PAPER

An Extension of MUSIC Exploiting Higher-Order Moments via Nonlinear Mapping

Yuya SUGIMOTO[†a)], Shigeki MIYABE[†], *Nonmembers*, Takeshi YAMADA[†], *Member*, Shoji MAKINO[†], *Fellow*, *and* Biing-Hwang JUANG[††], *Nonmember*

SUMMARY MUltiple SIgnal Classification (MUSIC) is a standard technique for direction of arrival (DOA) estimation with high resolution. However, MUSIC cannot estimate DOAs accurately in the case of underdetermined conditions, where the number of sources exceeds the number of microphones. To overcome this drawback, an extension of MUSIC using cumulants called $2q$-MUSIC has been proposed, but this method greatly suffers from the variance of the statistics, given as the temporal mean of the observation process, and requires long observation. In this paper, we propose a new approach for extending MUSIC that exploits higher-order moments of the signal for the underdetermined DOA estimation with smaller variance. We propose an estimation algorithm that nonlinearly maps the observed signal onto a space with expanded dimensionality and conducts MUSIC-based correlation analysis in the expanded space. Since the dimensionality of the noise subspace is increased by the mapping, the proposed method enables the estimation of DOAs in the case of underdetermined conditions. Furthermore, we describe the class of mapping that allows us to analyze the higher-order moments of the observed signal in the original space. We compare $2q$-MUSIC and the proposed method through an experiment assuming that the true number of sources is known as prior information to evaluate in terms of the bias-variance tradeoff of the statistics and computational complexity. The results clarify that the proposed method has advantages for both computational complexity and estimation accuracy in short-time analysis, i.e., the time duration of the analyzed data is short.

key words: *underdetermined DOA estimation, microphone array, MUSIC, 2q-MUSIC, higher-order statistics*

1. Introduction

Direction of arrival (DOA) estimation techniques have mainly been developed for applications in the radio communication field. They involve an array of antennas and are used as a processing step in many sensor systems including radar, sonar, and geodesic location systems [1]. Because of the importance of DOA estimation in many applications, a wide range of studies have already been reported. These techniques have also been adopted in microphone array signal processing targeting acoustic signals and have become an essential preprocessing technique for various applications such as noise reduction and source separation [2].

Early approaches to DOA estimation employed beamformer techniques utilizing the time differences between sig-nal arrivals [1], [2], such as generalized cross-correlation (GCC) [3] 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) method [4], 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. Over the past few decades, many DOA estimators with high resolution [5], [6] have been proposed. Among these methods, MUltiple SIgnal Classification (MUSIC) [7] 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 [8] was developed in the 1990s, and later this approach was generalized as $2q$-MUSIC [9], 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 [14]. However, the problems of estimation accuracy in short-time analysis and high computational complexity still remain because of the cumulants with a large variance and the complex calculation.

Meanwhile, for scenarios with more noise sources than microphones, a speech enhancement method called complementary beamforming [10] has been proposed, which has been applied to DOA estimation problems [11]. Complementary beamforming has also been extended and explained as the mapping of a signal onto a higher-dimensional space using a kernel function [12]. 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.

In this paper, we propose mapped MUSIC [13], an extension of MUSIC that utilizes higher-dimensional mapping for underdetermined DOA estimation with a higher resolution and a small variance. 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. In contrast to cumulants, moments suffer from estimation bias because they do not maintain additivity. However, their variance is considerably smaller than that of cumulants because of their simpler formulation. Therefore, mapped MUSIC with the mapping and conventional $2q$-MUSIC have a bias-variance tradeoff, and 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.

The rest of this paper is organized as follows. In Sect. 2, we state the problem. In Sect. 3, we review the conventional methods. In Sect. 4, we describe the proposed method. In Sect. 5, we compare the properties of the proposed method and the conventional methods. In Sect. 6, we evaluate the estimation performance of the proposed method. Finally, the paper is concluded in Sect. 7.

2. Problem Formulation

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

$$
\begin{aligned}
\mathbf{z}(\omega, t) &= [z_1(\omega, t), \cdots, z_M(\omega, t)]^T \\
&= \mathbf{A}(\omega)\mathbf{s}(\omega, t), \quad (1)
\end{aligned}
$$

$$
\mathbf{A}(\omega) = [\mathbf{a}_1(\omega), \cdots, \mathbf{a}_N(\omega)], \quad (2)
$$

$$
\mathbf{s}(\omega, t) = [s_1(\omega, t), \cdots, s_N(\omega, t)]^T, \quad (3)
$$

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

$$
\begin{aligned}
\mathbf{x}(\omega, t) &= [x_1(\omega, t), \cdots, x_M(\omega, t)]^T \\
&= \mathbf{z}(\omega, t) + \mathbf{n}(\omega, t), \quad (4)
\end{aligned}
$$

$$
\mathbf{n}(\omega, t) = [n_1(\omega, t), \cdots, n_M(\omega, t)]^T, \quad (5)
$$

where $n_j(\omega, t)$ is the zero-mean noise signal superimposed on the jth microphone.

In the signal model expressed by Eq. (4), each microphone observes a mixture of source signals and noise signals. The problem in this paper 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

$$
\mathbf{b}(\omega; \theta) = \frac{1}{\sqrt{M}} \left[\exp(-j\omega\tau_1), \cdots, \exp(-j\omega\tau_M) \right]^T,
$$
$$(6)$$

where τ_i $(i = 1, \cdots, M)$ are time delays from a reference point at each microphone, given by a signal from direction θ.

3. Conventional Methods: MUSIC and $2q$-MUSIC

This section reviews two conventional methods for high-resolution DOA estimation: MUSIC, which is based on subspace analysis exploiting the covariance, and $2q$-MUSIC, which is an extension of MUSIC exploiting $2q$th-order cumulants with an arbitrary positive integer q. Since the covariance is equivalent to the second-order cumulant, the description of $2q$-MUSIC in this section includes that of standard MUSIC as a special case when $q = 1$.

