
Sound sources	Japanese speech signals from JNAS [50] emitted from speakers 1 m apart from array
# of sources	5
Room size	Studio 1: $3.4 \times 4 \times 2.7$ [m] Studio 2: $5.5 \times 9 \times 2.6$ [m]
Noise type	Observed diffuse noise
Reverberation (T_{60})	Studio 1: 0.3, Studio 2: 0.8 [s]
SNR	10,20 [dB]
Signal length	1,2,3,5,10,20 [s]
Sampling frequency	16 [kHz]
Frequency bandwidth	0.2–5 [kHz]
Frame length	512 samples
Frame shift length	256 samples
Window function	Hanning window

the number of sources as the dimensionality of the signal subspace for the proposed method and $2q$-MUSIC. Also DOA estimation with MUSIC is performed similarly to in Sect. 4.4.1 as a reference. We performed 1,000 trials of the estimation with randomly chosen combinations of source positions and obtained the RMSE over all trials. Table 4.3 shows the experimental conditions.

Also, to evaluate the estimation performance in a more realistic environment, we conducted an experiment using convolutive mixtures with the impulse responses measured in two different studios shown in Figs. 4.8 and 4.9. To mimic noisy observation, we observed diffuse noise with the same microphone array in a noisy room containing several computers and fans. Table 4.4 shows the experimental conditions.

Figures 4.10 and 4.11 show the results of the simulation experiment and the experiment with the measured impulse responses. Throughout these results, standard MUSIC does not perform appropriately because of the setup of underdetermined case, and SRP-PHAT also performs poorly because of its low resolution. To the contrary, both mapped MUSIC and $2q$-MUSIC show the effectiveness for the underdetermined DOA estimation also in these practical conditions. The estimation accuracies of mapped MUSIC and $2q$-MUSIC improve with increasing signal length and deteriorate
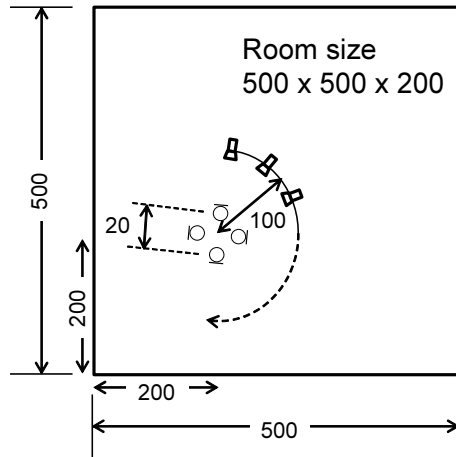
Figure 4.7: Environment for simulation experiment.

Table 4.3: Conditions for simulation experiment

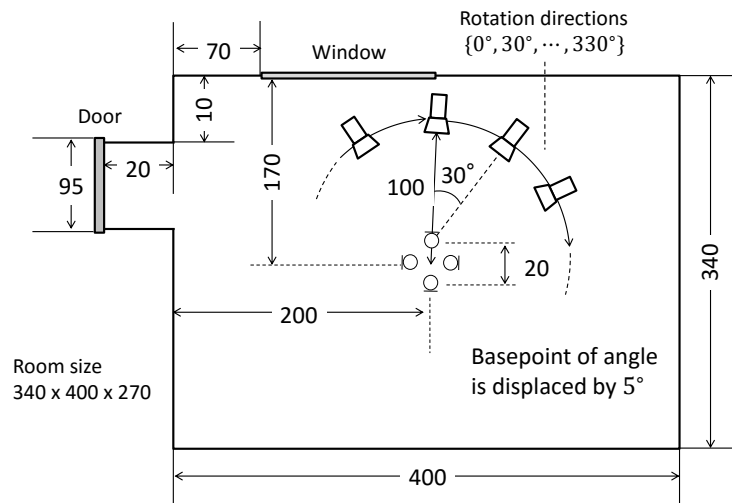|	
Microphone array	Circular array with radius of 0.1 m
# of microphones	4
Sound sources	Japanese speech signals from JNAS [50] emitted as point sources 1 m apart from array
# of sources	5
Room size	$5 \times 5 \times 2$ [m]
Noise type	Diffuse pink noise
Reverberation (T_{60})	0,150,500 [ms]
SNR	10,20 [dB]
Signal length	1,2,3,5,10,20 [s]
Sampling frequency	16 [kHz]
Frequency bandwidth	0.2–5 [kHz]
Frame length	512 samples
Frame shift length	256 samples
Window function	Hanning window

Figure 4.8: Recording environment of studio 1.
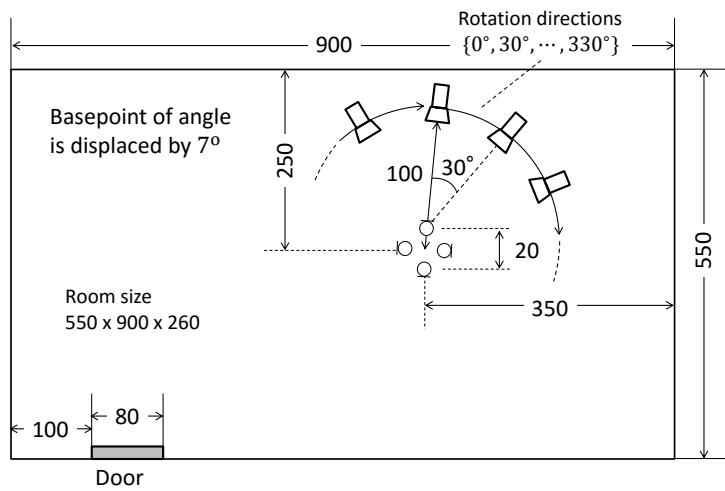
Figure 4.9: Recording environment of studio 2.

Table 4.5: Computational environment

CPU	Intel Corei7-3930K 3.2GHz
Memory	DDR3 64GB
OS	CentOS 6.6
Software	MATLAB R2010b

with increases in the reverberation time and noise. In most cases, mapped MUSIC performs better than $2q$-MUSIC. However, their difference becomes smaller with increasing signal length, and $2q$-MUSIC performs better for a signal length of 20 s in several cases. This behavior is consistent with the bias-variance tradeoff. These results ascertain the effectiveness of the proposed method for short-time analysis, which is particularly important for practical DOA estimation.

4.5.2   Evaluation of execution time

We also measured the execution times required to construct the statistical matrix, the moment matrix $\mathbf{R}_{2d}(\omega)$ in mapped MUSIC, and the cumulant matrix $\mathbf{C}_{2q}(\omega)$ in $2q$-MUSIC. Using the same conditions as in the DOA estimation experiments, we adopted four microphones and 4th- and 6th-order statistics. Using a system whose specifications are shown at Table 4.5, we conducted an evaluation by averaging 100 measurements for two signal lengths, 1 s and 10 s.

Figure 4.12 shows the number of multiplications and the practical execution time required for the construction of the statistical matrix. For both signal lengths, the proposed method has a shorter computation time than $2q$-MUSIC, although the computation time is not proportional to the number of multiplications. This is thought to be due to the slow loop processing of MATLAB. Although the measured speed strongly depends on the environment, this experiment indicates the lower computational complexity of the proposed method. Thus, mapped MUSIC is expected to be a useful tool for various applications requiring a small delay and rapid computation.

