

TR-S-0017

マイクロホンアレーを用いた適応型ビームフォーミング  
におけるアレー信号処理An array signal processing for adaptive  
beamforming with microphone array矢代 智也  
Tomoya Yashiro西浦 敬信  
Takanobu Nishiura

2001.2.23

現在、マイクロホンアレーを用いた遠隔発話音声の高音質（高SN比）受音が注目されている。しかしマイクロホンアレーを用いて音声を高音質に受音するためには話者方向推定が必要となる。そこで本報告では方向推定法として、遅延フィルタにより信号のパワーを利用する Delay and Sum Array(DSA)、信号の相関行列に対する逆行列を利用する Mimum Variance method (MV)、相関行列から推定した信号部分空間と雑音部分空間の直交性を利用する MUSIC の3手法により評価実験を行い、比較・検討を行う。その結果 DSA は残響下でも安定しており、マイクロホン数が多いほど精度が高くなることを確認した。次に MV は暗騒音に強く、マイクロホン間隔を大きくとるほど分解能が高くなる傾向があった。最後に MUSIC は特徴として鋭い指向性を持つため、音波数が既知の場合有用な方法であること確認した。今後の検討として残響に頑健なアレー信号処理法の検討が必要であると考えられる。

## 1 はじめに

現在の音声入力による機器の制御などにおいて、残響や雑音にとらわれない高音質な音声を受音できる事が望まれている。そのため、発話者はマイクロホンに接近して発話する必要がある。マイクロホンに接近して発話することにより、雑音は低減され、高い信号対雑音比 (SN比) が得られる。しかしその場合、発話者は常にマイクロホンに接近しているため、非常にマイクロホンに対し、煩わしさを感じる。従って、望むべくはマイクロホンから離れた発話に対する高音質受音である。そこで現在は発話者がマイクロホンを意識せずに任意の位置での高音質受音すなわちハンズフリー受音に関する研究が盛んに行われてきている。特に近年は多数のマイクロホン素子を利用して雑音、残響などの低減を抑えかつ信号対雑音比の高い音声受音の方法 (マイクロホンアレー) が注目されている。

高い SN 比で受音を行うためには、目的信号の到来方向に対して感度が高く、雑音の到来方向に対しては感度が低い指向性受音器の利用が有効である。以前の信号処理を利用しない従来形機器において、目的の方向を変化させるためには、受音器の向きを移動させる必要があった。この点に関して、本報告で扱うマイクロホンアレー信号処理を利用すれば機器を固定したままで電氣的に指向性ビームの方向制御が可能となり使いやすさの面で格段の向上が得られる。以上のように、高 SN 比での受音を行うためには指向性を持つ受音器が必要となり、これはマイクロホンアレーを用いることによって実現できるのである。

マイクロホンアレーを用い、高音質の受音が可能となるマイクロホンの指向性形成の種類は大きく分けると 2 つある。

- 遅延和アレー  
発話者に対して指向性を形成。
- 適応型アレー  
雑音や、残響に対して死角を形成。

ここでは遅延和アレー (以下 DSA) と、適応型アレーから最小分散法 (以下 MV)、MUSIC の計 3 手法を用いてビームフォームのための音源方向推定に対するの評価実験を行い、それによる評価と考察を述べる。

本報告の構成は以下の通りである。

2. で各手法の概略を述べ、3. で各手法の音源方向推定法についての説明を行い、4. で実際の評価実験の方法を説明する。5. では、その実験に対する評価方法の説明をする。また 6. で結果を、7. で得られた結果に対し考察や、改善点を述べる。

## 2 音源方向推定の概略

方向推定法については様々な技術が行われており、その中で音の到来方向が（マイクロホン数 $M-1$ ）以下である場合には、音場はパラメトリックモデルに適合し、高い分解能の方向推定を行うことができるためそのパラメトリックな方法の代表として、MV、MUSICの2手法を用いることにし、さらに時間領域のDFTに対応したノンパラメトリックな手法としてDSAを用いた。以下で、各手法の簡単なアルゴリズムを説明する。

- DSA

複数のマイクロホンからのそれぞれの信号を遅延フィルタに掛けた後、総和を求め、総和信号のパワーを調べる。遅延フィルタが角度の関数になっているため、フィルタを用いて全方向のパワー分布を調べることにより方向推定を検知する方法である。

- MV

複数のマイクロホンからのそれぞれの信号から、それぞれの相関関数の行列である相関行列を求める。次に相関行列の逆行列を求め、ある角度に擬似的な相関のピークを持つ方向制御ベクトルと乗算を行い、その方向制御ベクトルを全方向でサーチすることによるパワー分布から方向を推定する。

- MUSIC

MVと同様に相関行列を求めた後、固有値、固有ベクトルを求め、その固有ベクトルの雑音情報のみの部分を用いて、雑音相関行列を計算する。方向制御ベクトルと雑音相関行列の乗算をピークを求める式の分母におくことにより理想的な方向に方向制御ベクトルのピークを持つとき、MUSICの特徴である

「信号部分空間と雑音部分空間は直交する」

を用いると、方向制御ベクトルは理想方向で雑音相関関数との積によりその解は極限に0に近づくため、全方向で解の逆数をとることによりピーク分布を表す。

以上、次節では、さらに詳しい各推定方法の説明を行う。

### 3 各手法の説明

#### 3.1 Delay and Sum Array (DSA)

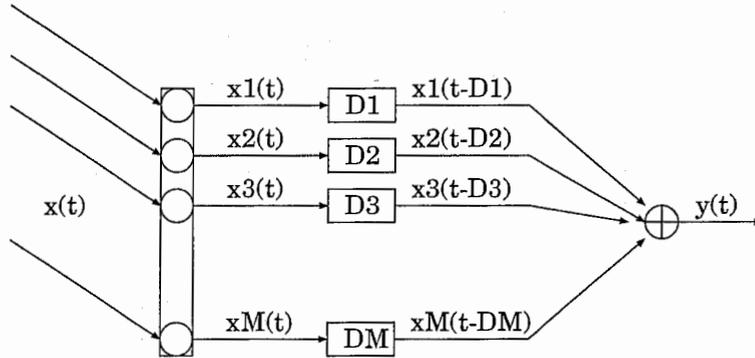


図 1: DSA の原理

ここでは各マイクロホンで受信した信号に遅延を付加した後総和をとる方法である、DSA について説明する。図 1 に DSA のアルゴリズムを示す。ここで  $x(t)$  はある方向から到来する時間軸に関する入力信号であり、等間隔におかれたマイクロホンに到達するときにそれぞれに遅延が付加され  $x_1(t), x_2(t), \dots, x_M(t)$  という信号になる。 $y(t)$  はある時点の遅延フィルタを通したそれぞれのマイクロホン信号の総和であり、遅延フィルタを変化させ、 $|y(t)|$  を求めることによりあらゆる方向でのパワー分布が現れる。

