

## 複数次数の高次モーメントを用いた非線形 MUSIC \*

☆杉本侑哉, 宮部滋樹, 山田武志, 牧野昭二 (筑波大),  
Fred Juang (ジョージア工科大)

### 1 はじめに

マイクロホンアレーを用いた信号処理において音源の方向推定 (DOA) は, 話者追跡や音声強調などの様々な応用技術の前処理として必要不可欠である. 高分解能に方向推定を行う代表的な手法として, 観測信号の部分空間分析に基づく MUSIC (Multiple Signal Classification) 法 [1] が知られている. しかし MUSIC は, 観測信号の伝達関数ベクトルと直交する雑音部分空間を推定する必要があり, 共分散行列の次元数に拘束される.  $N$  個の音源方向を推定するためには  $M (> N)$  個のマイクロホンが必要であり,  $N$  が  $M$  に近づくほど推定性能は低下する. したがって, 多数の音源を推定する場合, アレーの大規模が避けられないといった問題がある.

このような背景から, 音源数がマイクロホン数よりも多い劣決定条件においても高分解能に方向推定を行うために, いくつかの MUSIC 拡張手法が提案されている. 例えば, 90 年代初めの 4 次キウムラント分析に基づく手法 [2] や, 近年提案された上手法を任意の偶数次キウムラント分析へと一般化した  $2q$ -MUSIC [3] などが挙げられる. これらの手法は, 高次クロスキウムラントを用いてバーチャルに観測信号の次元数を増大させることで, MUSIC よりも高い解像度と劣決定条件での方向推定を達成している. また, 異なったアプローチでの MUSIC 拡張手法として, 我々は写像 MUSIC [4] を提案している. 写像 MUSIC は劣決定条件で音声強調を行う非線形ビームフォーマ [5, 6, 7] のアイデアを MUSIC に応用した手法であり, 任意の偶数次モーメント分析に相当する. 推定性能は従来のキウムラント分析に基づく手法とほぼ同等であるが, 観測の次元数を増大させるために高次元写像を用いることで大きく計算量が削減されている.

しかし従来の MUSIC 拡張手法には, [3] に示されているように, 推定の分解能を高めるためには統計量の次数を上げる必要があるが, 同時に統計的バイアスにより安定性が損われるといったトレードオフがあった. 本稿ではこれを解決し, さらに分解能を高めた方向推定を目的とした, 複数次数の高次モーメント分析に基づく方向推定手法を提案する. 提案手法は, 単一の偶数次モーメント分析のための写像を, 直和により簡便に拡張した写像を用いて複数次数モーメントの分析を行う. 単一の高い次数のモーメント分

析において劣化要因となる統計的バイアスの影響を, 低次モーメントとの同時分析によって緩和することで, 従来の MUSIC 拡張手法に比べて極めて高い分解能を達成する.

### 2 問題設定

本稿を通して, 信号は複素振幅として表現する. 観測信号を次のようにモデル化する.

$$\begin{aligned} \mathbf{x}(\omega, t) &= [x_1(\omega, t), \dots, x_M(\omega, t)]^T \\ &= \sum_{i=1}^N \mathbf{a}_i(\omega) s_i(\omega, t) + \mathbf{n}(\omega, t) \end{aligned} \quad (1)$$

$$\mathbf{n}(\omega, t) = [n_1(\omega, t), \dots, n_M(\omega, t)]^T \quad (2)$$

$$\mathbf{a}_i(\omega) = [a_{1,i}(\omega), \dots, a_{M,i}(\omega)]^T \quad (3)$$

ただし,  $t = 1, \dots, L$  はある処理区間における時間フレームインデックス,  $\omega$  は周波数,  $M$  は素子数,  $N$  は音源数,  $s_i(\omega, t)$  は  $i$  番目の音源信号の複素振幅,  $x_j(\omega, t)$  は  $j$  番目の素子における観測信号の複素振幅,  $n_j(\omega, t)$  は  $j$  番目の素子における観測雑音の複素振幅,  $a_{j,i}(\omega)$  は  $i$  番目の音源から  $j$  番目のマイクロホンへの伝達関数を表す. (1) に示す信号モデルにおいて, 各マイクでは複数の音源から到来する信号と, 雑音が線形に混ざり合っ観測される. 本稿における問題設定を, このような混合信号の分析により各音源  $s_1(\omega, t), \dots, s_N(\omega, t)$  の到来方向を推定することとし, 次式に示す任意の方向  $\theta$  のステアリングベクトル  $\mathbf{b}(\omega; \theta)$  ( $\|\mathbf{b}(\omega; \theta)\|_2 = 1$ ) から, 伝達関数ベクトル  $\mathbf{a}_1(\omega), \dots, \mathbf{a}_N(\omega)$  との誤差がより小さいステアリングベクトル  $\mathbf{b}(\omega; \theta_1), \dots, \mathbf{b}(\omega; \theta_N)$  を探索することにより行う.