4.6   Conclusion

In this chapter, we proposed mapped MUSIC, a high-resolution DOA estimator for underdetermined conditions. We also discussed the properties of the mapping function with degree d used to analyze the $2d$th-order cross-moments and presented efficient algorithms with which the 4th- and 6th-order moments can be calculated. Furthermore, we compared the characteristics of the proposed method and conventional $2q$-MUSIC utilizing $2q$th-order cumulants. We demonstrated the advantageousness of the proposed method via an experiment. Mapped MUSIC is expected to be a suitable option
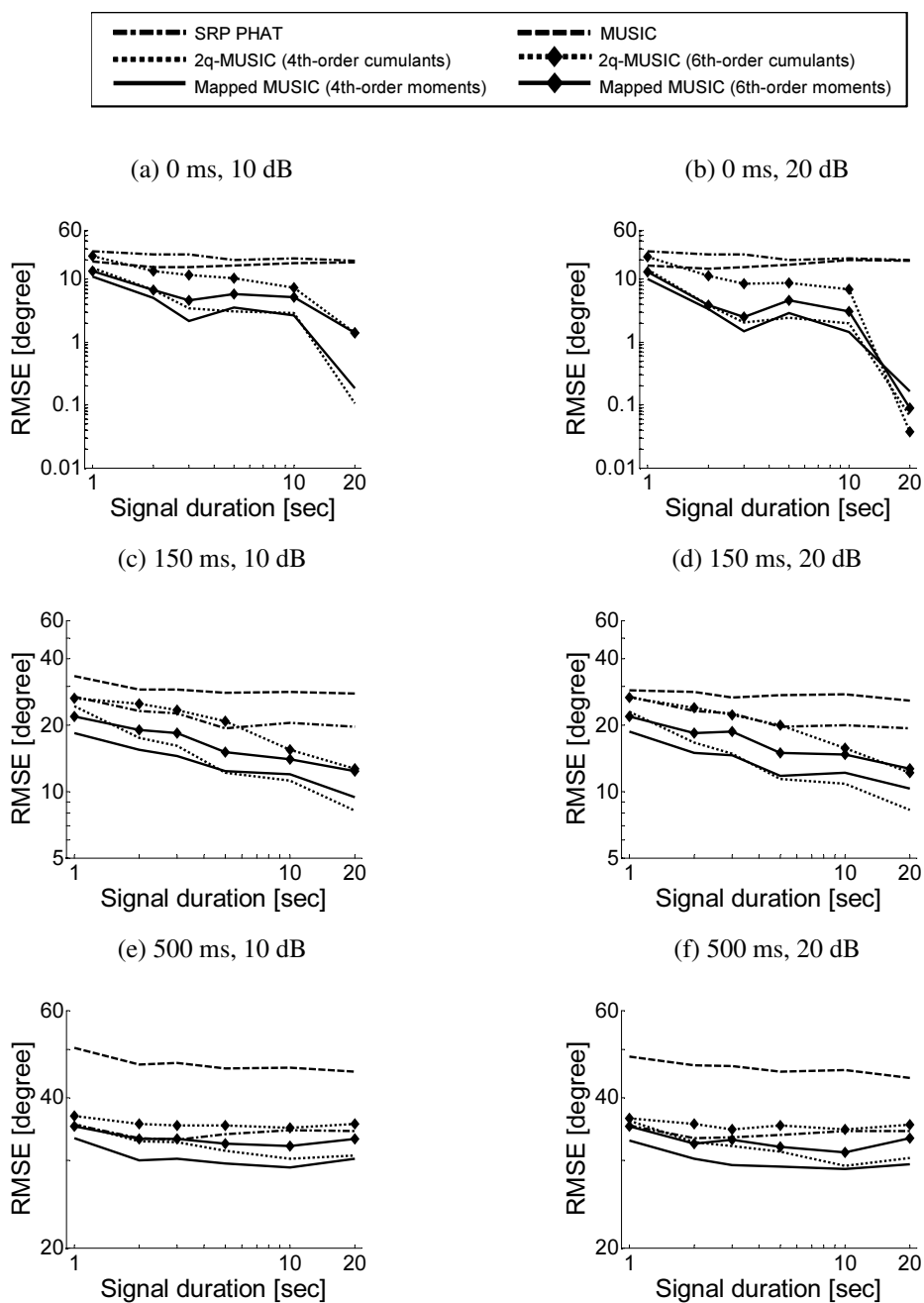
Figure 4.10: Results of simulation experiment. Each subcaption represents a combination of reverberation and SNR.
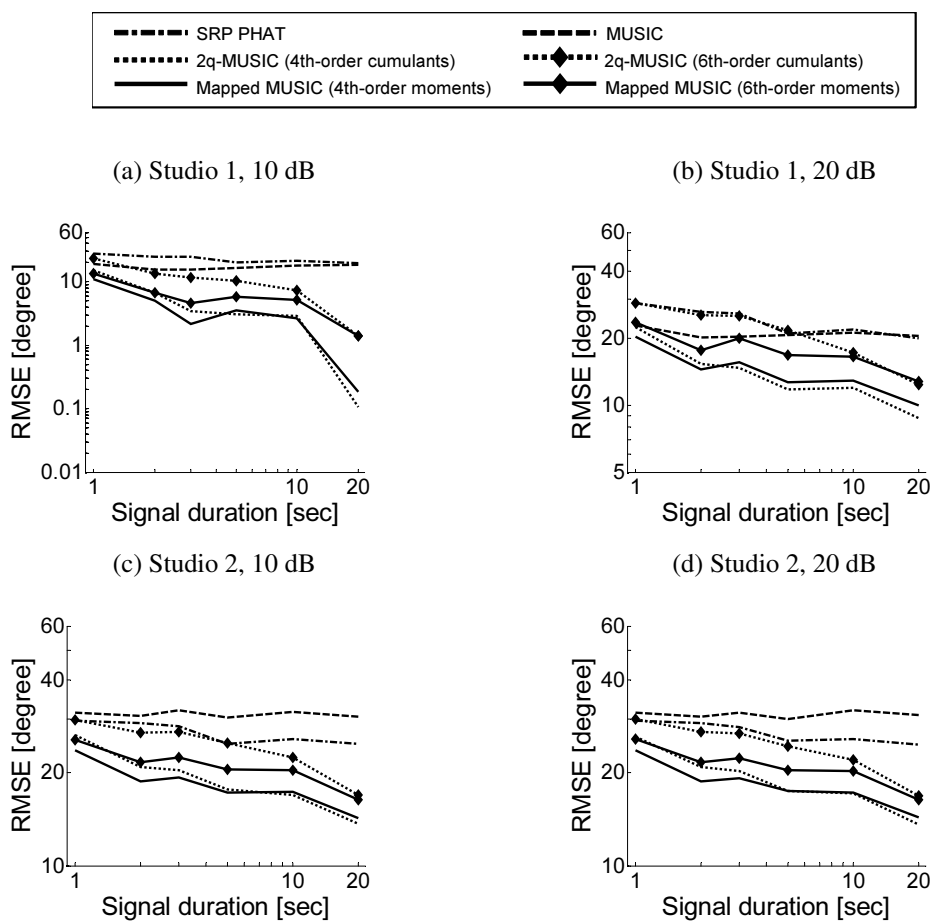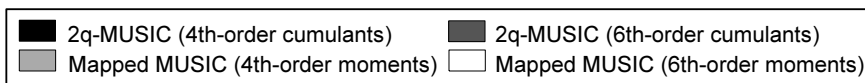
Figure 4.11: Results of experiment with the measured impulse response. Each subcaption represents a combination of recording room and SNR.

for many applications requiring short-time processing and rapid computation.
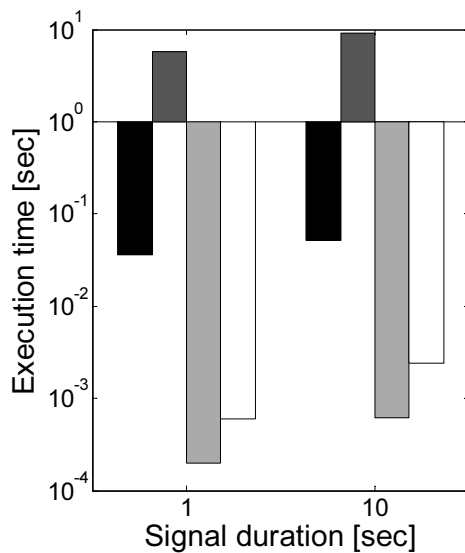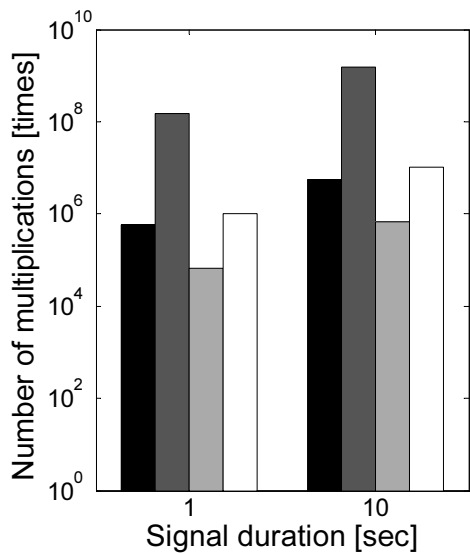
Figure 4.12: Number of multiplications and execution time.