$M$  個のマイクロホンへの入力信号に対し、次式で示される遅延量  $D_i$  がそれぞれに付加される。

$$i \text{ 番目のマイクロホン受信信号に付加される遅延 } D_i = D_0 - (i-1)\tau_L \quad i = 1, 2, \dots, M \quad (1)$$

$$\tau_L = (d \sin \theta_L) / c \quad (2)$$

ここで  $\theta_L$  は  $x(t)$  の目的角（感度を高くする方向）であり、 $d$  はマイクロホン間隔、 $c$  は音速を表す。また  $D_0$  は、因果律を満たすための固定遅延量である。

このとき、目的信号が  $\theta_L$  から到来していると仮定すると、各マイクロホンで受け取る信号は

$$x_i(t) = x_1(t - (i-1)\tau_L) \quad (3)$$

で表せるため、式 (2) を用いて表すと、

$$x_i(t - D_i) = x_1(t - (i-1)\tau_L - D_i) = x_1(t - D_0) \quad (4)$$

となり、各遅延器を通った信号は全て同相化されて、それらを加算しその時点での  $|y(t)|$  を求めることで最大パワーとなり、目的方向  $\theta_L$  は強調されるのである。

一方、目的方向以外の角度では、 $x_i(t)$  それぞれは、同相化しておらず、これらを加算し  $|y(t)|$  を求めても、パワーは小さく強調されない。

この実験では遅延を付加するのにデジタルフィルタを用いる。たとえば、信号を  $D$  だけ送らせる遅延器の周波数特性は  $H_D(\omega) = e^{-j\omega D}$  となる。ここでその  $D$  が遅延周期  $T$  の倍数分であるのなら離散信号をその数

サンプル分遅延させればよいことだが、そうでない場合は先ほどの周波数特性を逆フーリエ変換を行い、

$$\begin{aligned} h_D &= \frac{1}{4\pi f_{max}} \int_{-2\pi f_{max}}^{2\pi f_{max}} e^{-j\omega D} e^{j\omega t} d\omega \\ &= \frac{1}{2\pi f_{max}(t-D)} \sin(2\pi f_{max}(t-D)) \end{aligned} \quad (5)$$

となる。ただし、積分の前の定数は、 $h_D(D) = 1$  とするための正規化定数である。またこれは  $\frac{\sin A}{A}$  の形から sinc 関数であることがわかる。この式の  $h_D(t)$  を適当に切り出して標準化したものが遅延量  $D$  を付加する FIR フィルタとなる。

従って実験では遅延を付加するのに信号  $x(t)$  と  $h_D(t)$  を畳み込む、つまり  $X(\omega)$  と  $H_D(\omega)$  を乗算する。

このデジタルフィルタを用いて、ある遅延を掛けた場合での全信号のパワーは

$$|y(t-D)| = \left| \sum_{n=1}^M h_{(n-1)D} * xn(t - D_0 - (n-1)D) \right| \quad (6)$$

$$DOA = \underset{\theta}{\operatorname{argmax}}(y^2(\theta)) \quad y^2(\theta) : \theta \text{方向のパワー} \quad (7)$$

となり、この遅延量  $D$  が方向の関数となり、到来方向推定が検知できる。

### 3.2 Minimum Variance method (MV)

MV は適応型アレーを用いた 1 つの方法であり、注目している方向  $\theta$  以外から到来する音の影響を最小にする推定方法である。そのため、目的音の大きさを変化させずに雑音を最小化する適応型アレーが利用される。この条件を満たす適応型アレーのフィルタの、各周波数  $\omega$  の信号に対するゲインを  $w_i, i = 1, \dots, M$  と表す。このとき  $w_i$  は次式によって与えられる。

$$w(\theta)^* = [w_1, \dots, w_M] = \frac{d(\theta)^* R^{-1}}{d(\theta)^* R^{-1} d(\theta)} \quad (8)$$

ここで  $d(\theta)$  は

$$d(\theta) = [1, e^{-j\omega\tau}, e^{-j\omega 2\tau}, \dots, e^{-j\omega(M-1)\tau}]^T \quad (9)$$

$\omega$ : 角速度

$\tau = (d \sin \theta) / c$

$d$ : マイクロホン間隔  $c$ : 音速

と定義されている方向制御ベクトルである。また  $R$  は

$$R = E[x(t)x(t)^*] \quad (10)$$

$$x(t) = [X_1(t), \dots, X_M(t)]^T \quad (11)$$

と定義されている入力信号の相関行列 (correlation matrix) である。

各マイクロホンで受信された狭帯域信号  $X_i(t)e^{j\omega t}, i = 1, 2, \dots, M$  に、このフィルタゲイン  $w_i$  を乗算した後、総和をとってアレー出力  $y(\theta, t)$  とする。この式は式 (9) をもちいて、 $w(\theta)^* x(t)e^{j\omega t}$  と、表される。最小分散法は、このアレー出力  $y(\theta, t)$  のパワー  $P(\theta)$  を、 $\theta$  方向から到来する音のパワーの推定値とする。

$P(\theta)$  は、

$$\begin{aligned} P(\theta) &= E[|y(\theta, t)|^2] = E[|w(\theta)^* x(t)|^2] \\ &= w(\theta)^* E[x(t)x(t)^*] w(\theta) = w(\theta)^* R w(\theta) \\ &= \frac{1}{d(\theta)^* R^{-1} d(\theta)} \end{aligned} \quad (12)$$

と表される。

この MV は、到来方向が明確なピークとして推定された場合には、そのピーク値は到来音のパワーに比例した値になっている。従って、複数の音が到来しているとき、得られた推定値のピークの大小関係は、そのまま到来音のパワーの大小関係を表しているという特徴を持つ。

### 3.3 MULTIPLE SIGNAL CLASSIFICATION (MUSIC)

この手法も MV 法と同じく、相関行列  $R$  に依存する。

方向  $\theta_q$  から音波が到来しており、1 番目のマイクロホンで受信される複素振幅が  $S_q(t)$  であるとする場合、 $i$  番目のマイクロホンで受信される複素振幅は  $S_q(t)e^{-j\omega(M-1)\tau_q}$ ,  $\tau_q = (d \sin \theta_q)/c$  と表されるため、これをベクトルを用いて表すと

$$[S_q(t), S_q(t)e^{-j\omega\tau_q}, S_q(t)e^{-j\omega 2\tau_q}, \dots, S_q(t)e^{-j\omega(M-1)\tau_q}]^T \quad (13)$$

$$= S_q(t)d_q \quad (14)$$

$$d_q = [1, e^{-j\omega\tau_q}, e^{-j\omega 2\tau_q}, \dots, e^{-j\omega(M-1)\tau_q}]^T \quad (15)$$

