

## 音源方向推定法の室内残響下での性能評価\*

田中雅史・金田 豊・小島順治\*\*

(NTT ヒューマンインタフェース研究所)

(平成5年7月26日受付)

**内容梗概** 音源の方向推定を室内で行う際に推定が良好に行える条件を検討した。信号源の方向を推定する手法は種々提案されているが、これらを室内で利用する場合には、室内反射音の影響が無視できないと考えられる。反射音の量は直接音と間接音のエネルギーの比である D/R により定量化した。直線アレーを直方体室の側壁と床に並行に配置し、音源は一つという条件で、方向推定手法、音源の方向、部屋の容積、音源の指向性について実験を行った。その結果、ほぼ、実験の条件によらず D/R が -5 ~ -10 dB の間に推定が良好にできる下限があることが分かった。

**キーワード** 方向推定, 残響, マイクロホンアレー, D/R, 推定誤差

Direction estimation, Reverberation, Microphone array, D/R, Estimation error

## 1. はじめに

音源方向の推定技術は、異常音発生方向の検出や、話者の移動に追従できる受音システムの構築などをはじめとして、利用範囲の広い基本技術である。

音源方向の推定は、これまで多くの手法が提案されている。例えば、スペクトル分析法に基づいた手法<sup>1),2)</sup>、複数の受音点での音源からの信号の到来時間の差を利用する手法<sup>3)-6)</sup>、また、人の頭部などの場合のように回折による周波数特性の違いを利用する方法がある<sup>7)</sup>。

さて、室内でこれらの音源方向推定技術を利用する場合には室内反射音の影響を無視することはできない。すなわち、音源からマイクロホンに受音される直接音は音源方向の情報を含んでいるが反射音の多くは実際の音源方向とは異なった方向から到来するのでこれらは推定の妨害要因となる。そして、室内反射音が多く存在するような環境下では音源方向を良好に推定することができないという問題が発生する。

従来の研究では、反射音の量と推定誤差の関係、例えば、どの程度の反射音の量であれば音源方向推定が可能であるのかというようなことに関しては議論がなされてこなかった。しかし、この関係を明らかにする

ことは実際の室内音場において音源方向推定技術を利用していくうえで大変重要なことである。

そこで本報告では、直接アレーを直方体の部屋で側壁と床に並行に配置するという条件で、代表的な信号源方向推定手法であるスペクトル分析法に基づいた手法を対象として、室内反射音の大小と方向推定の限界の関係について検討した。2章において、まず音源方向推定法について説明する。本報告では最も一般的な遅延和によるビームフォーミングと共に分解能の高い手法についても検討の対象とした。3章において、推定が良好に行える条件を直接音と反射音のエネルギーの比をパラメータとして実験的に示す。4章において以上の結果をまとめる。

## 2. 音源方向の推定法

本論文で検討対象とした音源方向推定手法としてはレーダ等の分野で代表的な3手法、すなわち、遅延和法 (DSA: Delay Sum Array)、最小分散法 (MV: Minimum Variance)、MUSIC (Multiple Signal Classification) 法を選んだ。これらの方向推定法を、まず、アレーへの入力が狭帯域信号の場合について説明し、次に、広帯域に適用する場合について述べる。

**Fig. 1** に直接マイクロホンアレーの構成と音波が方向  $\theta_s$  からマイクロホンアレーに入射している様子を示した。音波は中心帯域  $f$  の狭帯域信号であるとす。各時刻  $k$  におけるマイクロホン1での音波の複素振幅を  $s_1(k, \theta_s, f)$  と表すことにする。このとき、各

\* Performance of sound source direction estimation methods under reverberant conditions.

\*\* Masashi Tanaka, Yutaka Kaneda and Junji Kojima (NTT Human Interface Laboratories, Musashino, 180)

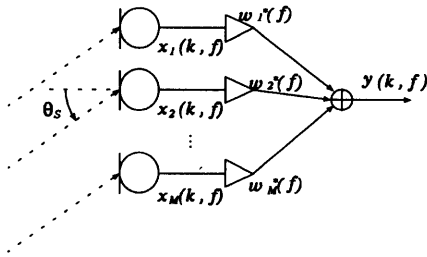


Fig.1 Schematic diagram of a linear microphone array beam former.