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

$$
\mathbf{C}_{2q}(\omega) = \left[c_{ij}(\omega) \right]_{ij}, \quad (7)
$$

$$
c_{ij}(\omega) \triangleq \mathrm{Cum}[x_{o_{i,1}}^{\oplus}(\omega, t), \cdots, x_{o_{i,q}}^{\oplus}(\omega, t),
$$
$$
x_{o_{j,q+1}}^{\ominus}(\omega, t), \cdots, x_{o_{j,2q}}^{\ominus}(\omega, t)], \quad (8)
$$

$$
x_{o_{i,l}}^{\oplus}(\omega, t) \triangleq \begin{cases} x_{o_{i,l}}(\omega, t) & (\text{if } l \text{ is odd}) \\ x_{o_{i,l}}^*(\omega, t) & (\text{if } l \text{ is even}) \end{cases}, \quad (9)
$$

$$
x_{o_{j,l}}^{\ominus}(\omega, t) \triangleq \begin{cases} x_{o_{j,l}}^*(\omega, t) & (\text{if } l \text{ is odd}) \\ x_{o_{j,l}}(\omega, t) & (\text{if } l \text{ is even}) \end{cases}, \quad (10)
$$

where $[\cdot]^*$ denotes the complex conjugate, $[\cdot]_{ij}$ denotes a matrix consisting of the argument in parentheses as its (i, j) entry, and $\mathrm{Cum}[\cdots]$ 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 $1, \cdots, M$:

$$
o_{i,l} = 1 + \left\lfloor \frac{i-1}{M^{q-l}} \right\rfloor \bmod M
$$

$$
\text{for } i = 1, \cdots, M^d \ \ l = 1, \cdots, q, \quad (11)
$$

$$o_{j,l} = 1 + \left\lfloor \frac{j-1}{M^{2q-l}} \right\rfloor \bmod M$$

$$\text{for } j = 1, \cdots, M^d \ l = q+1, \cdots, 2q, \qquad (12)$$

where $\lfloor \cdot \rfloor$ denotes the floor function and mod denotes the modulus. As described in [14], 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

$$\mathbf{C}_{2q}(\omega) = \begin{bmatrix} \mathbf{U}_s(\omega) & \mathbf{U}_n(\omega) \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta}_s(\omega) & \\ & \boldsymbol{\Delta}_n(\omega) \end{bmatrix} \begin{bmatrix} \mathbf{U}_s^H(\omega) \\ \mathbf{U}_n^H(\omega) \end{bmatrix},$$
$$(13)$$

where $[\cdot]^H$ denotes the complex conjugate transpose, 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 paper 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:

$$\mathbf{b}_q(\omega; \theta) \triangleq \begin{cases} \mathbf{b}(\omega; \theta) & (q = 1) \\ \mathbf{b}(\omega; \theta) \otimes \mathbf{b}_{q-1}^*(\omega; \theta) & (q \geq 2) \end{cases}, \quad (14)$$

$$f_{2q}(\omega; \theta) \triangleq \frac{1}{\left| \mathbf{b}_q^H(\omega; \theta) \mathbf{U}_n(\omega) \right|^2}, \qquad (15)$$

where $\otimes$ denotes the Kronecker product operator. Note that the DOA evaluation function $f_{2q}(\omega; \theta)$ is defined for each narrow band, and it is necessary to integrate the information from all the frequency bins to obtain a broadband DOA estimation. Among the various averaging operations that can be used to integrate the narrowband DOA estimation [15], we use the following geometric mean in this paper:

$$\overline{f_{2q}(\theta)} \triangleq \left[\prod_{\omega} f_{2q}(\omega; \theta) \right]^{\frac{1}{J}}, \qquad (16)$$

where J denotes the number of averaged frequency bins.

DOAs are estimated by finding the peaks of $\overline{f_{2q}(\theta)}$. 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 [16] for every entry $c_{ij}(\omega)$ of $\mathbf{C}_{2q}(\omega)$.

4. Proposed Method

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 $\boldsymbol{\phi} : \mathbb{C}^M \rightarrow \mathbb{C}^{M'}$ and conducts a similar analysis to MUSIC with the mapped observation vector $\boldsymbol{\phi}(\mathbf{x}(\omega, t))$. Mapped MUSIC is also a generalization of standard MUSIC, which corresponds to the special case that $\boldsymbol{\phi}(\mathbf{x}(\omega, t)) = \mathbf{x}(\omega, t)$. 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 $\boldsymbol{\phi}(\mathbf{x}(\omega, t))$:

1. The magnitude relation of the norm is retained.

$$||\boldsymbol{\phi}(\mathbf{x})||_2 \geq ||\boldsymbol{\phi}(\mathbf{y})||_2 \text{ if } ||\mathbf{x}||_2 \geq ||\mathbf{y}||_2. \qquad (17)$$

2. The origin remains intact.

$$\boldsymbol{\phi}(\mathbf{x}) = 0 \text{ if } \mathbf{x} = 0. \qquad (18)$$

3. The orthogonality between vectors is preserved.

$$\boldsymbol{\phi}^H(\mathbf{x})\boldsymbol{\phi}(\mathbf{y}) = 0 \text{ if } \mathbf{x}^H\mathbf{y} = 0. \qquad (19)$$