と表され、 $K$  個の音波が到来しているときの、受信信号の複素振幅ベクトル  $x(t) = [X_1(t), \dots, X_M(t)]^T$  は、それらの和として、

$$x(t) = \sum_{q=1}^K S_q(t)d_q + n_m(t) \quad (16)$$

と、表される。ここで、 $n_m(t)$  は、各マイクロホンにおける雑音を表すベクトルである。このとき、到来する各音波および雑音がそれぞれ無相関であると仮定すると、相関行列は、

$$R = E[x(t)x(t)^*] = \sum_{q=1}^K P_{s_q} d_q d_q^* + \rho I \quad (17)$$

となる。但し、 $P_{s_q}$  は信号のパワー  $E[|S_q(t)|^2]$  を、 $\rho$  は雑音のパワーを、 $I$  は単位行列を表す。また、 $d_1 \dots d_K$  はそれぞれ独立なベクトルであるので、右辺第一項のランクは  $K$  である。また、一項二項ともエルミート行列なので、式 (15) はユニタリ行列  $V$  によって対角化され、固有値が得られる。

$$V^* R V = \text{diag}[\lambda_1, \dots, \lambda_K, 0, \dots, 0] + \rho I \quad (18)$$

但し、 $\lambda_1, \dots, \lambda_K$  は式 (15) の固有値を表す。ここでユニタリ行列  $V$  は固有ベクトル  $v_1, \dots, v_M$  を用いて  $V = [v_1, \dots, v_M]$  と表すことができ  $VV^* = V^*V = I$  という性質から式 (15) の左に  $V$ 、右に  $V^*$  を掛けると

$$R = \sum_{q=1}^K (\lambda_q + \rho) v_q v_q^* + \rho \sum_{q=K+1}^M v_q v_q^* \quad (19)$$

と表される。ここで、ユニタリ行列の列要素である  $v_1, \dots, v_M$  は互いに直行しているので信号部分空間と、 $v_{K+1}, \dots, v_M$  を基底とした雑音部分空間 (noise subspace) とは直交する。従って信号部分空間の要素である信号の到来方向ベクトル  $d_1, \dots, d_K$  と  $v_{K+1}, \dots, v_M$  とは直交する。

このように MUSIC は、相関行列の雑音部分空間と音の到来方向ベクトルとの直交性を利用して、次式で到来方向を推定する方法である。

$$P(\theta) = \frac{1}{d(\theta)^* R_n d(\theta)} \quad (20)$$

但し、雑音相関行列  $R_n$  は、次式で示すように  $v_{K+1}, \dots, v_M$  より構成される。

$$R_n = \sum_{q=K+1}^M v_q v_q^* \quad (21)$$

つまり、 $d(\theta)$  が実際の音の到来方向ベクトルの時、 $R_n$  との性質から  $R_n$  と  $d(\theta)$  は直交するため、0 に限りなく近づく。その他の  $d(\theta)$  においては  $R_n$  と直交しないため 0 にはならない。

この方法では  $d(\theta)$  が実際の音の到来方向ベクトルの時、 $P(\theta)$  が鋭いピークを持つ上、実際の到来パワーに比例しないため音の到来方向の推定のみ有効である。また、この方法に限り、到来音波の数は既知でなければならない。その理由として相関行列の固有値の絶対値の大きい順に  $K+1$  番目から  $M$  番目に対応する固有ベクトルは、雑音のみの情報を含んでいる固有ベクトルとなり、それらのみを取り出すことによって雑音相関行列  $R_n$  が作られるからである。同様に固有値の大きい順に 1 番目から  $K$  番目までは目的信号情報のみを含んだ固有ベクトルに対応している。従って、音波数が既知でない場合、相関行列の固有値、固有ベクトルから雑音情報のみを取り出すことは非常に困難である。

### 3.4 各手法での指向性の比較

ここでは、推定方法の違いによる指向性の比較を行う。図 2 に、2 マイクロホン素子で 1 つの音源が  $90^\circ$

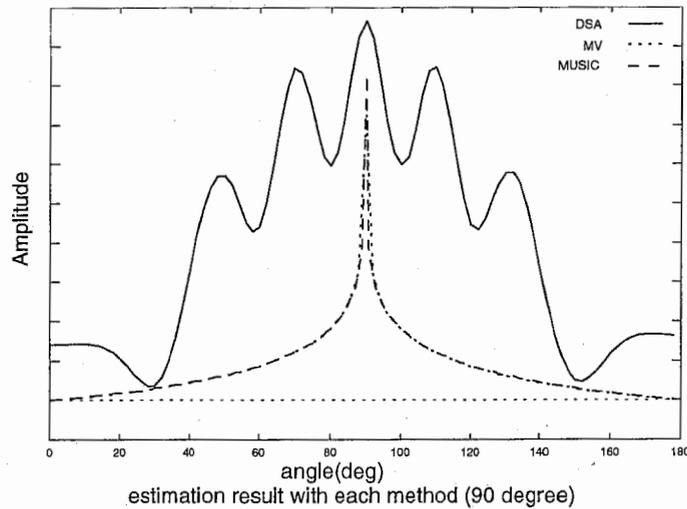


図 2: 3 手法を用いた指向性の比較

に存在するときの場合の各々の指向性を示す。

DSA は、理想的な  $90^\circ$  の方向に限らず、その側面方向にも小さいピークがみられる。これは信号の総和がパワーとなっているため、高まりあう部分と打ち消しあう部分が現れるためである。

MV は、理想の方向のみに鋭い指向性を形成している。このピーク値は到来信号のパワーに比例した高さとなるので、パワーが大きい信号を用いることによってより大きいピークが期待できる。

また、MUSIC は比較的 MV よりもさらに鋭い指向性を示しており、この方法では実際の到来パワーにピークが比例しないため、パワーが小さい信号の推定でも容易に行うことができる。

## 4 評価実験

本報告では、DSA、MV、MUSICの3手法の方向推定を定量的、定性的に比較するため計算機シミュレーションと実環境において評価実験を行う。そのためには実際のある方向から到来する発話信号を複数のマイクロホンで受音したデータが必要となるため、評価実験のためのデータ収録を実験室を用いて行った。

### 4.1 収録環境

実験室の概要を図3に示す。この実験では、吸音状態と反響状態において測定を行った。また、それぞれの条件で、実験室での収録環境を表1に示す。図に示すように30、60、90、120、150度でのデータ