Chapter 5

Extension of mapped MUSIC

This chapter proposes an extension of mapped MUSIC for joint analysis of the moments of multiple orders. By use of multiple maps of different orders simultaneously, this method achieves further improved estimation performance with the more extended subspace dimensionality. In this chapter, we describes the aim of this method and propose new map for joint analysis of moments of multiple orders. Experimental results of speech DOA estimation demonstrate that mapped MUSIC exploiting moments of multiple high orders achieves a higher accuracy than other MUSIC extensions. Section 5.1 describes an aim of joint analysis of moments of multiple high orders and introduce new map for that purpose. Section 5.2 conducts the experiment to evaluate efficiency of mapped MUSIC exploiting moments of multiple high orders. Section 5.3 concludes this chapter.

5.1 Joint analysis of cross moments of multiple orders

5.1.1 The aim of approach

In previous Chap. 4, the experimental results shows that 4th order moment analysis shows superior performance than that of 6th within a range of realistic time length, although an analysis of the higher order statistics yields the greater subspace dimensionality. These results suggest that the bad influence of bias is stronger than the beneficial increase of dimensionality when we raise the statistics order, and simply exploiting statistics of higher order is not a performance goal.

To overcome this shortcoming, here we propose a new mapping function for joint analysis of moments of multiple orders as the way to avoid the shortcoming and to further improve estimation performance. By evaluating moments of multiple orders at once, this approach aim to simultaneously realize the both advantages of the high and low orders, namely, high dimensionality and low variance. As we will show later in Sect. 5.2, this goal is found to be possible and can be accom-

plished by properly combining low-and-high-ordered moments in a nonlinearly expanded space.

5.1.2 Mapping $\phi_{d_1,\cdots,d_m}$ for joint analysis of moments

To utilize the advantages of high-resolution estimation with high-ordered moments and robust estimation with low-ordered moments, we propose the new map for joint analysis of cross moments of multiple orders. The proposed map $\phi_{d_1,\cdots,d_m}$ is given by a direct sum of maps $\phi_{d_1},\cdots,\phi_{d_m}$, proposed in Sect. 4.2, as

$$\phi_d\left(\mathbf{x}(\omega,t)\right) = \begin{cases} \mathbf{x}(\omega,t) & (d=1) \\ \mathbf{x}(\omega,t) \otimes \phi^*_{d-1}\left(\mathbf{x}(\omega,t)\right) & (d \geq 2) \end{cases}$$

$$= \left[\prod_{l=1}^{d} x^{\oplus}_{o_{1,l}}(\omega,t), \cdots, \prod_{l=1}^{d} x^{\oplus}_{o_{M^d,l}}(\omega,t)\right]^T, \tag{5.1}$$

$$x^{\oplus}_{o_{i,l}}(\omega,t) \triangleq \begin{cases} x_{o_{i,l}}(\omega,t) & (\text{if } l \text{ is odd}) \\ x^*_{o_{i,l}}(\omega,t) & (\text{if } l \text{ is even}) \end{cases}, \tag{5.2}$$

$$\phi_{d_1,\cdots,d_m} \triangleq \phi_{d_1} \odot \cdots \odot \phi_{d_m}, \tag{5.3}$$

$$\{\phi_{d_1},\cdots,\phi_{d_m}\} = \{\phi_d \mid d = d_1,\cdots,d_m\}$$

$$d_{i+1} > d_i \quad (i = 1,\cdots,m-1), \tag{5.4}$$

where $\odot$ denotes a direct sum. Mapped MUSIC with the map $\phi_{d_1,\cdots,d_m}$ gives joint analysis of multiple moments because the covariance matrix $\mathbf{R}^{\odot}_{d_1,\cdots,d_m}(\omega)$ of $\phi_{d_1,\cdots,d_m}$ contains the moments of multiple orders as

$$\mathbf{R}^{\odot}_{d_1,\cdots,d_m}(\omega) = E\left[\phi_{d_1,\cdots,d_m}(\mathbf{x}'(\omega,t))\phi^H_{d_1,\cdots,d_m}(\mathbf{x}'(\omega,t))\right]$$

$$= \begin{bmatrix} \mathbf{R}_{d_1,d_1}(\omega) & \mathbf{R}_{d_1,d_2}(\omega) & \cdots & \mathbf{R}_{d_1,d_m}(\omega) \\ \mathbf{R}_{d_2,d_1}(\omega) & \mathbf{R}_{d_2,d_2}(\omega) & \cdots & \mathbf{R}_{d_2,d_m}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{d_m,d_1}(\omega) & \mathbf{R}_{d_m,d_2}(\omega) & \cdots & \mathbf{R}_{d_m,d_m}(\omega) \end{bmatrix}, \tag{5.5}$$

$$\mathbf{R}_{d_i,d_j}(\omega) \triangleq E\left[\phi_{d_i}(\mathbf{x}'(\omega,t))\phi^H_{d_j}(\mathbf{x}'(\omega,t))\right] = [r_{pq}]_{pq}$$

$$(i,j = 1,\cdots,m \quad p = 1,\cdots,M^{d_i}$$

$$q = 1,\cdots,M^{d_j}), \tag{5.6}$$

$$r_{pq} = E\left[\left(\prod_{l=1}^{d_i} x'^{\otimes}_{o_{p,l}}\right)\left(\prod_{l=1}^{d_j} x'^{\otimes}_{o_{q,l}}\right)^*\right], \tag{5.7}$$

where $\mathbf{x}'(\omega, t)$ is an observed signal normalized properly (normalization rule is described later), and submatrices $\mathbf{R}_{d_i,d_j}(\omega)$ gives $(d_i + d_j)$th order cross moments. The dimensionality of $\mathbf{R}_{d_1,\cdots,d_m}^{\circledcirc}(\omega)$ is $\sum_{i=1}^{m} M^{d_i}$, which is still larger than that of the highest-dimensional single cross moment matrix $\mathbf{R}_{d_m,d_m}(\omega)$. In addition, by analyzing the low-ordered moments together with the high-ordered moments, the significance of the statistical bias in the analysis of high-ordered moments is relaxed. With the increased dimensionality and reduced significance of statistical bias, the mapped MUSIC with the proposed map $\phi_{d_1,\cdots,d_m}$ achieves further improvement of DOA estimation performance.

Note that we must take care of the scaling of the observed signal for the sufficient joint analysis of the moments of multiple orders. As we find in (5.5)–(5.7), $\mathbf{R}_{d_1,\cdots,d_m}^{\circledcirc}(\omega)$ contains moments of various orders, and it is obvious that the absolute values of the elements in the high-ordered moment matrices becomes significantly larger than those of the low-ordered moment matrices when the norm $||\mathbf{x}(\omega, t)||$ of the observation is large. In contrast, the absolute values of the elements in the high-ordered moments becomes considerably small when the norm $||\mathbf{x}(\omega, t)||$ is small. Therefore, if we utilize original observations as inputs of the mapping, respective analyses in each frequency bin behave differently because of the difference of the norm. For the sufficient joint analysis of moments of multiple orders, we must uniformly take into accounts all frequency bins and all the orders of moments must be weighted appropriately. As such a way of weighting, we normalize the observed signal with the L_{2d_m}-norm based on the maximum degree d_m within the proposed map $\phi_{d_1,\cdots,d_m}$ in each frequency bin as

$$\mathbf{x}'(\omega, t) = \frac{w\mathbf{x}(\omega, t)}{\sqrt[2d_m]{E\left[\frac{1}{M}\sum_{i=1}^{M}|x_i(\omega, t)|^{2d_m}\right]}}, \tag{5.8}$$

where w is a parameter to adjust the L_{2d_m}-norm of the observed signal. With large value of w, the proposed method analyzes the high-ordered moments more significantly, and vice versa.