$$\mathbf{b}(\omega; \theta) = \frac{1}{\sqrt{M}} [\exp(-j\omega\tau_1), \dots, \exp(-j\omega\tau_M)] \quad (4)$$

ただし,  $\tau_i$  ( $i = 1, \dots, M$ ) は方向  $\theta$  から到来する信号が, アレーの基準点と各素子の間に作る時間遅延量を表す.

### 3 写像 MUSIC

#### 3.1 写像 MUSIC による方向推定のアルゴリズム

写像 MUSIC [4] は観測信号  $\mathbf{x}(\omega, t)$  を非線形関数  $\phi: \mathbb{C}^M \rightarrow \mathbb{C}^{M'}$  ( $M' \geq M$ ) により写像し,  $M'$  次元

\* Non-linear MUSIC based on the analysis of multiple higher order moments. by Yuya, SUGIMOTO, Shigeki, MIYABE, Takeshi, YAMADA, Shoji, MAKINO (University of Tsukuba), Fred, Jduang (Georgia Institute of Technology)

ユークリッド空間においてベクトル  $\phi(\mathbf{x}(\omega, t))$  の共分散行列を用いて MUSIC と同様の分析を行う方向推定手法である。本章では、写像 MUSIC による推定のアルゴリズムについて述べる。まず、写像 MUSIC が正しく方向推定を行うためには、写像の前後で各素子間の相関情報が保持される必要がある。この条件は式 (5)–(7) により表される。

1. ノルムの大小関係を維持する。

$$\|\mathbf{x}\|_2 \geq \|\mathbf{y}\|_2 \rightarrow \|\phi(\mathbf{x})\|_2 \geq \|\phi(\mathbf{y})\|_2 \quad (5)$$

2. 原点を移動させない。

$$\mathbf{x} = \mathbf{0} \rightarrow \phi(\mathbf{x}) = \mathbf{0} \quad (6)$$

3. 信号間の直交性を維持する。

$$\mathbf{x}^H \mathbf{y} = 0 \rightarrow \phi^H(\mathbf{x}) \phi(\mathbf{y}) = 0 \quad (7)$$

ただし、 $[\cdot]^H$  は複素共役転置を表す。条件 (5)–(7) を満たす  $\phi$  により観測信号  $\mathbf{x}(\omega, t)$  を写像し、ベクトル  $\phi(\mathbf{x}(\omega, t))$  から共分散行列  $\mathbf{R}(\omega)$  を得る。

$$\mathbf{R}(\omega) = E[\phi(\mathbf{x}(\omega, t))\phi^H(\mathbf{x}(\omega, t))] \quad (8)$$

ただし、 $E[\cdot]$  は期待値演算、 $[\cdot]^H$  は複素共役転置を表す。共分散行列  $\mathbf{R}(\omega)$  の固有値分解より、次式を得る。

$$\mathbf{R}(\omega) = \mathbf{V}(\omega)\mathbf{E}(\omega)\mathbf{V}^H(\omega) \quad (9)$$

$$\begin{aligned} \mathbf{V}(\omega) &= [\mathbf{v}_1(\omega), \dots, \mathbf{v}_{M'}(\omega)] \\ \mathbf{V}^H(\omega)\mathbf{V}(\omega) &= \mathbf{I}_{M'} \end{aligned} \quad (10)$$

$$\begin{aligned} \mathbf{E}(\omega) &= \text{diag}[e_1(\omega), \dots, e_{M'}(\omega)] \\ e_1(\omega) &\geq \dots \geq e_{M'}(\omega) \end{aligned} \quad (11)$$

$$M' = \dim[\phi(\mathbf{x}(\omega, t))] \quad (12)$$

ただし、 $e_1(\omega), \dots, e_{M'}(\omega)$  は固有値、 $\mathbf{v}_1(\omega), \dots, \mathbf{v}_{M'}(\omega)$  は各固有値に対応する固有ベクトル、 $\mathbf{I}_{M'}$  は  $M'$  次元の単位行列、 $\text{diag}[\cdot]$  は引数を対角成分とする対角行列、 $\dim[\cdot]$  は次元数を表す。このとき、伝達関数  $\mathbf{a}$  に係る方向性の音源信号の写像と対応する固有ベクトルの本数を  $N'$  で表すと、

$$\begin{aligned} \text{span}[\phi(\mathbf{a}_i(\omega)) | i = 1, \dots, N] \\ \simeq \text{span}[\mathbf{v}_j(\omega) | j = 1, \dots, N'] \end{aligned} \quad (13)$$