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 $\boldsymbol{\phi}$ satisfying Eqs. (17)–(19). The covariance matrix of $\boldsymbol{\phi}(\mathbf{x}(\omega, t))$ is expressed as

$$\mathbf{R}(\omega) = E[\boldsymbol{\phi}(\mathbf{x}(\omega, t))\boldsymbol{\phi}^H(\mathbf{x}(\omega, t))], \qquad (20)$$

where $E[\cdot]$ denotes the expectation of the argument. The following equations are obtained by the eigendecomposition of the covariance matrix $\mathbf{R}(\omega)$:

$$\mathbf{R}(\omega) = \mathbf{V}(\omega)\mathbf{E}(\omega)\mathbf{V}^H(\omega), \qquad (21)$$

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

$$\mathbf{V}^H(\omega)\mathbf{V}(\omega) = \mathbf{I}_{M'}, \qquad (22)$$

$$\mathbf{E}(\omega) = \mathrm{diag}[e_1(\omega), \cdots, e_{M'}(\omega)],$$
$$e_1(\omega) \geq \cdots \geq e_{M'}(\omega), \qquad (23)$$
$$M' = \dim[\boldsymbol{\phi}(\mathbf{x}(\omega,t))], \qquad (24)$$

where $\mathbf{v}_1(\omega), \cdots, \mathbf{v}_{M'}(\omega)$ are the eigenvectors associated with the respective eigenvalues $e_1(\omega), \cdots, e_{M'}(\omega)$, $\mathbf{I}_i$ denotes the i-dimensional identity matrix, $\mathrm{diag}[\cdot]$ is a diagonal matrix with the arguments in the diagonal entries, 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,

$$\mathcal{S}(\omega) \triangleq \mathrm{span}[\mathbf{v}_1(\omega), \cdots, \mathbf{v}_K(\omega)], \qquad (25)$$

where $\mathrm{span}[\cdot]$ denotes the subspace spanned by the argument vectors. In contrast to the cumulants used in $2q$-MUSIC, moments do not maintain additivity and the subspace dimensionality is not defined by the number of sources N. Also, our preliminary experiment shows that the best performance of the proposed method is obtained by setting K larger than N. However, the parameter optimization complicates the discussion, and we utilize the suboptimal setting $K \leftarrow N$ in the experiment reported in this paper for simplicity. 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 $\mathbf{v}_{K+1}(\omega), \cdots, \mathbf{v}_{M'}(\omega)$, which span the noise subspace orthogonal to the signal subspace:

$$\boldsymbol{\phi}^H(\mathbf{a}_i(\omega))\mathbf{v}_j(\omega) \approx 0$$
$$\text{for } i = 1, \cdots, N \ \ j = K+1, \cdots, M'. \qquad (26)$$

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 $\boldsymbol{\phi}(\mathbf{b}(\omega; \theta))$. 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. (26), for mapped MUSIC:

$$f_{\mathrm{map}}(\omega; \theta) \triangleq \frac{1}{\sum_{j=K+1}^{M'} \left| \boldsymbol{\phi}^H(\mathbf{b}(\omega; \theta))\mathbf{v}_j(\omega) \right|^2}, \qquad (27)$$

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:

$$\overline{f_{\mathrm{map}}(\theta)} \triangleq \left[\prod_\omega f_{\mathrm{map}}(\omega; \theta) \right]^{\frac{1}{J}}. \qquad (28)$$

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. (17)–(19), but its properties change with the choice of mapping. In this paper, to quantitatively evaluate the properties of the mapping, we focus on the mapping $\boldsymbol{\phi}_d : \mathbb{C}^M \to \mathbb{C}^{M^d}$, which gives a $2d$th-order cross-moment matrix as its covariance matrix. The mapping function $\boldsymbol{\phi}_d(\mathbf{x}(\omega,t))$ is defined recursively as

$$\boldsymbol{\phi}_d(\mathbf{x}(\omega,t)) \triangleq \begin{cases} \mathbf{x}(\omega,t) & (d = 1) \\ \mathbf{x}(\omega,t) \otimes \boldsymbol{\phi}_{d-1}^*(\mathbf{x}(\omega,t)) & (d \geq 2) \end{cases}. \qquad (29)$$

According to this definition, each entry of the map $\boldsymbol{\phi}_d(\mathbf{x}(\omega,t))$ is given as the product of d observed signals corresponding to its entry index,

$$\boldsymbol{\phi}_d(\mathbf{x}(\omega,t)) = \left[\prod_{l=1}^{d} x_{o_{1,l}}^\oplus(\omega,t), \cdots, \prod_{l=1}^{d} x_{o_{M^d,l}}^\oplus(\omega,t) \right]^T. \qquad (30)$$

The entries of $\boldsymbol{\phi}_d(\mathbf{x}(\omega,t))$ are similar to the first half of those in Eq. (8); thus, the tuple of d indices $\{o_{k,1}, \cdots, o_{k,d}\}$ is given in the same manner as Eq. (11):

$$o_{k,l} = 1 + \left\lfloor \frac{k-1}{M^{d-l}} \right\rfloor \mod M$$
$$\text{for } k = 1, \cdots, M^d \ \ l = 1, \cdots, d. \qquad (31)$$

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

$$\mathbf{R}_{2d}(\omega) = E\left[\boldsymbol{\phi}_d(\mathbf{x}(\omega,t))\boldsymbol{\phi}_d^H(\mathbf{x}(\omega,t)) \right]$$
$$= \left[r_{ij}(\omega) \right]_{ij}, \qquad (32)$$
$$r_{ij}(\omega) \triangleq E\left[\prod_{l=1}^{d} x_{o_{i,l}}^{l\oplus}(\omega,t) \prod_{l=1}^{d} x_{o_{j,l}}^{l\ominus}(\omega,t) \right]. \qquad (33)$$

The increase in dimensionality of the covariance matrix $\mathbf{R}_{2d}(\omega)$ from M to M^d with the mapping $\boldsymbol{\phi}_d : \mathbb{C}^M \to \mathbb{C}^{M^d}$ enhances the expressiveness of the noise subspace and enables us to estimate the DOAs in underdetermined cases. Note that the actual dimensionality after mapping becomes less than M^d. This is because the set of M^d mapped vectors $\boldsymbol{\phi}_d(\mathbf{x}_i)$, $i = 1, \cdots, M^d$ 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 $\boldsymbol{\phi}_d(\mathbf{x}_i)$ is an increasing function of d.