5.2 Evaluation experiments of DOA estimation accuracy

To explore the possibility of joint analysis of moments, this section conduct experiments to evaluate DOA estimation accuracy. We evaluated mapped MUSIC with the proposed map $\phi_{1,2,3}$ and following conventional methods, MUSIC, mapped MUSIC ($d = 2, 3$) and $2q$-MUSIC ($q = 2, 3$) for comparison.

5.2.1 Experimental condition

We conducted an evaluation of the DOA estimation of each method by using mixed signals that were created as convolutive mixtures of the Japanese speeches and the impulse responses, measured at
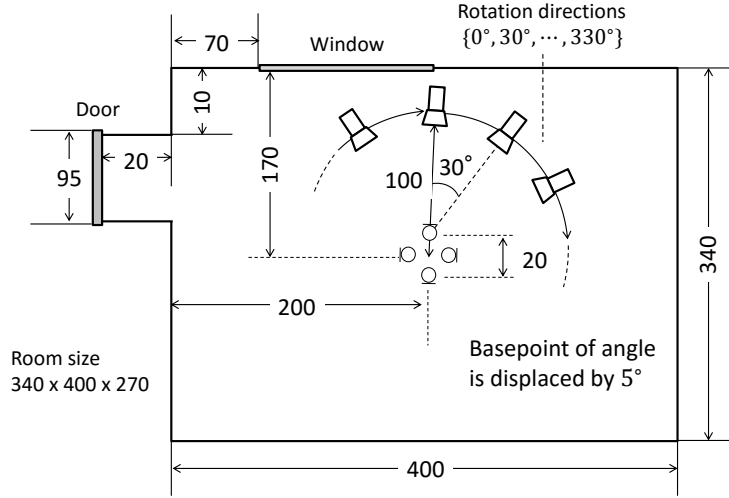
Figure 5.1: Recording environment.

the environment shown in Fig.5.1. To emulate noisy observation, diffused pink noises [52] was superimposed to the observed signals. To verify estimation performance for both overdetermined and underdetermined environments, we also evaluated the performance with two different number of sources. Table 5.1 shows the other experimental conditions.

For the evaluation, we employed the root mean squared error (RMSE) between the estimated direction and the true sound direction. We evaluated 100 combinations of the positions of three or five sources selected randomly from the directions $\{0°, 30° \cdots, 330°\}$. When the estimation score had fewer peaks than sources, we added a penalty equal to the average error for all directions. This experiment employed the signal subspace dimensionality parameter of mapped MUSIC $K = 15$ as the number of source is three, and $K = 30$ for five sources. We also employed normalization parameter $w = 2$ in (5.8) for both of the number of sources. DOA estimation with MUSIC when $M \leq N$ is performed by regarding the one-dimensional subspace associate with the minimum eigenvalue as the noise subspace in every frequency bin.

5.2.2 Experimental results

Figures 5.2–5.3 show the experimental results obtained under different number of sources. Throughout all the conditions, the proposed method performs the best, and mapped MUSIC performs slightly better than $2q$-MUSIC. When we compare the results of mapped MUSIC ($d = 2$) and that of ($d = 3$), the former shows the better performance, and this tendency also appears as to $2q$-MUSIC. As examined in Sect. 4.4, these results suggest that the latter is biased greater than the former be-

Table 5.1: Experimental conditions

Microphone array	Circular array with radius of 0.1 m
# of microphones	4
Sound sources	Japanese speech signals from JNAS [50] emitted from speakers 1 m apart from array
# of sources	3, 5
Room size	$3.4 \times 4 \times 2.7$ [m]
Noise type	Diffused pink noise
Reverberation (T_{60})	0.3 [s]
SNR	20 [dB]
Signal length	1 [s]
Sampling frequency	16 [kHz]
Frequency bandwidth	0.2–5 [kHz]
Frame length	512 samples
Frame shift length	256 samples
Window function	Hanning window

cause of the shortage of snapshots, and the observation of one second is not enough to reduce the influence of statistical bias derived from the high-ordered statistics. While, the comparison between the proposed method and mapped MUSIC ($d = 2$) denotes the effectiveness of joint analysis of moments of multiple orders. Although the high-ordered moments is inferior as for the sole use, we can confirm its usefulness to support the low-ordered moments. Furthermore, the proposed method can be expected to keep showing the best performance even if the time length becomes larger because the information of the high-ordered becomes more robust. From these discussion, the effectiveness of the proposed method utilizing the advantage of low-ordered and high-ordered moments simultaneously is verified.

5.3 Conclusion

This chapter proposed extended mapped MUSIC realizes to estimate DOAs with higher resolution than other conventional MUSIC extensions. We proposed new map as a direct sum of maps of multiple degrees, and showed that the proposed method corresponds to joint moments analysis of multiple orders taking advantage of both resolution with high-ordered moments and robustness with
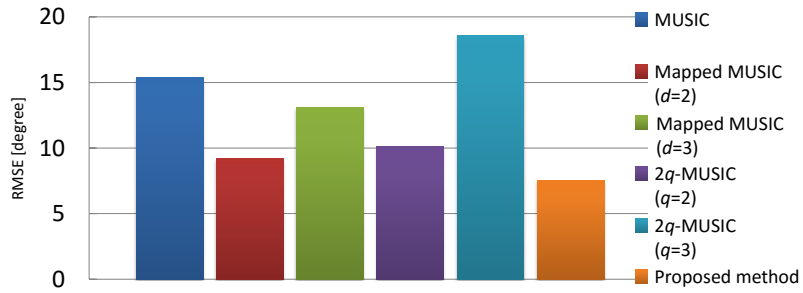
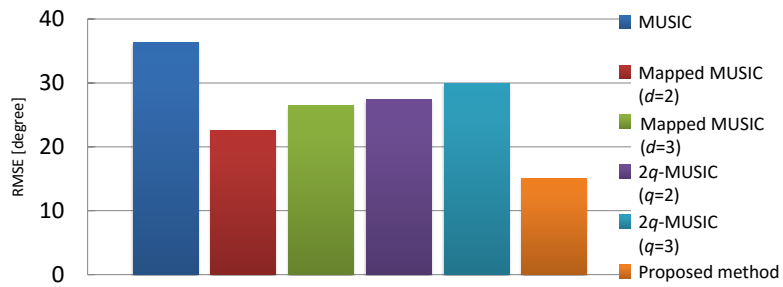Figure 5.2: Experimental results for 3 sources within 1 second.

Figure 5.3: Experimental results for 5 sources within 1 second.

low-ordered moments. The experiment comparing with the other MUSIC extensions based on single high order statistics clarified the effectiveness of the joint analysis and excellent DOA estimation performance by the proposed method.

Chapter 6

Conclusion

6.1 Thesis summary

This dissertation describes our study for underdetermined DOA estimation with high resolution.

Chapter 1 described background and purpose of this study.

Chapter 2 described problem statement of the DOA estimation of sound sources using microphone array. We showed the observation signal model assumed in this dissertation. Subsequently, we described specific estimation procedure using steering vector.

Chapter 3 reviewed conventional methods. We described estimation algorithms of delay-and-sum beamformer and its extensions based on inter-channel correlation function. We also described estimation algorithm of MUSIC and $2q$-MUSIC based on subspace method using eigenvalue analysis.

Chapter 4 proposed mapped MUSIC, our first work for this study. We described estimation algorithm of mapped MUSIC based on MUSIC-like analysis with the nonlinear map. Subsequently, we proposed the suitable map for the analysis of moments of arbitrary even order. Through the comparison with $2q$-MUSIC, we showed a superiority of mapped MUSIC for short-time analysis assuming practical estimation environment.