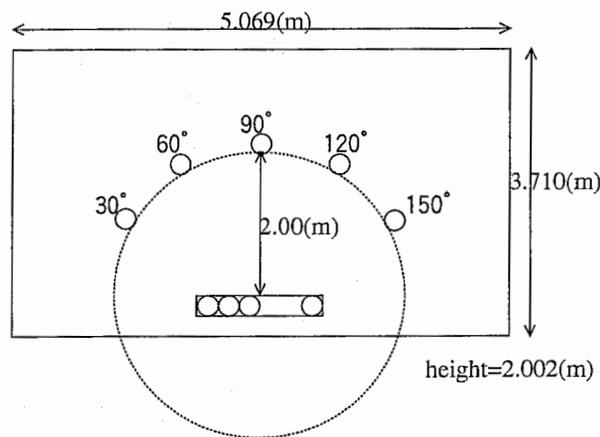


図3: 収録環境

にはそれぞれ話者データとしてATR音声データベースsetAのMTT,FTK,MHT,MTK,FKNを用いた。

表1: 収録条件

	吸音状態	反響状態
残響時間	0.27(sec)	0.75(sec)
幅、奥行き、高さ	5.069, 3.710, 2.002(m)	
騒音レベル	25.1dBA	
サンプリング周波数	16(kHz)	
室温	23( $^{\circ}$ C)	
テストデータ	ATR 音声データベース setA	
マイクロホンアレーの仕様		
素子数	15	
1素子間隔	2.125(cm)	

## 4.2 比較報告

本実習では実環境データだけでなくシミュレーションデータにおける性能評価実験も行った。実験に用いたデータは、2種類の実環境データに加え、それぞれの推定方法の特徴がより明確になるために、到来音波を擬似的に平面波に近づけ、暗騒音、残響を取り除いたデータとそれに暗騒音のみ付加したデータの2種類も用いた。

表2：各々のデータ情報

名称	暗騒音	残響
Simulation data 1	×	×
Simulation data2	○	×
Real data 1	○	$T_{[60]} = 0.27sec$
Real data2	○	$T_{[60]} = 0.75sec$

## 5 評価方法

DSA、MV、MUSIC のおのおのについて ATR 音韻バランス単語 216 語を全て、5つの角度（話者）、4種類の環境によるデータで2素子1音源、3素子1音源、3素子2音源の3つの条件につきシミュレーションを行った。1音源推定については、おのおの方法、部屋で角度ごとの方向推定結果216個の平均値と、分散に基づくエラーバーを用いて図5から図16に示した。

2音源推定については同時に同じ単語を違う方向から推定することはとても困難であるため、無作為に違う単語2つを発話させるように、64種類のデータを用いて推定をした。

### 5.1 DSA

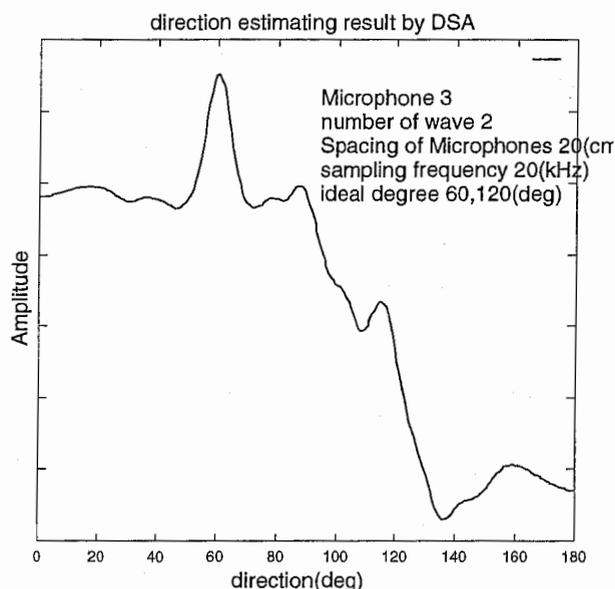


図4: DSA を用いた方向推定 (60°、120°)

DSA の方向推定において、図4のように幾つものピークが現れることが多い。そこで、方向推定の方法として、図のグラフを角度軸で微分し、幾つかの角度軸との交点での角度を取り上げ、その幾つかの交点の中から、元の図のピークの大きい角度を選ぶこととした。また、2音源の場合は方向推定グラフのピークを全て取り出し、ピークの大きいものから2つの角度を選ぶこととした。

### 5.2 MV、MUSIC

MV法の推定方向において、まれに理想音源数より推定音源数が少ない事や方向推定のピークが現れないという事がある。推定方法としては理想音源数1つ1つにつき、平均と分散を求める方法を用いているのでその場合には、その推定データは無視した。従って、おのおの推定結果について可測定数も調べた。

## 6 実験結果

1音源の方向推定に関しては、 $30^\circ$ ,  $60^\circ$ ,  $90^\circ$ ,  $120^\circ$ ,  $150^\circ$  の5方向を推定し、 $30^\circ$ ,  $60^\circ$ ,  $90^\circ$  と  $120^\circ$ ,  $150^\circ$  のグラフをそれぞれ4種類の環境について表す。また、その下には上の図と対応した可測定数を表すグラフをそれぞれ示した。「2マイクロホン、1音源」と「3マイクロホン、1音源」は216個中の可測定数であり、「3マイクロホン、2音源」は64個中の可測定数である。また表中の「216」は1音源の場合、「64」は2音源の場合に全て測定できた事を示す。