(ただし、 $\text{span}[\cdot]$  は引数ベクトルの張る空間を表す) 写像空間における雑音部分空間  $\mathcal{N}(\omega)$  は次式で定義される。

$$\mathcal{N}(\omega) \triangleq \text{span}[\mathbf{v}_k(\omega) | k = N' + 1, \dots, M'] \quad (14)$$

このとき  $N'$  は、写像関数  $\phi$  と音源信号数  $N$  に依存して、 $N$  から増加した数となるが、伝達関数ベクトルの写像  $\phi(\mathbf{a}_i(\omega))$  と雑音部分空間  $\mathcal{N}(\omega)$  の間の直交

性は条件 (5)–(7) によって保証されるため、次の条件を満たす。

$$\begin{aligned} |\phi^H(\mathbf{a}_i(\omega))\mathbf{v}_k(\omega)|^2 &\ll 1 \\ (i = 1, \dots, N \quad k = N' + 1, \dots, M') \end{aligned} \quad (15)$$

ステアリングベクトルが音源の伝達関数ベクトルを模擬する仮定  $(\mathbf{b}(\omega; \theta_i) \simeq \mathbf{a}_i(\omega) \quad (i = 1, \dots, N))$  と (15) より、雑音部分空間へのステアリングベクトルの写像  $\phi(\mathbf{b}(\omega; \theta))$  の直交射影を評価することで音源方向探索が可能である。以上より、ステアリングベクトルの方向  $\theta$  が音源方向  $\theta_i$  と一致した時にピークを形成する、MUSIC スコア  $f(\omega; \theta)$  が次式により定義できる。

$$f(\omega; \theta) \triangleq \frac{1}{\sum_{k=N'+1}^{M'} |\phi^H(\mathbf{b}(\omega; \theta))\mathbf{v}_k(\omega)|^2} \quad (16)$$

なお、(16) に定義した  $f(\omega; \theta)$  は狭帯域での MUSIC スコアであり、広帯域の推定結果を得るためにはこれらの結果を周波数方向に平均する必要がある。周波数平均の方法としてはいくつか考えられるが [8]、本稿では以下の幾何平均を採用する。

$$\overline{f(\theta)} \triangleq \left[ \prod_{\omega} f(\omega; \theta) \right]^{\frac{1}{K}} \quad (17)$$

ただし、 $K$  は平均演算に用いる周波数ビンの数を表す。推定音源方向は  $\overline{f(\theta)}$  のピーク探索により与えられる。

### 3.2 偶数次モーメント分析のための写像 $\phi_d$

[4] において、我々は任意の  $2d$  次クロスモーメント分析のための写像  $\phi_d: \mathbb{C}^M \rightarrow \mathbb{C}^{M^d}$  を提案している。

$$\phi_d(\mathbf{x}(\omega, t)) \triangleq \left[ \prod_{l=1}^d x_{c_{1l}}^{l \otimes}(\omega, t), \dots, \prod_{l=1}^d x_{c_{M^d l}}^{l \otimes}(\omega, t) \right]^T \quad (18)$$

$$x^{l \otimes} \triangleq \begin{cases} x & (\text{if } l \text{ is odd}) \\ x^* & (\text{if } l \text{ is even}) \end{cases} \quad (19)$$

ただし、 $[\cdot]^*$  は複素共役を表す。ここで、 $c_{kl}$  は  $\phi_d$  の計算に用いる  $\mathbf{x}$  の要素番号を指定するインデックスであり、 $\{1, \dots, M\}$  から  $d$  要素選択する全  $M^d$  通りの重複順列を、任意の順番で行ベクトルとして並べた  $M^d \times d$  インデックス行列  $\mathbf{C}$  の  $(k, l)$  成分として与えられる。

$$\mathbf{C} \triangleq [c_{kl}]_{kl} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 2 \\ & & \vdots & \\ M & M & \dots & M \end{bmatrix} \quad (20)$$