Chapter 5 proposed extension of mapped MUSIC, our second work for this study. To further improve the DOA estimation accuracy, we proposed new mapping function for joint analysis of moments of multiple orders. Through the experimental of DOA estimation using speech signals, we showed the efficiency of this method which combines both advantages of the high and low orders of statistics.

In this dissertation, we present mapped MUSIC for the underdetemined DOA estimation along the acoustic signal scenario, and they achieves superior estimation performance than the conventional DOA estimator. However, our concept is proposed for DOA estimation with arbitrary sensor array

and general array signal processing, and the tendency of analysis using higher-dimensional mapping can be changed easily with the mapping function. Therefore, the concept has large possibility to be introduced for wide variety of applications by designing the suitable map, and this study contributes for the development of array signal processing.

6.2   Future direction

While an estimation of the number of sources is an important aspect of DOA estimation, this dissertation avoids the discussion and evaluate DOA estimation performance under the assumption that the true number of sources is given in advance. This is the considerable shortcoming in this dissertation, and we need to propose an idea to automatically determine the proper dimensionality of subspace for a more practical usage. For the purpose, we assume thresholding, based on the estimated eigenvalue distribution in the covariance matrix of the mapping, is effective. W e are working on a statistical approach assuming probability distribution of source amplitude to estimate the covariance matrix of the mapping.

Also, with regard to the our second work, we confirmed that we can achieve the higher estimation accuracy by changing the combination of multiple orders of moments. This result suggest that the suitable combination ought to be selected for the better estimation according to the signal property and recording condition, such as signal time length. To propose the criterion for the choice of the proper combination is another our future task.

Appendix A

A.1 Kernel MUSIC

In the field of pattern recognition, kernel method is a popular approach, which enhances the linear separativity of classifiers with the nonlinear mapping to the higher-dimensional Hilbert space given by kernel function. Here we describe an alternative expression of mapped MUSIC, named kernel MUSIC that utilizes the kernel method.

A.1.1 Formulation of kernel MUSIC for generic kernel

It is well known that positive-definite kernel : $\mathbb{C}^M \times \mathbb{C}^M \to \mathbb{C}$ expresses the inner product on the Hilbert space $\mathcal{H}$ spanned by the mapping $\phi : \mathbb{C}^M \to \mathcal{H}$ as

$$k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle_{\mathcal{H}}, \tag{A.1}$$

where $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$ and $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the inner product on the Hilbert space $\mathcal{H}$. The kernel function is often used in the classification problem to enhance linear separability of the feature vectors, because the Hilbert space $\mathcal{H}$ given by the non-linear kernel function generally has higher dimension than $\mathbb{C}^M$. By replacing the eigenvalue problem of covariance matrix $\mathbf{C}_2(\omega) = E[\mathbf{x}(\omega, t)\mathbf{x}^{\mathrm{H}}(\omega, t)]$ in MUSIC with the eigenvalue problem of covariance linear operator $\mathbf{\Gamma}_\omega : \mathcal{H} \to \mathcal{H}, \mathbf{\Gamma}_\omega = E[\langle \cdot, \phi(\mathbf{x}(\omega, t)) \rangle_{\mathcal{H}} \langle \phi(\mathbf{x}(\omega, t)), \cdot \rangle_{\mathcal{H}}]$ on the Hilbert space $\mathcal{H}$ and steering vector $\mathbf{b}(\omega; \theta)$ with its mapping $\phi(\mathbf{b}(\omega; \theta))$, kernel MUSIC with the higher resolution of DOA estimation is obtained.

Since mapping function ϕ cannot be described analytically in general, covariance linear operator $\mathbf{\Gamma}_\omega$ cannot as well. However, as utilized in kernel PCA algorithm [53], the eigenvalues $\lambda_i(\omega)$ and the eigenvectors $\boldsymbol{v}_i(\omega) \in \mathcal{H}$ of $\mathbf{\Gamma}_\omega$ in the following eigenvalue problem,

$$\mathbf{\Gamma}_\omega(\boldsymbol{v}_i(\omega)) = \lambda_i(\omega)\boldsymbol{v}_i(\omega),$$
$$\langle \boldsymbol{v}_i(\omega), \boldsymbol{v}_j(\omega) \rangle_{\mathcal{H}} = \delta_{ij}, \quad \lambda_1(\omega) \geq \lambda_2(\omega) \geq \cdots \geq 0, \tag{A.2}$$

48

(δ_{ij} is Kronecker delta) can be expressed by the eigen decomposition of the $L \times L$ Grammian matrix $\mathbf{K}(\omega)$ with the observed signals $\mathbf{x}(\omega, t), t = 1, \cdots, L$ as

$$\lambda_i(\omega) = d_i(\omega)/L, \tag{A.3}$$

$$\boldsymbol{v}_i(\omega) = \frac{1}{\sqrt{d_i(\omega)}} \sum_{j=1}^{L} u_{ji}(\omega)\boldsymbol{\phi}(\mathbf{x}(\omega, j)), \tag{A.4}$$

$$\mathbf{K}(\omega) = [k(\mathbf{x}(\omega, i), \mathbf{x}(\omega, j))]_{ij} = \mathbf{U}(\omega)\mathbf{D}(\omega)\mathbf{U}^H(\omega), \tag{A.5}$$

$$\mathbf{U}(\omega) = [u_{ji}(\omega)]_{ji}, \quad \mathbf{U}^H(\omega)\mathbf{U}(\omega) = \mathbf{I}_L, \tag{A.6}$$

$$\mathbf{D}(\omega) = \mathrm{diag}\,[d_1(\omega), \cdots, d_L(\omega)],$$

$$d_1(\omega) \geq \cdots \geq d_L(\omega) \geq 0. \tag{A.7}$$

Denoting the rank of $\mathbf{K}(\omega)$ as $r(\omega)$, DOA evaluation function $g_{\mathcal{H}}(\omega)$ in the Hilbert space $\mathcal{H}$ is given as following because $\mathrm{span}[\upsilon_i(\omega)]_{i=N+1}^{r(\omega)}$ corresponds to noise subspace:

$$\begin{aligned} g_{\mathcal{H}}(\omega) &\triangleq \frac{1}{\sum_{i=N+1}^{r(\omega)} |\langle \boldsymbol{\phi}(\mathbf{b}(\omega; \theta)), \boldsymbol{v}_i(\omega)\rangle_{\mathcal{H}}|} \\ &= \frac{1}{\sum_{i=N+1}^{r(\omega)} \left| \sum_{j=1}^{L} \frac{u_{ji}(\omega)}{\sqrt{d_i(\omega)}} k\,(\mathbf{b}(\omega; \theta), \mathbf{x}(\omega, j)) \right|}. \end{aligned} \tag{A.8}$$

A.1.2 Kernel function corresponding to map ϕ_d

For the sake of ease, subsequent discussion omit ω and t from the description of observed signals $\mathbf{x}(\omega, t)$. Since kernel MUSIC analyzes correlation in the Hilbert space $\mathcal{H}$, its property is altered drastically by the choice of the mapping ϕ. Here we take up analyzing arbitrary even order moments, with kernel MUSIC described in Section A.1.1, by choosing kernel function $k_d(\cdot, \cdot)$ so that it corresponds to the inner products of the map ϕ_d proposed in Chap. 4. Then the kernel function k_d ought to be expressed as following equation,

$$\begin{aligned} k_d(\mathbf{x}, \mathbf{y}) &= <\phi_d(\mathbf{x}), \phi_d(\mathbf{y})> \\ &= \begin{cases} |\mathbf{x}^H\mathbf{y}|^{d-1}\mathbf{x}^H\mathbf{y} & \text{(if } d \text{ is odd)} \\ |\mathbf{x}^H\mathbf{y}|^d & \text{(if } d \text{ is even)} \end{cases}. \end{aligned} \tag{A.9}$$

k_d is positive-definite kernel and reproduces the higher dimensional Euclidian space $\mathbb{C}^{M^d - \mathrm{null}[\phi_d]}$. From the equality to inner product of the map ϕ_d, kernel MUSIC with k_d obtains identical estimation results to mapped MUSIC analyzing covariance matrix of the explicitly described map ϕ_d. In regard to computational complexity, whereas mapped MUSIC requires vast amount of computation in a more higher order analysis, kernel MUSIC shown in Eqs. (A.3)–(A.8) has a merit that

computational complexity of DOA estimation is mainly dominated by the number of snapshots L $(t = 1, \cdots , L)$ and remains almost constant for all order of moment analysis.