Here we discuss a similarity between the proposed method with the map $\boldsymbol{\phi}_d$ and $2q$-MUSIC based on cumulant analysis. From Eqs. (14) and (29), the mapped steering vector $\boldsymbol{\phi}_d(\mathbf{b}(\omega; \theta))$ 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. 5.1, 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 $\boldsymbol{\phi}$ and $\boldsymbol{\phi}'$ give exactly the same DOA estimation when their inner products are identical,

$$\boldsymbol{\phi}^H(\mathbf{x})\boldsymbol{\phi}(\mathbf{y}) = \boldsymbol{\phi}'^H(\mathbf{x})\boldsymbol{\phi}'(\mathbf{y}), \ \forall \mathbf{x}, \mathbf{y} \in \mathbb{C}^M, \tag{34}$$

and the analyses using these two mapping functions give identical results. Therefore, the mapping function $\boldsymbol{\phi}_d$ given by Eq. (29) 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 $\boldsymbol{\phi}_d$ of two arbitrary complex vectors $\mathbf{x}$ and $\mathbf{y}$ is given by

$$\boldsymbol{\phi}_d^H(\mathbf{x})\boldsymbol{\phi}_d(\mathbf{y}) = \begin{cases} \left|\mathbf{x}^H\mathbf{y}\right|^{d-1}\mathbf{x}^H\mathbf{y} & (d \text{ is odd}) \\ \left|\mathbf{x}^H\mathbf{y}\right|^d & (d \text{ is even}) \end{cases}. \tag{35}$$

Note that the inner product becomes real when the degree of the mapping d is even. In this instance, a real-valued mapping $\boldsymbol{\phi}'_d : \mathbb{C}^M \to \mathbb{R}^{M^d}$ satisfying Eq. (34) 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 $\boldsymbol{\phi}'_2 : \mathbb{C}^M \to \mathbb{R}^{M^2}$ can be employed:

$$\boldsymbol{\phi}'_2(\mathbf{x}) \triangleq \left[\boldsymbol{\phi}'^T_{\text{abs}}(\mathbf{x}), \boldsymbol{\phi}'^T_{\text{re}}(\mathbf{x}), \boldsymbol{\phi}'^T_{\text{im}}(\mathbf{x})\right]^T, \tag{36}$$

$$\boldsymbol{\phi}'_{\text{abs}}(\mathbf{x}) \triangleq [\forall |x_i|^2 | 1 \le i \le M]^T, \tag{37}$$

$$\boldsymbol{\phi}'_{\text{re}}(\mathbf{x}) \triangleq \sqrt{2}[\forall \text{Re}[x_i x_j^*] | 2 \le i \le M, 1 \le j \le i - 1]^T, \tag{38}$$

$$\boldsymbol{\phi}'_{\text{im}}(\mathbf{x}) \triangleq \sqrt{2}[\forall \text{Im}[x_i x_j^*] | 2 \le i \le M, 1 \le j \le i - 1]^T. \tag{39}$$

Moreover when $d > 2$, there are several redundant entries of equal value in $\boldsymbol{\phi}_d$, meaning that the rank of its covariance matrix becomes less than M^d. Hence, by designing an alternative mapping that satisfies Eq. (34) and omitting the redundant entries in the original mapping $\boldsymbol{\phi}_d$, we can remove the redundant computations. The mapping in the case of $d = 3$ can be simplified in this manner. We define the mapping $\boldsymbol{\phi}'_3 : \mathbb{C}^M \to \mathbb{C}^{\frac{M^3+M^2}{2}}$ as

$$\boldsymbol{\phi}'_3(\mathbf{x}) \triangleq \left[\boldsymbol{\phi}'^T_{3a}(\mathbf{x}), \boldsymbol{\phi}'^T_{3b}(\mathbf{x}), \boldsymbol{\phi}'^T_{3c}(\mathbf{x}), \boldsymbol{\phi}'^T_{3d}(\mathbf{x})\right]^T, \tag{40}$$

$$\boldsymbol{\phi}'_{3a}(\mathbf{x}) \triangleq [\forall |x_i|^2 x_i^* | 1 \le i \le M]^T, \tag{41}$$

$$\boldsymbol{\phi}'_{3b}(\mathbf{x}) \triangleq [\forall x_i x_j^{*2} | 1 \le i, j \le M, j \ne i]^T, \tag{42}$$

$$\boldsymbol{\phi}'_{3c}(\mathbf{x}) \triangleq \sqrt{2}[\forall |x_i|^2 x_j^* | 1 \le i, j \le M, j \ne i]^T, \tag{43}$$

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

$$(i, j, \text{ and } k \text{ are different}). \tag{44}$$

5. Comparison of Proposed Method 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.

5.1 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 [17]. 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 [14], 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 1 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 $s(\omega, t)$
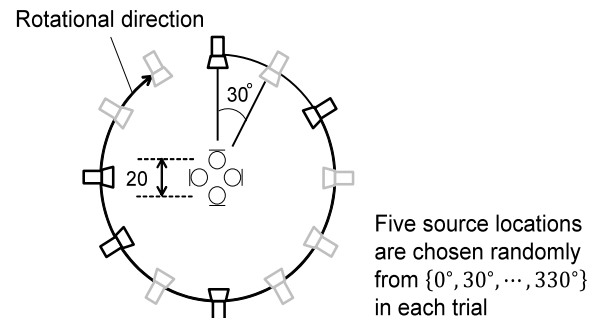

Fig. 1 Experimental environment for statistical comparison.