ただし、 $[\cdot]_{ij}$  は  $(i, j)$  要素が引数により与えられる行列を表す。写像  $\phi_d$  を用いた写像 MUSIC は、キュムラント分析に基づく従来の MUSIC 拡張手法と比べて、少ない計算量でほぼ同程度の推定性能を得ることができる。

## 4 提案手法

より高い分解能での方向推定のために、複数次数のクロスモーメントを同時分析するように拡張した写像 MUSIC を提案する。提案手法は、3章で述べた写像 MUSIC に、以下の写像  $\phi_{d_1, \dots, d_m}$  を適用することで実現される。

$$\phi_{d_1, \dots, d_m} \triangleq \phi_{d_1} \oplus \dots \oplus \phi_{d_m} \quad (21)$$

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

$$d_{i+1} > d_i \quad (i = 1, \dots, m-1) \quad (22)$$

ただし、 $\oplus$  は直和を表す。 $\phi_{d_1, \dots, d_m}$  は、それぞれ次数の異なる  $2d_i$  次モーメント分析のための写像  $\phi_{d_i}$  ( $i = 1, \dots, m$ ) の直和として与えられ、 $\phi_{d_1, \dots, d_m}$  の共分散行列  $\mathbf{R}_{d_1, \dots, d_m}^\oplus(\omega)$  は次式のように複数次数のモーメントを含むことから、 $\phi_{d_1, \dots, d_m}$  を用いた写像 MUSIC はこれらの複数次数モーメント同時分析に相当する。

$$\begin{aligned} \mathbf{R}_{d_1, \dots, d_m}^\oplus(\omega) &= E[\phi_{d_1, \dots, d_m}(\mathbf{x}'(\omega, t))\phi_{d_1, \dots, d_m}^H(\mathbf{x}'(\omega, t))] \\ &= \begin{bmatrix} \mathbf{R}_{d_1, d_1}(\omega) & \mathbf{R}_{d_1, d_2}(\omega) & \dots & \mathbf{R}_{d_1, d_m}(\omega) \\ \mathbf{R}_{d_2, d_1}(\omega) & \mathbf{R}_{d_2, d_2}(\omega) & \dots & \mathbf{R}_{d_2, d_m}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{d_m, d_1}(\omega) & \mathbf{R}_{d_m, d_2}(\omega) & \dots & \mathbf{R}_{d_m, d_m}(\omega) \end{bmatrix} \end{aligned} \quad (23)$$

$$\begin{aligned} \mathbf{R}_{d_i, d_j}(\omega) &\triangleq E[\phi_{d_i}(\mathbf{x}'(\omega, t))\phi_{d_j}^H(\mathbf{x}'(\omega, t))] = [r_{pq}]_{pq} \\ (i, j &= 1, \dots, m \quad p = 1, \dots, M^{d_i} \\ q &= 1, \dots, M^{d_j}) \end{aligned} \quad (24)$$

$$r_{pq} = E \left[ \left( \prod_{l=1}^{d_i} x'_{c_{pl}} \otimes \right) \left( \prod_{l=1}^{d_j} x'_{c_{ql}} \otimes \right)^* \right] \quad (25)$$

ただし、 $\mathbf{x}'(\omega, t) = [x'_1(\omega, t), \dots, x'_M(\omega, t)]^T$  は正規化された観測信号（正規化の規則は後述）、小行列  $\mathbf{R}_{d_i, d_j}(\omega)$  は  $(d_i + d_j)$  次クロスモーメント行列を表す。このとき行列  $\mathbf{R}_{d_1, \dots, d_m}^\oplus(\omega)$  全体と、単一で最大次数の小行列  $\mathbf{R}_{d_m, d_m}(\omega)$  との次元数を比較すると、それぞれ  $\sum_{i=1}^m M^{d_i}$  と  $M^{d_m}$  となる。次元数の差がどの程度の大きさかは  $\phi_{d_1, \dots, d_m}$  の設計に依存するが、これは単一次数モーメントを分析する写像 MUSIC と比べて、提案手法がより大きな観測の次元拡張を行っていることを示している。さらに  $\mathbf{R}_{d_1, \dots, d_m}^\oplus(\omega)$  の分析では、低次のモーメントを高次のモーメントと同時に分析することで、特に高次モーメントで問題となる統計的バイアスの影響が抑えられる。このように、 $\phi_{d_1, \dots, d_m}$  を用いた提案手法による複数次数の高次モーメント同時分析は、従来の単一次数の高次統計量に基づいた分析に対して、より大きな次元拡張と統計的バイアスの抑圧を達成している。