Appendix B

B.1 Additional experimental results in Sect. 4.4.1

In Sect. 4.4.1, we show the experimental result of error transition using CGGD. For the further detailed investigation, this section shows additional experimental results using different probability distributions to generate pseudorandom numbers. Table B.1 are the lists of probability distributions, in which α denotes scale parameter, β denotes shape parameter, and $\Gamma(\cdot)$ denotes gamma function. Likewise CGGD with shape parameter $\beta = 0.315$ that gives similar property to the human speech, we determined the parameters of each probability distribution to fit the complex distributions of human speech.

Figures B.1–B.2 are the results of error transitions obtained from narrowband of 3 kHz. Also in these case using different probability distributions, we can confirm bias-variance tradeoff.

Table B.1: Probability distributions utilized in the experiment.

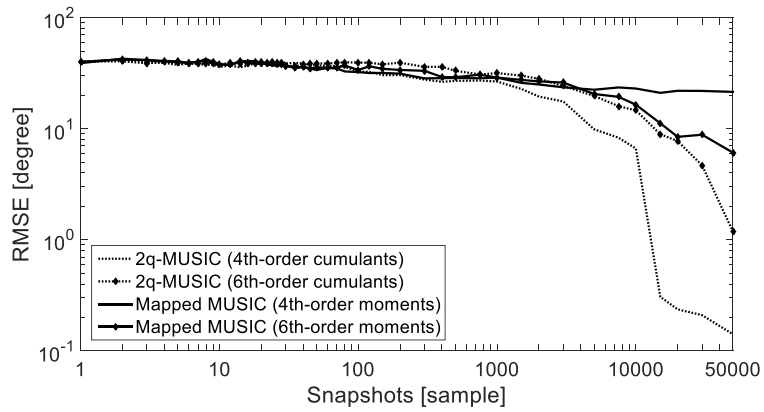| Distribution type	Density function	Kurtosis
Complex gamma	$p(s) = \dfrac{s^{\beta-1}\exp\left(-\frac{s}{\alpha}\right)}{2\pi\Gamma(\beta)\alpha^{\beta}}$	$\dfrac{1}{\beta} - 1$
Weibull	$p(s) = \dfrac{\beta|s|^{2\beta-2}\exp\left(-\frac{|s|^{2\beta}}{\alpha^{\beta}}\right)}{\pi\alpha^{\beta}}$	$\dfrac{2\beta\Gamma\left(\frac{2}{\beta}\right)}{\Gamma\left(\frac{1}{k}\right)^{2}} - 2$

Figure B.1: Error transition with increasing number of snapshots. Sources are given as complex gamma distribution with $\beta = 0.5$ and $\alpha = 1$.
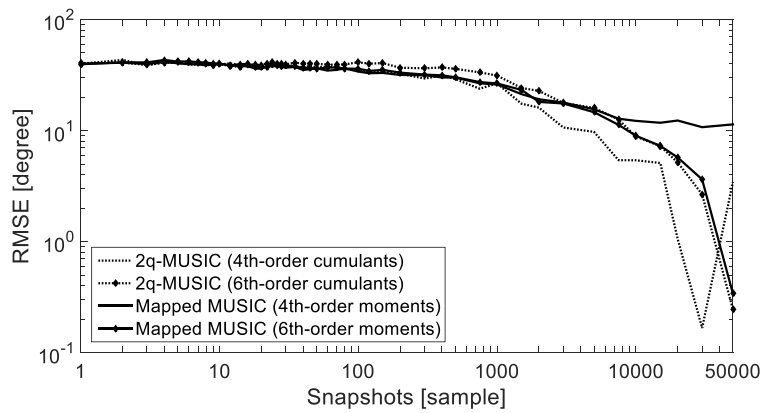

Figure B.2: Error transition with increasing number of snapshots. Sources are given as Weibull distribution with $\beta = 0.3$ and $\alpha = 1$.

B.2   Derivation of alternative mapping

In Sect. 4.3, we present alternative maps for the efficient computation, ϕ_2':

$$\phi_2'(\mathbf{x}) \triangleq \left[\phi_{\mathrm{abs}}'^T(\mathbf{x}), \phi_{\mathrm{re}}'^T(\mathbf{x}), \phi_{\mathrm{im}}'^T(\mathbf{x}) \right]^T, \tag{B.1}$$

$$\phi_{\mathrm{abs}}'(\mathbf{x}) \triangleq [\forall |x_i|^2 | 1 \le i \le M]^T, \tag{B.2}$$

$$\phi_{\mathrm{re}}'(\mathbf{x}) \triangleq \sqrt{2}[\forall \mathrm{Re}[x_i x_j^*] | 2 \le i \le M, 1 \le j \le i-1]^T, \tag{B.3}$$

$$\phi_{\mathrm{im}}'(\mathbf{x}) \triangleq \sqrt{2}[\forall \mathrm{Im}[x_i x_j^*] | 2 \le i \le M, 1 \le j \le i-1]^T, \tag{B.4}$$

and ϕ_3':

$$\phi_3'(\mathbf{x}) \triangleq \left[\phi_{3a}'^T(\mathbf{x}), \phi_{3b}'^T(\mathbf{x}), \phi_{3c}'^T(\mathbf{x}), \phi_{3d}'^T(\mathbf{x}) \right]^T, \tag{B.5}$$

$$\phi_{3a}'(\mathbf{x}) \triangleq [\forall |x_i|^2 x_i^* | 1 \le i \le M]^T, \tag{B.6}$$

$$\phi_{3b}'(\mathbf{x}) \triangleq [\forall x_i x_j^{*2} | 1 \le i, j \le M, j \ne i]^T, \tag{B.7}$$

$$\phi_{3c}'(\mathbf{x}) \triangleq \sqrt{2}[\forall |x_i|^2 x_j^* | 1 \le i, j \le M, j \ne i]^T, \tag{B.8}$$

$$\phi_{3d}'(\mathbf{x}) \triangleq \sqrt{2}[\forall x_i x_j^* x_k^* | 1 \le i, j, k \le M]^T,$$
$$(i, j, \text{ and } k \text{ are different}), \tag{B.9}$$

whose inner product is the identical to that of the map ϕ_d,

$$\phi_d(\mathbf{x}) \triangleq \begin{cases} \mathbf{x} & (d = 1) \\ \mathbf{x} \otimes \phi_{d-1}^*(\mathbf{x}) & (d \ge 2)x \end{cases}. \tag{B.10}$$

Here we confirm the equality of the inner product between these mappings. For example, when