Table 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: $\beta = 0.05$–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]

whose directivities are given by steering vectors $\mathbf{b}(\omega; \theta)$ with θ randomly chosen from $\theta = \{0°, 30°, \cdots, 330°\}$. Source signals $s(\omega, t)$ are generated from a circularly symmetric complex generalized Gaussian distribution (CGGD) [18] whose probability density function is given as

$$p\left(s(\omega, t)\right) = \frac{\beta \exp\left(-\left(\frac{|s(\omega,t)|}{\alpha}\right)^\beta\right)}{2\pi\alpha^2 \Gamma\left(\frac{2}{\beta}\right)}, \tag{45}$$

where $\exp(\cdot)$ 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, 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 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 1 summarizes the experimental conditions.

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

$$\text{RMSE} \triangleq \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|\hat{\theta}_i - \theta_i\right|^2}, \tag{46}$$

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 between estimated DOAs and true DOAs, we chose the combination that minimizes the error. The results are shown in Fig. 2. First, both $2q$-MUSIC and the proposed mapped MUSIC improve estimation accuracy with
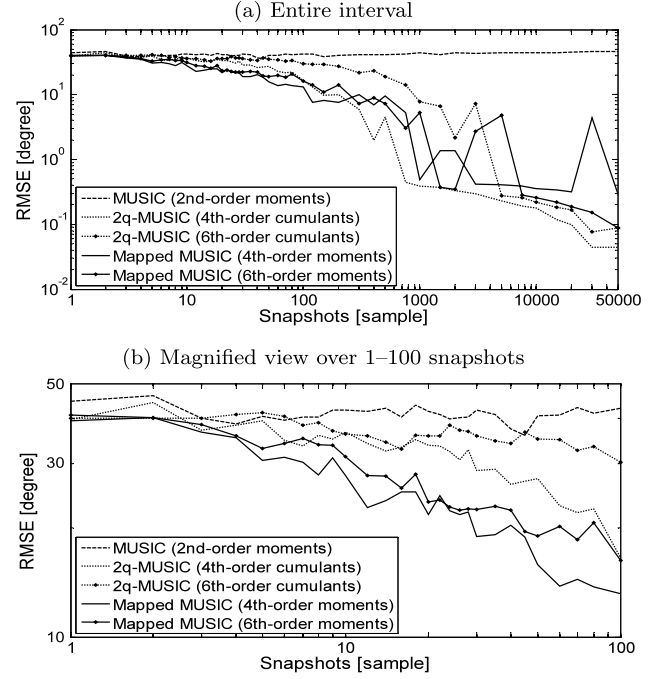

Fig. 2 Error transition with increasing number of snapshots ($\beta = 0.315$). Mapped MUSIC and $2q$-MUSIC show efficacy for underdetermined DOA estimation. The RMSE is evaluated over 100 trials. In the range of 1–100 snapshots, mapped MUSIC shows better performance and $2q$-MUSIC outperforms mapped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

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. 2, 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. 3. 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.

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

(a) 3 kHz
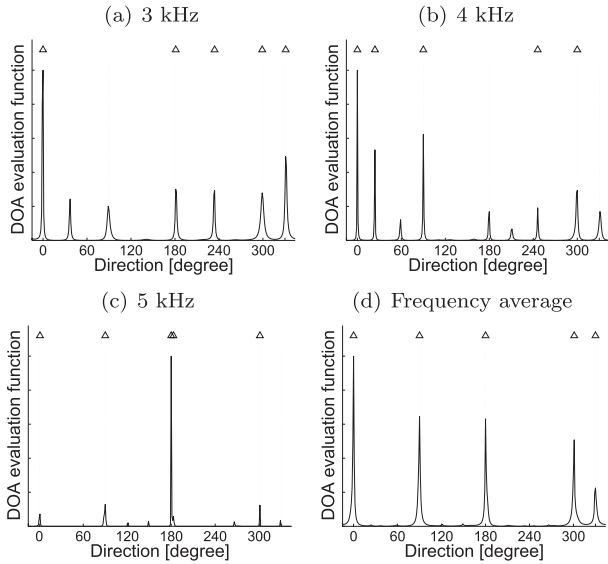
(b) 4 kHz

(c) 5 kHz

(d) Frequency average

Fig. 3 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.

(a) Snapshot 10

(b) Snapshot 100

(c) Snapshot 1000

Fig. 4 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.

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. 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. Thus, mapped MUSIC performs well with sparse signals such as speech.

5.2 Comparison of Computational Complexity

In this section, we discuss the computational complexity of the proposed method. As described in Sect. 3, 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 $\boldsymbol{\phi}_{2d}(\omega)$ given once. Table 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 mul-
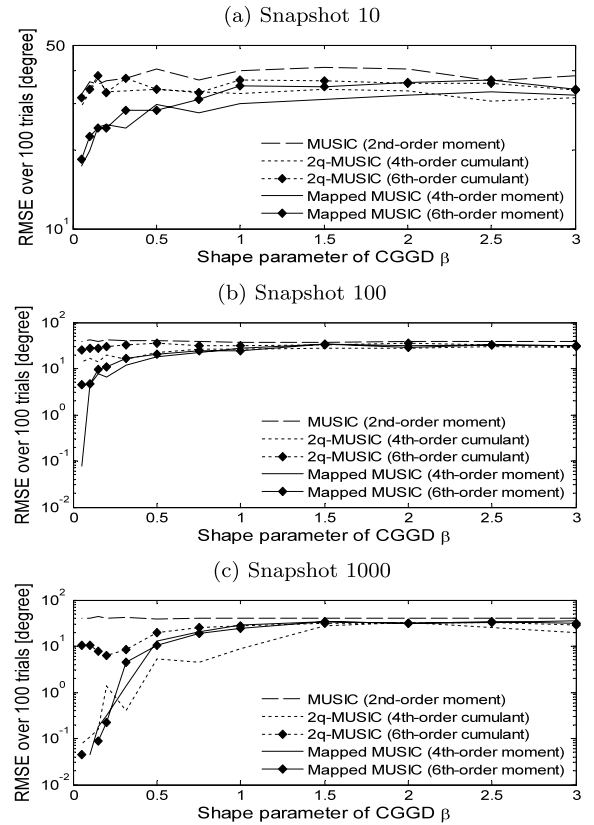
Table 2 Computational complexity.