## 6.1 2 マイクロホン素子、1 音源

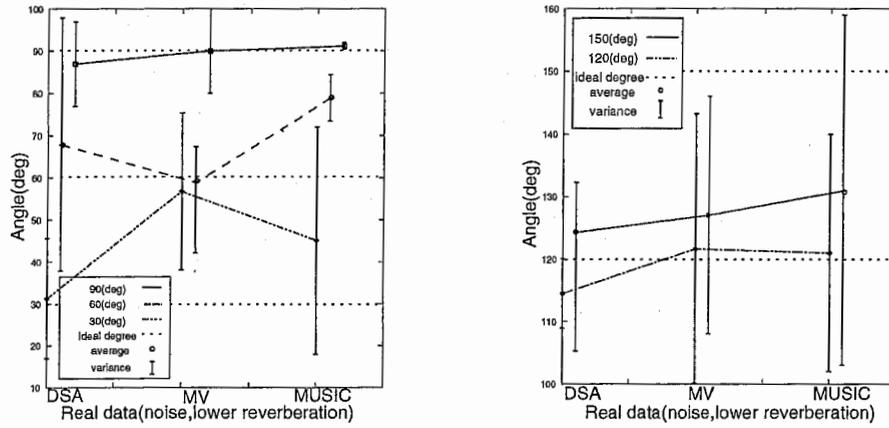


図 5: 実環境データ 1 を用いた方向推定結果

表 3: 実環境データ 1 での可測定数

	DSA	MV	MUSIC
30(deg)	216	208	194
60(deg)	216	197	207
90(deg)	216	278	214

	DSA	MV	MUSIC
120(deg)	216	214	218
150(deg)	216	208	206

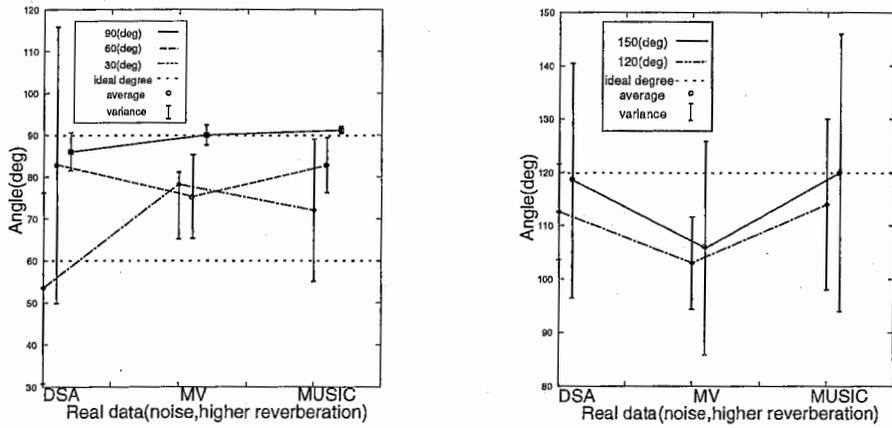


図 6: 実環境データ 2 を用いた方向推定結果

表 4: 実環境データ 2 での可測定数

	DSA	MV	MUSIC
30(deg)	216	122	208
60(deg)	216	58	216
90(deg)	216	86	214

	DSA	MV	MUSIC
120(deg)	216	152	215
150(deg)	216	185	213

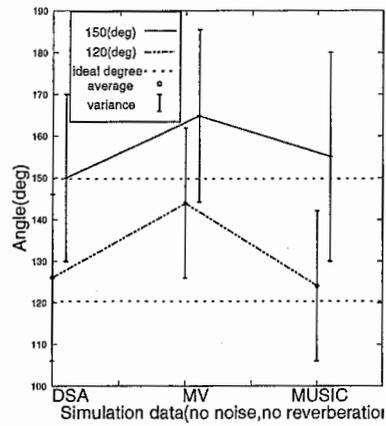
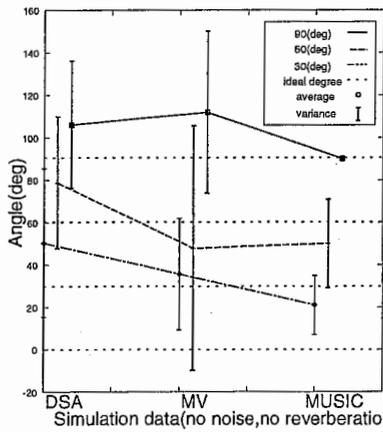


図 7: シミュレーションデータ 1 を用いた方向推定結果

表 5: シミュレーションデータ 1 での可測定数

	DSA	MV	MUSIC
30(deg)	216	172	184
60(deg)	216	171	200
90(deg)	216	8	216

	DSA	MV	MUSIC
120(deg)	216	182	215
150(deg)	216	206	213

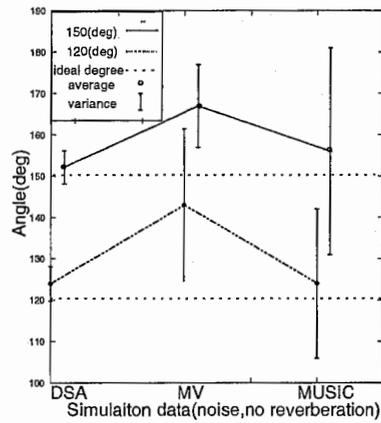
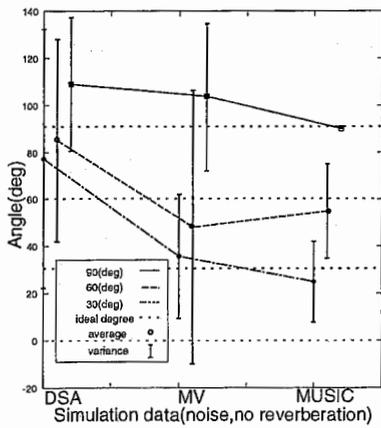


図 8: シミュレーションデータ 2 を用いた方向推定結果

表 6: シミュレーションデータ 2 での可測定数

	DSA	MV	MUSIC
30(deg)	216	172	208
60(deg)	216	172	216
90(deg)	216	13	1214

	DSA	MV	MUSIC
120(deg)	216	183	215
150(deg)	216	206	213

## 6.2 3マイクロホン、1音源

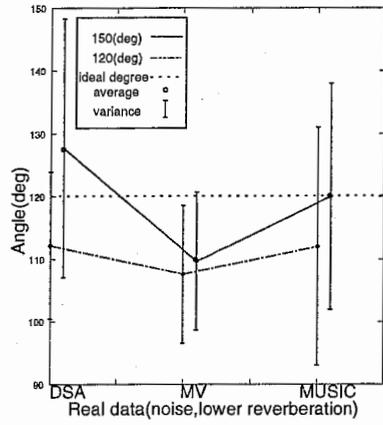
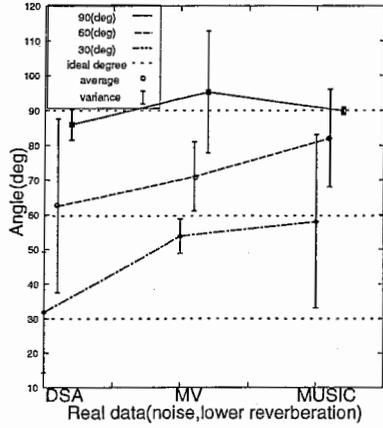


図 9: 実環境データ 1 を用いた方向推定結果

表 7: 実環境データ 1 での可測定数

	DSA	MV	MUSIC
30(deg)	216	216	216
60(deg)	216	216	216
90(deg)	216	216	216

	DSA	MV	MUSIC
120(deg)	216	216	216
150(deg)	216	216	216

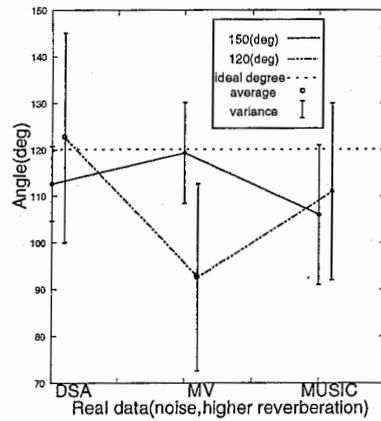
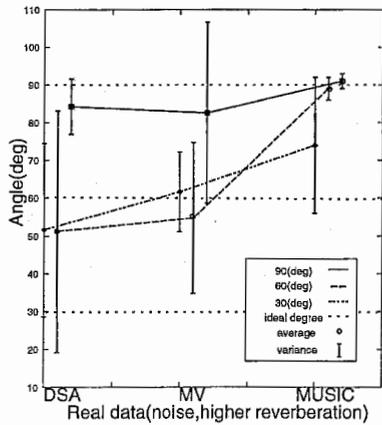