$d = 2$,

$$\phi_2^H(\mathbf{x})\phi(\mathbf{y}) = |x^*y|^2$$

$$= \mathbf{y}^H\mathbf{x}\mathbf{x}^H\mathbf{y}$$

$$= \sum_{i=1}^{M} y_i^* x_i \sum_{j=1}^{M} x_j^* y_j$$

$$= \sum_{i=1}^{M} |x_i^* y_i|^2 + \sum_{i=1}^{M} \sum_{j \neq i}^{M} y_i^* x_i x_j^* y_j$$

$$= \sum_{i=1}^{M} |x_i^*|^2 |y_i|^2 + 2\text{Re}\left[\sum_{i=2}^{M} \sum_{j=1}^{i-1} y_i^* x_i x_j^* y_j\right]$$

$$= \sum_{i=1}^{M} |x_i|^2 |y_i|^2 + 2\sum_{i=2}^{M} \sum_{j=1}^{i-1} \text{Re}\left[x_i x_j^*\right] \text{Re}\left[y_i^* y_j\right] - 2\sum_{i=2}^{M} \sum_{j=1}^{i-1} \text{Im}\left[x_i x_j^*\right] \text{Im}\left[y_i^* y_j\right]$$

$$= \sum_{i=1}^{M} |x_i|^2 |y_i|^2 + 2\sum_{i=2}^{M} \sum_{j=1}^{i-1} \text{Re}\left[x_i x_j^*\right] \text{Re}\left[y_i y_j^*\right] - 2\sum_{i=2}^{M} \sum_{j=1}^{i-1} \text{Im}\left[x_i x_j^*\right] \text{Im}\left[-y_i y_j^*\right]$$

$$= \sum_{i=1}^{M} |x_i|^2 |y_i|^2 + 2\sum_{i=2}^{M} \sum_{j=1}^{i-1} \text{Re}\left[x_i x_j^*\right] \text{Re}\left[y_i y_j^*\right] + 2\sum_{i=2}^{M} \sum_{j=1}^{i-1} \text{Im}\left[x_i x_j^*\right] \text{Im}\left[y_i y_j^*\right]$$

$$= \phi_{\text{abs}}^{'T}(\mathbf{x})\phi_{\text{abs}}^{'}(\mathbf{y}) + \phi_{\text{re}}^{'T}(\mathbf{x})\phi_{\text{re}}^{'}(\mathbf{y}) + \phi_{\text{im}}^{'T}(\mathbf{x})\phi_{\text{im}}^{'}(\mathbf{y})$$

$$= \phi_2^{'T}(\mathbf{x})\phi_2^{'}(\mathbf{y}). \tag{B.11}$$

We omit the derivation when $d = 3$, but a equivalence of the inner product between ϕ_3 and $\phi_3^{'}$ is also proofed in the same manner as (B.11).

Acknowledgements

References

[1] H. L. Van Trees, "Detection, Estimation, and Modulation Theory, Optimum Array Processing", Wiley-Interscience, 2004.

[2] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach", *IEEE Trans. Signal Process. Mag.*, vol.13, no.4, pp. 67–94,1996.

[3] R. Zekavat and R. M. Buehrer, "Handbook of Position Location: Theory, Practice and Advances", John Wiley and Sons, 2011.

[4] P. S. Naidu, "Sensor Array Signal Processing", CRC Press, 2010.

[5] Z. Chen, G. Gokeda, and Y. Yu, "Introduction to Direction-of-Arrival Estimation", *Artech House*, 2010.

[6] M. Brandstein and D. Ward, eds., "Microphone arrays: signal processing techniques and applications", *Springer*, 2001.

[7] E. J. Hannan and P. J. Thomson, "The estimation of coherence and group delay", *Biometrika*, vol.58, pp.469–481, 1971.

[8] V. H. MacDonald and P. M. Schultheiss, "Optimum passive bearing estimation", *J. Acoust. Soc. America*, vol.46, pp.37–43, 1969.

[9] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay", *IEEE Trans. Acoust, Speech Signal Process.*, vol.24, pp.320–327, 1976.

[10] H. F. Silverman, *et al.*, "Performance of real-time source-location estimators for a large-aperture microphone array", *IEEE Trans. Speech Audio Process.*, vol.4, no.13, pp. 593–606, 2005.

[11] J. Capon, "High resolution frequency-wave number spectral analysis", *Proc. IEEE*, vol.57, pp.1408–1518, 1969.

[12] J. Makhoul, "Linear prediction: A tutorial review", *Proc. IEEE*, vol.63, no.4, pp.561–580, 1975.

[13] Y. Ogawa and N. Kikuma, "High-Resolution Techniques in Signal Processing Antennas", *IEICE Trans. Commun.*, vol.E78-B, No.11, pp.1435–1442, 1995.

[14] W. F. Gabriel, "Spectral Analysis and Adaptive Array Supperresolution Techniques", *Proc. IEEE*, vol.68, No.6, pp.654–666, 1980.

[15] S. U. Pillai, "Array Signal Processing", Springer-Verlag New York Inc, 1989.

[16] S. S. Reddi, "Multiple Source Location A Digital Approach", *IEEE Transs. on AES*, Vol.15, No.1, 1979.

[17] R. Kumaresan and D. Tufts, "Estimating the Angles of Arrival of Multiple Plane Waves", *IEEE Trans. on AES*, Vol.AES-19, pp.134–139, 1983.

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

[19] R. Roy and T. Kailath, "ESPRIT - Estimation of Signal Parameter via Rotational Invariance Techniques", *IEEE Trans. on ASSP*, vol.37, No.7, pp.984–995, 1989.

[20] B. Porat and B. Friedlander, "Direction finding algorithms based on higher-order statistics", *IEEE Trans. Signal Process.*, vol.39, no.9, pp.2016–2024, 1991.

[21] P. Chevalier, A. Ferreol and L. Albera, "High resolution direction finding from higher-order statistics: The $2q$-MUSIC algorithm", *IEEE Trans. Signal Process.*, vol.54, no.8, pp.2986–2997, 2006.

[22] P. Chevalier, *et al.*, "On the Virtual Array Concept for Higher Order Array Processing", *IEEE Trans. Signal Process.*, vol.53, no.4, pp.1254–1271, 2005.

[23] G. Chardon, *et al.*, "Near-field acoustic holography using sparse regularization and compressive sampling principles", *J.Acoust. Soc. Am.*, vol.132, no.3, pp. 1521-1534, 2012.

[24] S. Koyama, S. Shimauchi, and H. Ohmuro, "Sparse sound field representation in recording and reproduction for reducing spatial aliasing artifacts", *Proc. ICASSP*, pp.4476-4480, 2014.

[25] D. Malioutov, M. C. etin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays", *IEEE Trans. Signal Process.*, vol.53, no.8, pp. 3010-3022, 2005.

[26] A. Asaei, *et al.*, "Modelbased sparse component analysis for reverberant speech localization", *Proc. ICASSP*, pp.1453-1457, 2014.

[27] T. Noohi, N. Epain, and C. T. Jin, "Super-resolution acoustic imaging using sparse recovery with spatial priming", *Proc. ICASSP*, pp.2414-2418, 2015.

[28] S. Rickard and O. Ylmaz, "On the approximate W-disjoint orthogonality of speech", *Proc. ICASSP*, pp.529–532, 2002.

[29] H. Saruwatari, *et al.*, "Speech Enhancement Using Nonlinear Microphone Array Based on Complementary Beamforming", *IEICE Trans. Fundam.*, Vol.E82-A, No.8, pp.1501–1510, 1999.

[30] H. Kamiyanagida, *et al.*, "Direction of Arrival Estimation Using Nonlinear Microphone Array", *IEICE Trans. Fundam.*, Vol.E84-A, No.4, pp.999–1010, 2001.

[31] S. Miyabe, *et al.*, "Kernel-based nonlinear independent component analysis for underdetermined blind source separation", *Proc. ICASSP*, pp.1641–1644, 2009.

[32] Y. Sugimoto, *et al.*, "Underdetermined DOA estimation by the non-linear MUSIC exploiting higher-order moments", *Proc. IWAENC 2012*, E-03, 2012.

[33] Y. Sugimoto, *et al.*, "An extension of MUSIC exploiting higher-order moments via nonlinear mapping", *IEICE Trans. Fundam.*, Vol.E99-A, No.6, pp.1152–1162, 2016.

[34] Y. Sugimoto, *et al.*, "Employing Moments of Multiple High Orders for High-Resolution Underdetermined DOA Estimation Based on MUSIC", *Proc. WASPAA'2013*, pp.1–4, 2013.

[35] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol.33, pp.387-392, 1985.