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 $\boldsymbol{\phi}_{2d}$]	$4L(M^2 + M^4)$
$\mathbf{R}_6(\omega)$ [with $\boldsymbol{\phi}_{2d}$]	$4L(M^2 + M^3 + M^6)$
$\mathbf{R}_4(\omega)$ [with $\boldsymbol{\phi}'_{2d}$]	$L(M + 3M^2 + M^4)$
$\mathbf{R}_6(\omega)$ [with $\boldsymbol{\phi}'_{2d}$]	$L(3M^2 + 5M^3 + M^4 + 2M^5 + M^6)$
$\mathbf{R}_4(\omega)$ [entry-wise]	$12LM^4$
$\mathbf{R}_6(\omega)$ [entry-wise]	$20LM^6$

tiplications with the alternative map $\boldsymbol{\phi}'_{2d}$ ($d = 2, 3$), 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. 5 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 $\boldsymbol{\phi}_{2d}$ and $\boldsymbol{\phi}'_{2d}$, respectively, with increasing number of snapshots from 0 to 10000 when the number of microphones is four.
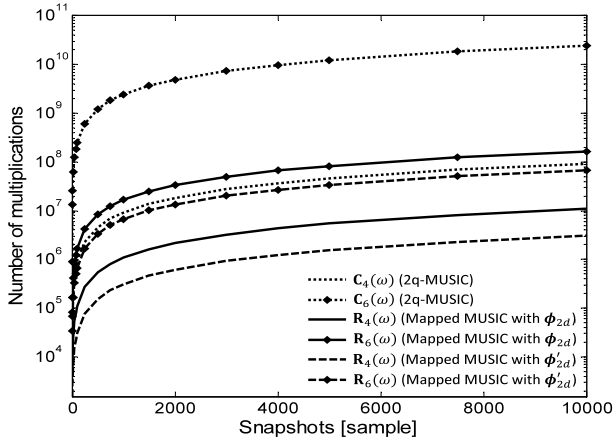
Fig. 5 Transitions of the numbers of multiplications required to construct the statistical matrix with increasing number of snapshots. The number of microphones is four and the number of snapshots ranges from 0 to 10000. The proposed method has lower computational complexity, and using the computationally efficient map ϕ'_{2d} markedly reduces the computational complexity.

6. Experiments to Evaluate Practical Estimation Performance

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.

6.1 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 [4], standard MUSIC, and $2q$-MUSIC. For this experiment, we created speech mixtures by convoluting Japanese speech samples [19] and room impulse responses assuming point sources in the environment shown in Fig. 6 using the image method [20]. We prepared seven different reverberant conditions and diffuse pink noises [21] 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. 5.1, we use 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. 5.1 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 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
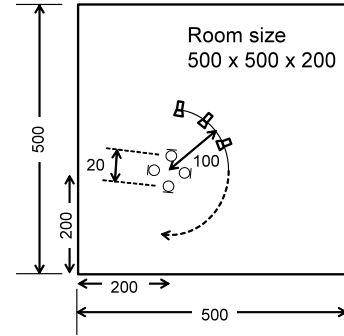

Fig. 6 Environment for simulation experiment.

Table 3 Conditions for simulation experiment.

Microphone array	Circular array with radius of 0.1 m
# of microphones	4
Sound sources	Point sources 1 m apart from array
# of sources	3, 5
Room size	$5 \times 5 \times 2$ [m]
Noise type	Diffuse pink noise
SNR	10,20 [dB]
Reverberation (T_{60})	0,150,500 [ms]
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

Table 4 Conditions for experiment using impulse response measured in real room.

Sound sources	Speakers 1 m apart from array
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]

in two different studios. To mimic noisy observation, we observed diffuse noise with the same microphone array in a noisy room containing several computers and fans. The other conditions followed those in the simulation experiment. Table 4 shows the experimental conditions that were different from those in the simulation experiment.