図 10: 実環境データ 2 を用いた方向推定結果

表 8: 実環境データ 2 での可測定数

	DSA	MV	MUSIC
30(deg)	216	216	216
60(deg)	216	216	216
90(deg)	216	214	216

	DSA	MV	MUSIC
120(deg)	216	216	216
150(deg)	216	216	216

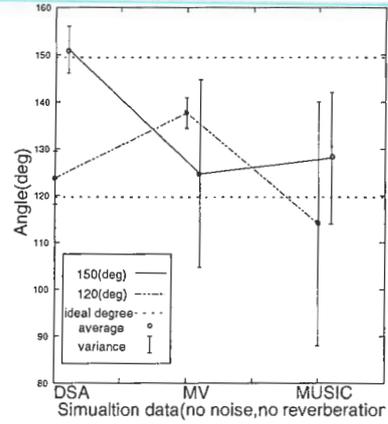
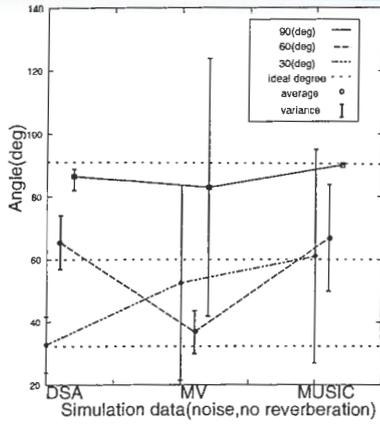


図 11: シミュレーションデータ 1 を用いた方向推定結果

表 9 : シミュレーションデータ 1 での可測定数

	DSA	MV	MUSIC
30(deg)	216	216	213
60(deg)	216	216	215
90(deg)	216	216	216

	DSA	MV	MUSIC
120(deg)	216	216	216
150(deg)	216	216	216

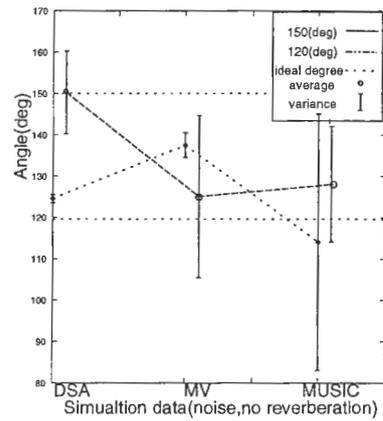
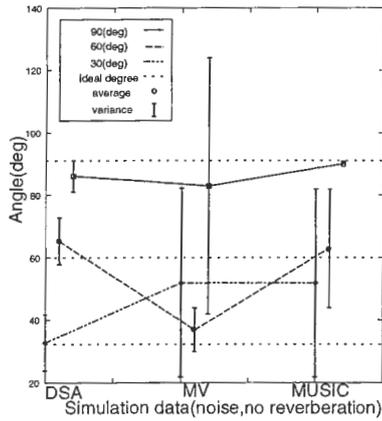


図 12: シミュレーションデータ 2 を用いた方向推定結果

表 10 : シミュレーションデータ 2 での可測定数

	DSA	MV	MUSIC
30(deg)	216	216	214
60(deg)	216	216	216
90(deg)	216	216	216

	DSA	MV	MUSIC
120(deg)	216	216	216
150(deg)	216	216	216

### 6.3 3マイクロホン、2音源

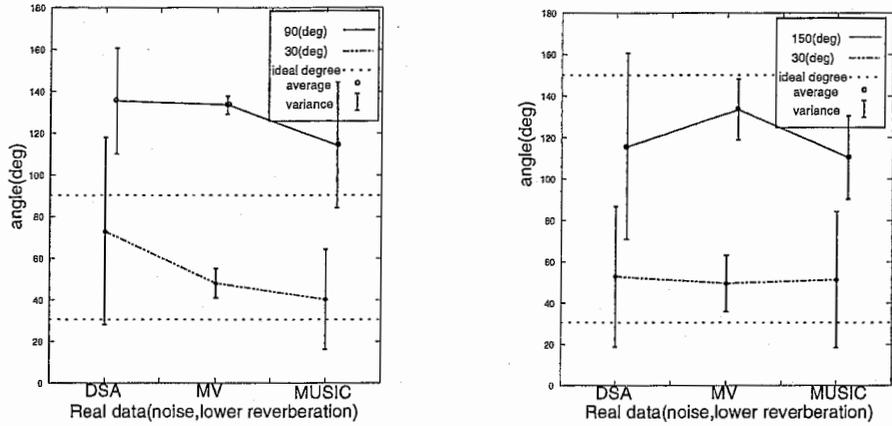


図 13: 実環境データ1を用いた方向推定結果 (左 30°, 90°, 右 30°, 150°)

表 11: 実環境データ1での可測定数

	DSA	MV	MUSIC
30(deg),90(deg)	64	64	13

	DSA	MV	MUSIC
30(deg),150(deg)	64	62	10

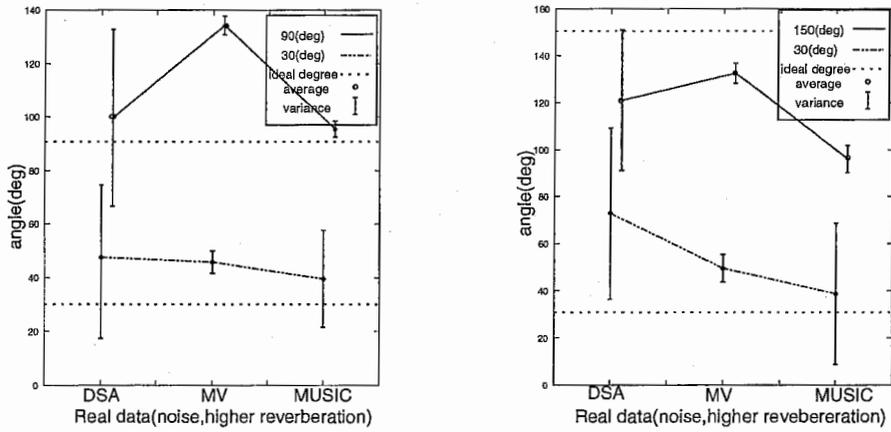


図 14: 実環境データ2を用いた方向推定結果 (左 30°, 90°, 右 30°, 150°)

表 12: 実環境データ2での可測定数

	DSA	MV	MUSIC
30(deg),90(deg)	64	64	10

	DSA	MV	MUSIC
120(deg),150(deg)	64	64	3

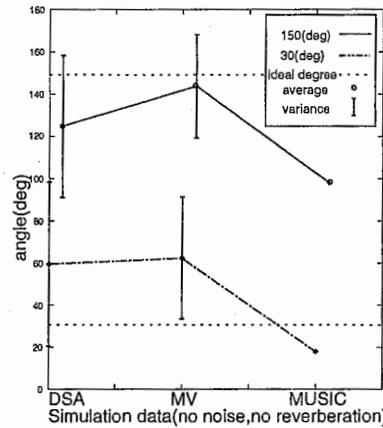
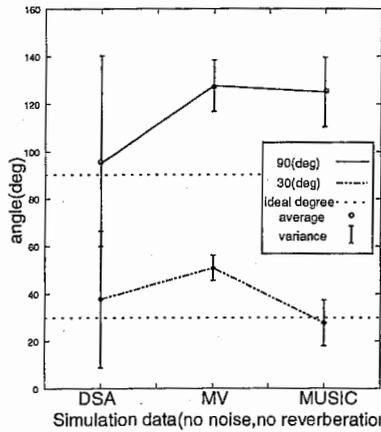