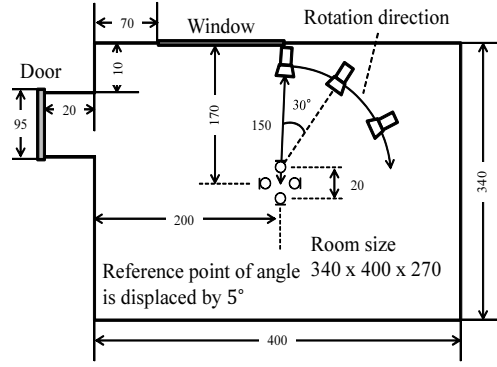


Fig. 1 Recording environment.

一方、異なる次数のモーメントを同時分析する場合は観測信号のスケーリングが問題となる。(23)–(25)に示すように、 $\mathbf{R}_{d_1, \dots, d_m}^\oplus(\omega)$  は様々な次数のモーメントを含んでいるが、モーメント次数の高い小行列の要素ほど低次小行列のそれに対して、観測信号ベクトルのノルム  $\|\mathbf{x}(\omega, t)\|$  の大小に伴って極端に値が増減する。観測信号をそのまま用いると、大きい信号成分が含まれる帯域とそれ以外の帯域では、複数次数モーメントの分析の基準が異なったものとなるため、モーメント同時分析を正しく行うためには異なる次数の要素全てを適切に重み付ける必要がある。本稿では各周波数ビンにおいて、 $\phi_{d_1, \dots, d_m}$  の要素中で最大次数となる  $d_m$  を基準として、 $L_{2d_m}$ -ノルムによりこの重み付けを行う。

$$\mathbf{x}'(\omega, t) = \frac{w\mathbf{x}(\omega, t)}{2^{d_m} \sqrt{E \left[ \frac{1}{M} \sum_{i=1}^M |x_i(\omega, t)|^{2d_m} \right]}} \quad (26)$$

ただし、 $w$  は  $L_{2d_m}$ -ノルム基準の正規化を調整するためのパラメータである。提案手法による分析では、 $w$  が大きい場合は高次モーメントを重視したものとなり、 $w$  が小さい場合はその逆となる。

## 5 実験

本章では、提案手法の有効性を検証するために行った実験について述べる。推定性能の比較のための従来手法として、写像 MUSIC ( $d = 2, 3$ ) と  $2q$ -MUSIC ( $q = 2, 3$ ) を用いた。

### 5.1 実験条件

図1に示す環境で収録した伝達関数による日本語音声の畳み込み混合を用いて、提案手法の性能評価を行った。観測信号には、拡散性雑音 [9] を模擬したピンクノイズを異なる3つのSNRで重畳した。優決定条件と劣決定条件それぞれの下での推定性能を検証するために、異なる2つの音源数で評価を行った。他の実験条件については表1に示す。

方向推定の性能評価には、 $\{0^\circ, 30^\circ, \dots, 330^\circ\}$  の12方向にランダムに配置した3または5音源を100通

Table 1 Experimental conditions

Sensor array configuration		Circular array with radius of 0.1 m	
Sound sources		Speakers 1.5 m apart from array	
# of sound sources	3, 5	# of sensors	4
Sampling frequency	16 kHz	SNR	20 dB
Room size	3.4 m × 4 m × 2.7 m		
Reverberation	T <sub>60</sub> of 0.3 s		
Speech duration	1 s		
Frame length	512 samples		
Frame shift length	256 samples		
Window function	Hamming window		

り用意し、真の音源方向と推定方向の間の2乗平均平方根誤差 (RMSE) を用いた。音源数分のピーク探索結果が得られず正しく推定方向を得られなかった場合は、全ての推定方向のエラーを平均したペナルティを与えた。また、(13)–(16) に示す写像空間の信号部分空間の次元数  $N'$  は、3 音源では  $N' = 15$ 、5 音源では  $N' = 30$  を用い、(26) に示す正規化パラメータ  $w$  は  $w = 2$  を用いた。なお、 $M \leq N$  における MUSIC の方向推定は、最小固有値に対応する一次元部分空間を雑音部分空間とみなして行った。