Figures 7 and 8 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 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
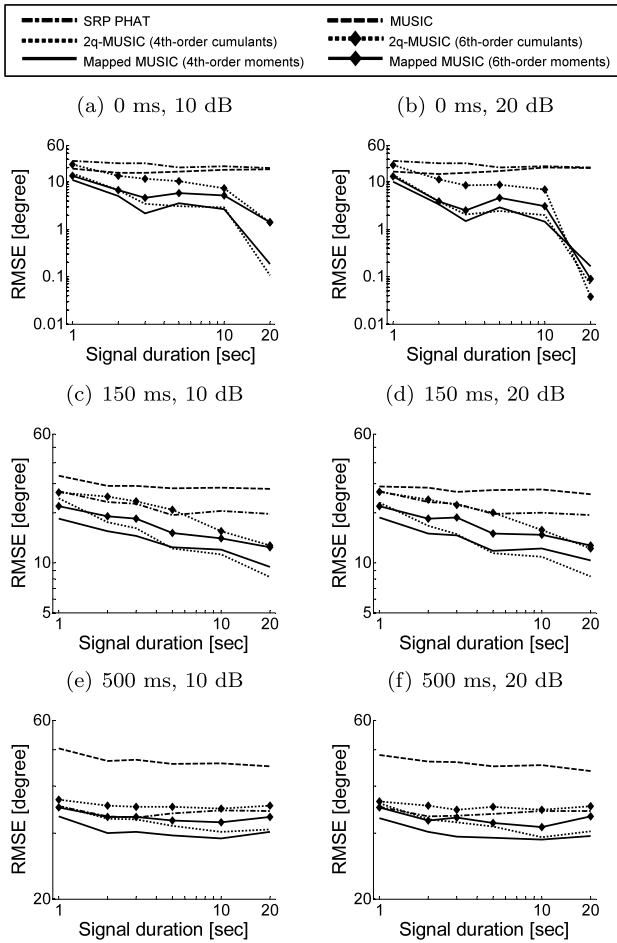
Fig. 7 Results of simulation experiment. Each subcaption represents a combination of reverberation and SNR.
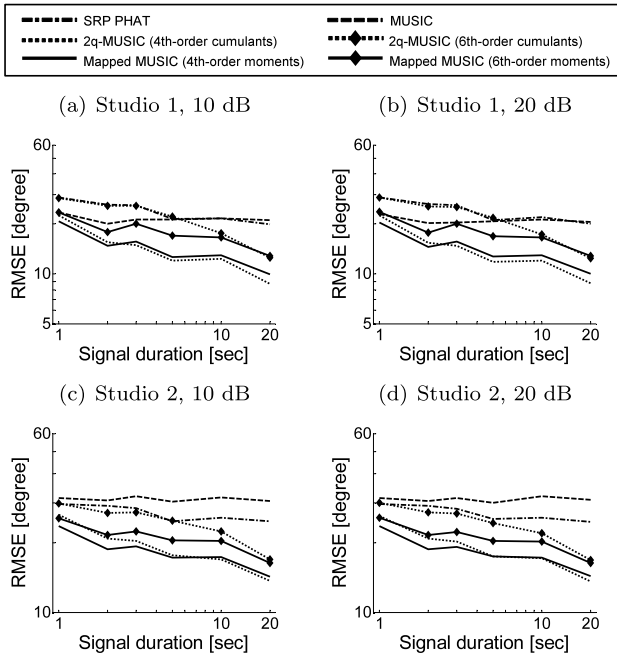

Fig. 8 Results of experiment with the measured impulse response. Each subcaption represents a combination of recording room and SNR.

Table 5 Computational environment.

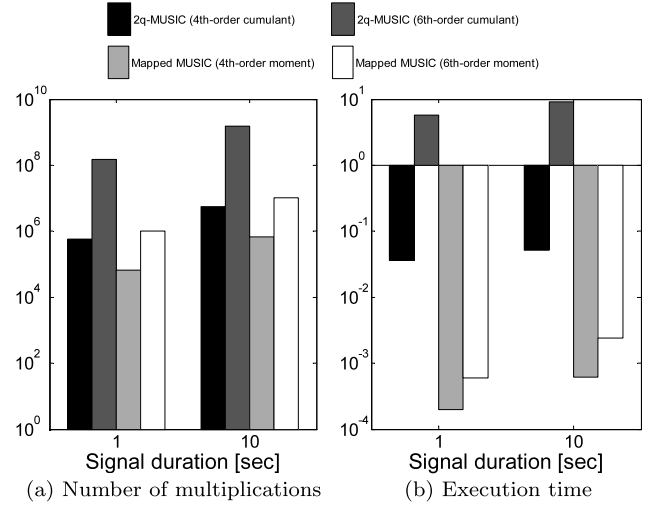| CPU	Intel Corei7-3930K 3.2 GHz
Memory	DDR3 64GB
OS	CentOS 6.6
Software	MATLAB R2010b

Fig. 9 Number of multiplications and execution time.

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.

6.2 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 5, we conducted an evaluation by averaging 100 measurements for two signal lengths, 1 s and 10 s.

Figure 9 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.

7. Conclusion

In this paper, 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 for many applications requiring short-time processing and rapid computation.

Acknowledgments

References

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

[2] M. Brandstein and D. Ward, eds., Microphone Arrays: Digital Signal Processing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[3] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. Acoust., Speech, Signal Process., vol.24, no.4, pp.320–327, 1976.

[4] H.F. Silverman, Y. Yu, J.M. Sachar, and W.R. Patterson, "Performance of real-time source-location estimators for a large-aperture microphone array," IEEE Trans. Speech Audio Process., vol.13, no.4, pp.593–606, 2005.

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

[6] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," Proc. IEEE, vol.57, no.8, pp.1408–1418, 1969.

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

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

[9] 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.

[10] H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Speech enhancement using nonlinear microphone array based on complementary beamforming," IEICE Trans. Fundamentals, vol.E82-A, no.8, pp.1501–1510, Aug. 1999.

[11] H. Kamiyanagida, H. Saruwatari, K. Takeda, F. Itakura, and K. Shikano, "Direction of arrival estimation using nonlinear microphone array, IEICE Trans. Fundamentals, vol.E84-A, no.4, pp.999–1010, April 2001.

[12] S. Miyabe, B.-H. Juang, H. Saruwatari, and K. Shikano, "Kernel-based nonlinear independent component analysis for underdetermined blind source separation," Proc. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, pp.1641–1644, 2009.

[13] Y. Sugimoto, S. Miyabe, T. Yamada, S. Makino, and F. Juang, "Underdetermined DOA estimation by the non-linear MUSIC exploiting higher-order moments," Proc. IWAENC 2012, International Workshop on Acoustic Signal Enhancement pp.1–4, 2012.