図 15: シミュレーションデータ 1 を用いた方向推定結果 (左 30, 90°, 右 30, 150°)

表 1 3 : シミュレーションデータ 1 での可測定数

	DSA	MV	MUSIC
30(deg),90(deg)	64	64	14

	DSA	MV	MUSIC
120(deg),150(deg)	64	64	1

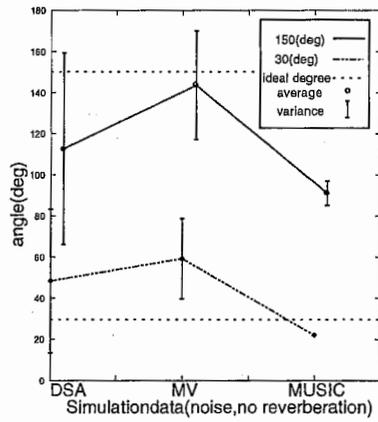
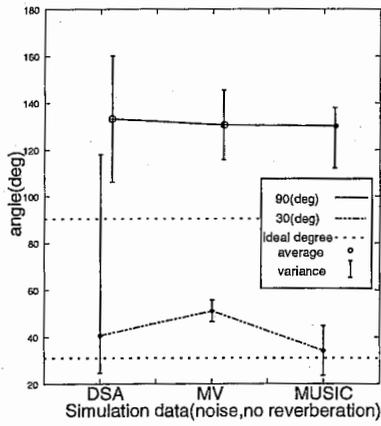


図 16: シミュレーションデータ 2 を用いた方向推定結果 (左 30, 90°, 右 30, 150°)

表 1 4 : シミュレーションデータ 2 での可測定数

	DSA	MV	MUSIC
30(deg),90(deg)	64	64	17

	DSA	MV	MUSIC
120(deg),150(deg)	64	64	2

## 7 考察

### 7.1 Delay and Sum Array (DSA)

本実験の DSA ビームフォーミングの特徴として、推定方向以外にも多くのピークが現れる事が挙げられる。これは、音声信号が波を持つためであり、他の波同士のパワーを計算してしまうことで生まれる。

部屋に残響がある場合、この DSA の方法を用いると、入力信号にその信号自身の反射音が重畳してしまい、その反射音と元の信号との時間的波形による重なりが生じ、それが大きいパワーになることによって誤った方向にピークが生じてしまう。実環境データでシミュレーションをしたところ、残響が高い部屋でのデータは残響が少ない部屋のデータに比べ全体的に分解能の高い  $90^\circ$  方向に集まった。これは明らかに残響の影響によるものであると考えられる。残響を除いたシミュレーションデータの結果をみると、 $120^\circ$ 、 $150^\circ$  方向にはほぼ理想的な方向推定ができていた。

次に暗騒音による影響を検討する。2種類のシミュレーションデータは、片方は暗騒音、残響ともに無く、もう一方は残響はないが、暗騒音は多少加えられている。暗騒音ありの時の方向推定結果と暗騒音なしの結果を比較した場合、やはり暗騒音が存在する環境下においての推定結果は精度が低い。DSA は暗騒音であろうともそれをある方向のパワーと捉えてしまうので誤差の原因となってしまうのであると考えられる。

マイクロホン数が多い場合の DSA は少ない場合と比べ、明らかに精度が高いことが結果からも分かる。少ないマイクロホンでは大きいパワーとして捉えられてしまった雑音や残響の影響も、数を増やすことにより、真に等しい信号波形のみパワーと比べ減衰して見えるため、方向推定のピークの精度向上につながると考えられる。4種類のデータそれぞれは2マイクロホン1音源に比べ、3マイクロホン1音源は推定誤差が少ないのがわかる。従って一般的に、この方法ではマイクロホン数が多い方が精度がよいということがこの結果から明らかになった。

ここでマイクロホン3素子のまま、音源数を2に増やしたときの結果を考察する。精度的にみた場合、結果はあまりよくないというのがみてとれる。 $30^\circ$ 、 $90^\circ$  の推定では  $120^\circ$  当たりに高いピークがでることが多く、また  $30^\circ$ 、 $150^\circ$  推定では  $90^\circ$  当たりに現れた。これは2つの異なった方向からの信号の和となる信号がそれぞれのマイクロホンに受信された場合、入力としてのそれぞれ2つの信号のパワーがピークとして現れると同時に、お互いのパワーも高いピークとして現れてしまい、その方向が推定方向として検出されてしまうからであると考えられる。ここでいえることは、DSA を用いる場合には、音源数と比べ、多くのマイクロホンを必要とするということである。マイクロホンは多ければ多いほど原信号のパワーを高め、高いピークとすることで雑音や、残響そして他信号同士のパワーを減衰させる事が可能であるといえるであろうと考えられる。

最後に、この方法では計算機を用いて方向推定する場合、推定角度ごとにそのときの全パワーの総和を計算するためとても時間がかかった。この計算時間は音声信号データの長さに比例するため、長い音声データの方向推定より、インパルスなどのごく短い信号データの推定に用いるべきだと考えた。

## 7.2 Minimum Variance method (MV)

MVはDSAと比べると、サイドローブが現れず、音源数だけのピークが現れる性質がある。これは、相関行列が収録環境に関する相関関数を持っており、それに方向制御ベクトルを用いることによって方向推定を行うからである。しかしこの方法では残響、暗騒音といった不必要な情報までもがこの相関行列の情報として加わってしまうため、高い残響環境ではよい結果が得られなかった。

2マイクロホン素子1音源の場合、実環境1のデータでの推定結果から $30^\circ$ 、 $150^\circ$ といった分解能が特に小さい方向からの推定誤差が大きくなっていった。ここから残響の特性としては、分解能の高さに反比例して影響することが考えられる。これは残響の相関が分解能の小さい $30^\circ$ 、 $150^\circ$ のそれぞれの高い相関を持つ地点から離れて現れる為だと考えられた。相関行列を用いる場合、相関のピークの情報として複素数を取り扱うため例を挙げて説明する。図17の左に、真の信号の相関関数に残響の相関関数が含まれたものを示す。