## 5.2 実験結果

図 2–3 に音源数ごとに得られた各手法の RMSE を示す。全ての条件の下で提案手法が最も優れており、写像 MUSIC は 2q-MUSIC に対してやや良好な結果となっている。写像 MUSIC の次数  $d = 2$  と  $d = 3$  の結果を比較すると、前者の方が優れた性能を示し、2q-MUSIC の  $q = 2$  と  $q = 3$  の比較についても同様の傾向が現れている。これらの結果は、[3] で検証されているように、次数の高い統計量ほどより大きなバイアスを受けることによるもので、今回の実験に用いた時間長ではスナップショットの数が不十分であったと考えられる。他方、提案手法と写像 MUSIC の比較は、複数次数のモーメント同時分析の有効性を示唆している。今回の条件では、6 次モーメントのみを用いた分析 ( $d = 3$ ) はやや劣った結果となっているが、提案手法の結果からこれを 4 次モーメント ( $d = 2$ ) の補佐的な情報として利用することで、推定性能が向上することを確認できる。また、提案手法は次数の高いモーメントの方が優れた推定結果を示す十分なスナップショット数の信号を分析する場合でも、安定化した高次モーメントとともに低次モーメントの情報を利用することとなるため、やはり単一次数のモーメント分析よりも優れた推定結果が期待できる。

## 6 おわりに

本稿では、劣決定条件に適用可能な新たな高分解能音源方向推定手法として、写像 MUSIC の拡張手法

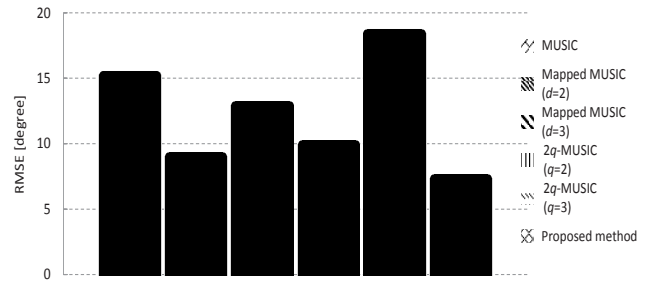


Fig. 2 Experimental results for 3 sources within 1 sec.

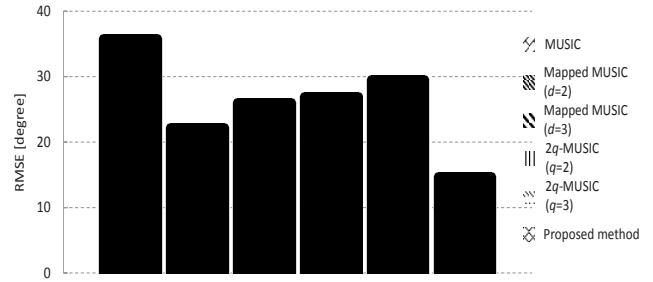


Fig. 3 Experimental results for 5 sources within 1 sec.

を提案した。従来の単一偶数次モーメント分析のための写像について示し、提案手法の方向推定アルゴリズムがこれを直和して得られる新たな写像に基づくことを述べた。提案手法による分析は、観測信号の複数次数の高次モーメント同時分析に相当しており、高次モーメントの分解能の高さと低次モーメントの推定における安定性の両方の利点を備えたものであることを述べた。単一の高次統計量分析に基づいた従来手法との推定性能比較実験によって、複数次数の高次統計量の同時分析が有効であり、提案手法により極めて高い推定性能が得られることを確認した。

謝辞 この研究は、科研費 23760325 の支援を受けたものである。

## 参考文献

- [1] R. O. Schmidt, *IEEE Trans. Antennas Propag.*, vol.AP-34, pp.276–280, 1986.
- [2] B. Porat *et al.*, *IEEE Trans. Signal Process.*, vol.39, pp.2016–2024, 1991.
- [3] P. Chevalier *et al.*, *IEEE Trans. Signal Process.*, vol.54, pp.2986–2997, 2006.
- [4] Y. Sugimoto *et al.*, *Proc. IWAENC*, 2012.
- [5] H. Saruwatari *et al.*, *IEICE Trans. Fundam.*, vol.E82-A, pp.1501–1510, 1999.
- [6] H. Kamiyanagida *et al.*, *IEICE Trans. Fundam.*, vol.E84-A, pp.999–1010, 2001.
- [7] S. Miyabe *et al.*, *Proc. ICASSP*, pp.1641–1644, 2009.
- [8] M. Wax *et al.*, *IEEE Trans. Acoust. Speech Signal Process.*, vol.32, pp.817–827, 1984.
- [9] I. A. McCowan *et al.*, *IEEE Trans. Speech Audio Process.*, vol.11, pp.709–716, 2003.