[14] P. Chevalier, L. Albera, A. Ferreol, and P. Comon, "On the virtual array concept for higher order array processing," IEEE Trans. Signal Process., vol.53, no.4, pp.1254–1271, 2005.

[15] M. Wax, T.-J. Shan, and T. Kailath, "Spatio-temporal spectral analysis by eigenstructure methods," IEEE Trans. Acoust., Speech, Signal Process., vol.32, no.4, pp.817–827, 1984.

[16] P.M. Cullagh, Tensor Methods in Statistics, Monographs on Statistics and Applied Probability, Chapman & Hall, 1987.

[17] C.L. Nikias and J.M. Mendel, "Signal processing with higher-order spectra," IEEE Signal Process. Mag., vol.10, no.3, pp.10–37, 1993.

[18] M. Novey, T. Adali, and A. Roy, "A complex generalized Gaussian distribution — Characterization, generation, and estimation," IEEE Trans. Signal Process., vol.58, no.3, pp.1427–1433, 2010.

[19] 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.

[20] J.B. 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.

[21] 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.

Yuya Sugimoto was born in Kochi, Japan. He received his B.Eng. and M.Eng. degrees from Tsukuba University in 2009 and 2012, respectively. He is now pursuing Ph.D. degree at the Graduate School of the University of Tsukuba. His research interests include directional estimation and speech diarization in array signal processing. He is a member of ASJ.

Shigeki Miyabe received his B.E. degree from Kobe University in 2003 and received his M.E. and Ph.D. degrees from Nara Institute of Science and Technology (NAIST) in 2005 and 2007, respectively. From 2008 to 2009, he was a Visiting Scholar with Georgia Institute of Technology. In 2009, he joined the Graduate School of Information Science and Technology, University of Tokyo, as a researcher and became Assistant Professor in 2010. He has been an Assistant Professor at the Life Science Center of Tsukuba Advanced Research Alliance, University of Tsukuba, since 2011. He received the Awaya Prize from the Acoustic Society of Japan in 2012. He is a member of IEEE and ASJ.

Takeshi Yamada was born in Osaka, Japan. He received his B. Eng. degree from Osaka City University, Japan, in 1994 and his M. Eng. and Dr. Eng. degrees from Nara Institute of Science and Technology, Japan, in 1996 and 1999, respectively. He is presently an Associate Professor with the Faculty of Engineering, Information and Systems, University of Tsukuba, Japan. His research interests include speech recognition, sound scene understanding, multichannel signal processing, media quality assessment, and e-learning. He is a member of IEEE, IEICE, IPSJ, ASJ, and JLTA.

Shoji Makino received his B.E., M.E., and Ph.D. degrees from Tohoku University, Japan, in 1979, 1981, and 1993, respectively. He joined NTT in 1981. He is now a Professor at University of Tsukuba. His research interests include adaptive filtering technologies and realization of acoustic echo cancellation, blind source separation of convolutive mixtures of speech. He received the ICA Unsupervised Learning Pioneer Award in 2006, the IEEE MLSP Competition Award in 2007, the TELECOM System Technology Award of the TAF in 2004, the Achievement Award of the IEICE in 1997, the Outstanding Technological DevelopmentAward of theASJ in 1995, the Paper Award of the IEICE in 2005 and 2002, the Paper Award of the ASJ in 2005 and 2002. He is the author or co-author of more than 200 articles in journals and conference proceedings and is responsible formore than 150 patents. He was a Keynote Speaker at ICA2007 and a Tutorial speaker at ICASSP2007. He is a member of the Award Committee of the IEEE James L. Flanagan Speech & Audio Processing Award. He is a member of the Awards Board and the Conference Board of the IEEE SP Society. He is an Associate Editor of the IEEE Transactions on Speech and Audio Processing and an Associate Editor of the EURASIP Journal on Applied Signal Processing. He is the Chair of the Technical Committee on Blind Signal Processing of the IEEE CAS Society and a member of the Technical Committee on Audio and Electroacoustics of the IEEE SP Society. He was the Chair of the Technical Committee on Engineering Acoustics of the IEICE and the ASJ. He is a member of the International ICA Steering Committee and a member of the International IWAENC Standing committee. He is the General Chair of the WASPAA2007, the General Chair of the IWAENC2003, the Organizing Chair of the ICA2003, and is the designated Plenary Chair of ICASSP2012. He is an IEEE Fellow, an IEICE Fellow, a council member of the ASJ, and a member of the EURASIP.

Biing-Hwang Juang received his Ph.D. from the University of California, Santa Barbara in 1981. He had worked at Speech Communications Research Laboratory (SCRL) and Signal Technology, Inc. (STI) on a number of Government-sponsored research projects. Notable accomplishments during the period include development of vector quantization for voice applications, voice coders at extremely low bit rates, 800 bps and around 300 bps, and robust vocoders for use in satellite communications. He subsequently joined the Acoustics Research Department of Bell Laboratories, working in the area of speech enhancement, coding and recognition. He became Director of Acoustics and Speech Research at Bell Labs in 1996, and Director of Multimedia Technologies Research at Avaya Labs (a spin-off of Bell Labs) in 2001. His group continued the long heritage of Bell Labs in speech communication research, including, most notably, the invention of electret microphone, network echo canceller, a series of speech CODECs, and key algorithms for signal modeling and automatic speech recognition. In the past few years, he and his group developed a speech server for applications such as AT&T's advanced 800 calls and the Moviefone, the Perceptual Audio Coder (PAC) for digital audio broadcasting in North America (in both terrestrial and satellite systems), and a world-first real-time full-duplex hands-free stereo teleconferencing system. He has published extensively, including the book "Fundamentals of Speech Recognition", co-authored with L.R. Rabiner, and holds about twenty patents. He joins Georgia Tech in 2002.