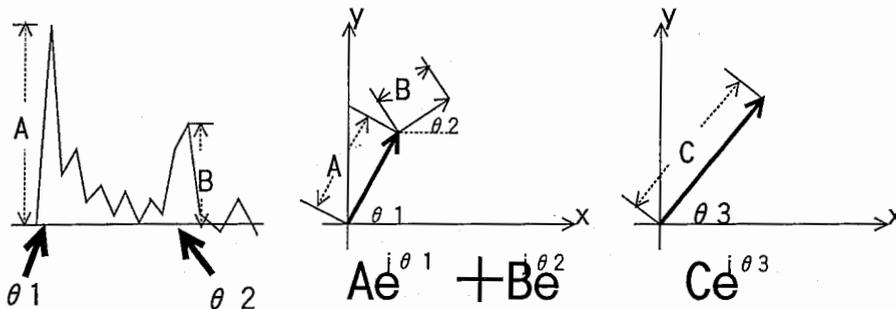


図17: 残響が及ぼす推定誤差

真の信号同士の相関は角度 $\theta_1$ でAのピークを持つものとする。またここで、残響による信号同士の相関が角度 $\theta_2$ でBのピークを持つものとし、その他のピークは十分小さいと考えると全体的な相関関数の複素表現では真中の図となる。2つのピークが複素ベクトルとして和算され、その結果図4右のように、角度 $\theta_3$ でCのピークとして現れ、その角度で認識されてしまうものと考えられる。

次に、2種類のシミュレーションデータ結果を比較すると、ほぼ推定方向、分散とも変わらない様子が見られた。具体例を挙げると図7, 8の $30^\circ$ ,  $60^\circ$ ,  $90^\circ$ である。これは、暗騒音同士の細かい相関ピークがが入力信号の相関ピークに対し小さく、無視できるものであると考えられるためである。たとえ暗騒音と入力信号が相関を持とうとも、暗騒音のパワーがとても小さいため、無視できるほどの小さい相関ピークとなってしまふ事が考えられる。

音源は1つのままでマイクロホン素子を3つにしたとき、精度が突然下がってしまったため、方向推定の波形を観察したところ図18のようなビームフォーミングが現れた。これは明らかに推定結果に誤りがあることを示している。この原因としては幾つかの事が考えられる。

- 相関行列の誤り
- 雑音や残響による相関行列の変形
- その他の環境による影響

ここで、シミュレーションデータの結果をみると、同じように分散の度合いが大きくなっていった。つまり2方

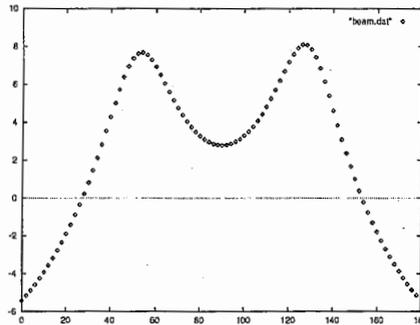


図 18: MV の誤った推定

向にピークが立ってしまい、推定方向が安定せず、分散の大きい推定になってしまったのである。従って、やはり相関行列の誤りであることが最大の原因であると考えられた。

3 マイクロホンの場合でもシミュレーションデータ 1 と 2 を比較したところ、ほぼ変わらない推定結果であった。結論としてはやはりノイズには強いという傾向がみられた。その上、残響がある場合に比べて精度も向上していた。

3 マイクロホン素子、2 音源の場合では同じようにほぼ左右対称のピークが現れる事が多く、残念ながら正確性の低い方向推定になってしまった。

最後に、この MV 法は全体的に残響により全方向推定が分解能の高い  $90^\circ$  方向によってしまう事、ノイズの相関パワーは入力信号の相関パワーに比べて無視できるためノイズに強いことが確認できた。また、マイクロホン数-音源数=1 で、特にマイクロホン間隔を大きくとった場合はより多くのマイクロホンを用いた場合に比べ精度が向上していた。従って相関行列はマイクロホン間隔を大きくとることにより精度が高まる事が確認できた。

### 7.3 MUSIC

2 マイクロホン 1 音源の場合に、実環境データ 1 と 2 を比較してみると、データ 2 の結果は 5 方向全体的に  $90^\circ$  に引き寄せられている。つまり、高い残響下での方向推定は難しいことがわかる。その原因を予想するに、残響は、発話信号が壁や床に当たり跳ね返ってできる擬似的音源であるといえるので仮想的にマイクロホンアレーに不特定多数の音波が受信されることになり、到来音波が既知である必要がある MUSIC は信号部分空間情報までも雑音相関関数に含まれてしまうため誤った方向推定をしてしまうと考えられる。

また、暗騒音に関してシミュレーションデータ 1 と 2 の比較の差がほとんどみられないことから、MV と同様に暗騒音に対してはあまり影響を受けることなく精度の高い方向推定が行えると考えられる。MUSIC は、収録環境内にある雑音を相関行列として持ち、その雑音行列と直交するように信号の行列を検知するため、高い雑音下でも正確な方向推定が行えるであろうと考えられるからである。

3 マイクロホン、2 音源の場合、MV と同じく、よい方向推定が行えたとは言えないような結果となった。MUSIC の実装に当たって、相関行列を作るまでは MV と同じ相関行列を用いたので MV での相関行列に誤りがあった可能性が大きい。また可測定数が非常に少なかったため、MUSIC を用いての複数音源の推定は難解であると感じた。

MUSIC を用いての全ての  $90^\circ$  方向の推定に注目してみると、3 手法の中では一番指向性が鋭く、またほとんど分散もみられない。従って、分解能が高い方向への指向性は正比例的に鋭くなるのがこの MUSIC の

特徴でもあることが考えられる。

一般的に MUSIC での方向推定ピークは鋭く、3手法の中では一番はっきりしていると考えられる。もし、到来音波数が既知であった場合、高い残響下でなければ、この手法が一番適しているということが、この MUSIC の結論である。

## 8 まとめ

本報告では、近年注目されている高音質 (SN 比の高い) 受音のためのアレー信号処理による 3 種類の方向推定法の比較、検討を行った。DSA は他の 2 方法に比べ、残響下に強い傾向があった。またマイクロホン数を増やすことにより、よりよい指向性が得られることもわかった。但し、指向性に小さいピークも現れるため誤差が発生しやすい事も確認できた。

MV は比較的ノイズに強い傾向があり、しかし残響下では大きな誤差が生まれたことから残響に影響を受けやすいことがわかった。また、マイクロホン間隔は相関行列の分解能を左右する 1 つのパラメータであることが確認できた。

MUSIC はピークが鋭いため、方向検知が容易であるが、残響下では擬似的に音波が増えるため MV とおなじく残響に影響を受けやすい事がわかった。

マイクロホンとスピーカーの間隔を広げると、マイクロホンへの入力信号が平面波に近づき、比較的精度の高い方向推定がおこなえる事がわかった。

最後に展望として、残響の影響を考慮に入れ、高い残響下でも正しく推定が行えるようなアレー信号処理法の検討をしたいと思う。

## 9 参考文献

- 大賀寿郎、山崎芳男、金田豊、”音響システムとデジタル処理、” (社) 電子情報通信学会、pp.173-217、1995.