[36] P. Chen, M. Wicks, and R. Adve, "Development of a procedure for detecting the number of signal in a radar measurement", *IEE Proceedings on Radar Sonar and Navigation*, vol.148, pp.219-226, 2001.

[37] P. Danes and J. Bonnal "Information-Theoretic Detection of Broadband Sources in a Coherent Beamspace MUSIC Scheme", *Proc. of IROS*, pp.1976-1981, 2010.

[38] M. H. Hansen and B. Yu, "Minimum Description Length Model Selection Criteria for Generalized Linear Models", Statistics and Science: A Festschrift for Terry Speed, Vol.40, pp.145–163, 2003.

[39] K. Yamamoto, *et al.*, "Detection of Overlapping Speech in Meeting using Support Vector Machines and Support Vector Regression", *IEICE Trans. Fundamentals*, Vol.E89-A, No.8, pp.2158-2165, 2006.

[40] X. Zhai, *et al.*, "TDOA Estimation by Mapped SRP Based on Higher-Order Moment Analysis", *Proc. APSIPA2014*, pp.1–4, 2014.

[41] T. Pham and B. M. Sadler, "Adaptive wideband aeroacoustic array processing", *IEEE Trans. Acoust., Speech, Signal Process.*, pp.817-827, 1984.

[42] F. Asano, *et al.*, "Detection and separation of speech event using audio and video information fusion and its application to robust speech interface", *EURASIP J. Applied Signal Processing*, vol.2004, no.11, pp.1727–1738, 2004.

[43] M. Wax, T.-J. Shan and T. Kailath, "Spatio-temporal analysis by eigenstructure methods", *IEEE Trans. Acoust. Speech Signal Process.*, vol.32, pp.817–827, 1984.

[44] P. M. Cullagh, "Tensor methods in statistics", *Monographs on Statistics and Applied Probability*, 1987.

[45] C. L. Nikias and J. M. Mendel, "Signal Processing with higher-order spectra", *IEEE Signal Process. Magazine*, vol.10, pp.10–37, 1993.

[46] C. L. Nikias and A. P. Petropulu, "Higher-order Spetra Analysis", PTR Prentice Hall, 1993.

[47] M. Novey, *et al.*, "A complex generalized Gaussian distribution–characterization generation and estimation", *IEEE Trans. Signal Process.*, vol.58, pp.1427–1433, 2010.

[48] I. Tashev and A. Acero, "Statistical modeling of the speech signal", *Proc. IWAENC*, pp.70–79, 2010.

[49] O. Yilmaz, and S. Rickard, "Blind separation of speech mixtures via time-frequency masking", *IEEE Trans. Signal Process.*, vol.52, no.7, pp.1830–1846, 2004.

[50] K. Itou, *et al.*, "JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research", *Proc. IPSJ*, 5H–10, vol.2, pp.225–226, 2003.

[51] J. A. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics", *J. Acoust. Soc. Am.*, vol.65, no.4, pp.943-950, 1979.

[52] I. A. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence", *IEEE Trans. Speech Audio Process.*, vol.11, no.6, pp.709–716, 2003.

[53] B. Scholkopf, *et al.*, "Support vector methods in learning and Feature Extraction", *Aust. J. Intell. Infom. Process. Syst.*, vol.1, pp.3–9, 1998.

Research Activities

- Journal Publication

    – Yuya Sugimoto, Shigeki Miyabe, Takeshi Yamada, Shoji Makino, and Biing-Hwang (Fred) Juang, " An extension of MUSIC exploiting higher-order moments via nonlinear mapping, " IEICE Trans. Fundam., Vol.E99-A, No.6, pp.1152-1162, 2016.

- International Conference
  <With refereeing>

    – Yuya Sugimoto, Shigeki Miyabe, Takeshi Yamada, Shoji Makino, and Biing-Hwang (Fred) Juang, " Underdetermined DOA Estimation by the Non-linear MUSIC Exploiting Higher-Order Moments, " Proc. IWAENC ' 12, pp.1–4, Germany, Sep. 2012. <Award: Selected paper>

    – Yuya Sugimoto, Shigeki Miyabe, Takeshi Yamada, Shoji Makino, and Biing-Hwang (Fred) Juang, " Employing Moments of Multiple High Orders for High-Resolution Underdetermined DOA Estimation Based on MUSIC, " Proc. WASPAA ' 2013, pp.1–4, USA, Oct. 2013.

    – Michiaki Katoh, Yuya Sugimoto, Shigeki Miyabe, Shoji Makino, Takeshi Yamada, and Nobuhiko Kitawaki, " Visualization of Conversation Flow in Meetings by Analysis of Direction of Arrivals and Continuousness of Utterance," Proc. TJASSST11, Sun-No.343, pp.1–4, Tunisia, Nov. 2011.

    – Xiao-Dong Zhai, Yuya Sugimoto, Shigeki Miyabe, Takeshi Yamada, and Shoji Makino, "TDOA Estimation by Mapped Steered Response Power Based on Higher-Order Moment Analysis," Proc. APSIPA 2014, pp.1–4, Cambodia, Dec. 2014.

  <Without refereeing>

    – Yuya Sugimoto, Shigeki Miyabe, Takeshi Yamada, Shoji Makino, and Biing-Hwang (Fred) Juang, " Underdetermined DOA estimation by the extended MUSIC based on nonlinear higher-dimensional mapping, " ICT International Exchange Workshop, 0108-1, pp.1–4, Japan, Nov. 2014.

- Domestic Conference

– Yuya Sugimoto, Michiaki Katoh, Shoji Makino, and Takeshi Yamada, " Scattered Signal Detection by Frequency Analysis of Spatial Spectrum, " Annu. Conf. Acoustic Society of Japan, 3-P-7(c), pp.877–878, Mar. 2011.

– Yuya Sugimoto, Shigeki Miyabe, Takeshi Yamada, and Shoji Makino, " High-resolution DOA Estimation with Non-linear MUSIC Based on Higher-order Covariance, " Annu. Conf. Acoustic Society of Japan, 3-1-6, pp.763–766, Mar. 2012. <Award: Best student paper>

– Yuya Sugimoto, Shigeki Miyabe, Takeshi Yamada, and Shoji Makino, " Underdetermined DOA Estimation by the Non-linear MUSIC Based on Higher-Order Moment Analysis, " IEICE Technical report, no. EA2012-41, pp.49–54, Jun. 2012.

– Yuya Sugimoto, Shigeki Miyabe, Takeshi Yamada, Shoji Makino, and Biing-Hwang (Fred) Juang, " Non-linear MUSIC based on the analysis of multiple higher order moments, " Annu. Conf. Acoustic Society of Japan, 1-1-11, pp.601–604, Sep. 2013.

– Michiaki Katoh, Yuya Sugimoto, Shoji Makino, Takeshi Yamada, and Nobuhiko Kitawaki, " Scattered Speech Signal Detection by Principal Component Analysis for Spatial Power Spectrum, " IEICE Technical report, no. EA2010-47, pp.25–30, Aug. 2010.

– Michiaki Katoh, Yuya Sugimoto, Shoji Makino, Takeshi Yamada, and Nobuhiko Kitawaki, " Comparison between Principal Component Analysis and Frequency Analysis for Scattered Signal Detection for Spatial Spectrum, " Annu. Conf. Acoustic Society of Japan, 3-P-8(d), pp.879–880, Mar. 2011.

– Michiaki Katoh, Yuya Sugimoto, Shigeki Miyabe, Shoji Makino, Takeshi Yamada, and Nobuhiko Kitawaki, " Visualizing Conversation Flow of Meeting Recording with Speech Classification Based on Continuousness of Utterance, " Annu. Conf. Acoustic Society of Japan, 3-P-20, pp.197–200, Sep. 2011.

– Xiao-Dong Zhai, Yuya Sugimoto, Shigeki Miyabe, Takeshi Yamada, and Shoji Makino, " Time Difference of Arrival estimation by the non-linear GCC based on higher-order moment analysis, " Annu. Conf. Acoustic Society of Japan, 1-Q-23, pp.747–750, Sep. 2013.