マイクロホンで受信される音波は周波数  $f$  と入射角  $\theta_s$  に依存した位相差を持つ。ここで、1 番目のマイクロホンでの受信信号を基準とした  $i$  番目のマイクロホンの受信信号の位相を  $\phi_i(\theta_s, f)$  とする。そうすると、各マイクロホンで受信される信号の複素振幅  $x_i(k, f)$ ,  $i=1, \dots, M$  ( $M$  はマイクロホンの総数) は

$$x_i(k, f) = s_i(k, \theta_s, f) e^{j\phi_i(\theta_s, f)} + n_i(k, f) \quad (1)$$

と表される。ここで、 $n_i(k, f)$  は雑音を表す。Fig.1 に示したアレー処理系では、受信した信号  $x_i(k, f)$  を  $w_i(f)$  の重み付けをした後、すべてのマイクロホンに関しての和をとってアレーの出力を得る。

$$y(k, f) = \sum_{i=1}^M x_i(k, f) w_i^*(f) \quad (2)$$

ただし、 $*$  は複素共役を表す。

DSA 法では、雑音の影響が無視できれば、アレーの出力のパワーは重みが  $w_i(f) = \exp(j\phi_i(\theta_s, f))$  のときに最大になることを利用し音源の方向を推定する。音波がある方向から到来していると仮定し、それに応じて重み  $w_i(f)$  を変化させアレーの出力のパワーが最大となる方向を探す。 $w_i(f) = \exp(j\phi_i(\theta, f))$  とすることはアレーの指向性の面から見ると、仮想した音源方向  $\theta$  に指向性のビームの最大値が向くようにすることに相当する。方向  $\theta$  の音源を仮定した時の DSA 法に基づいた  $\theta$  方向からの入射パワーの推定値  $P_{\text{DSA}}(\theta, f)$  は次式で表される。

$$\begin{aligned} P_{\text{DSA}}(\theta, f) &= \frac{1}{K} \left| \sum_{k=1}^K y(k, f) \right|^2 \\ &= \frac{1}{K} \left| \sum_{i=1}^M \sum_{k=1}^K x_i(k, f) w_i(k, f) w_i^*(f) \right|^2 \\ &= \frac{1}{K} \left| \sum_{i=1}^M \sum_{k=1}^K e^{-j\phi_i(\theta, f)} x_i(k, f) \right|^2 \quad (3) \end{aligned}$$

この式を行列、ベクトル表示するために、次のベクトル、行列を定義する。

$$\begin{aligned} \mathbf{d}(\theta, f) &= (1, e^{-j\phi_2(\theta, f)}, \dots, e^{-j\phi_M(\theta, f)})^H \quad (4) \\ \mathbf{x}(k, f) &= (x_1^*(k, f), x_2^*(k, f), \dots, \end{aligned}$$

$$x_M^*(k, f))^H \quad (5)$$

$$\mathbf{R}(f) = \frac{1}{K} \sum_{k=1}^K [\mathbf{x}(k, f) \mathbf{x}(k, f)^H] \quad (6)$$

ここで、 $H$  は共役転置を表す。また、定義より  $1 = e^{-j\phi_1(\theta, f)}$  である。 $\mathbf{d}(\theta, f)$  は周波数  $f$ 、方向  $\theta$  の音源のマイクロホン上での複素振幅の比からなるベクトル、 $\mathbf{R}(f)$  は受信した信号のマイクロホン間の共分散行列である。これらを用いると  $P_{\text{DSA}}(\theta, f)$  は次のように表される<sup>8)</sup>。

$$\begin{aligned} P_{\text{DSA}}(\theta, f) &= \frac{1}{K} \sum_{k=1}^K |\mathbf{d}(\theta, f)^H \mathbf{x}(k, f)|^2 \\ &= \mathbf{d}(\theta, f)^H \mathbf{R}(f) \mathbf{d}(\theta, f) \quad (7) \end{aligned}$$

MV 法でのアレーの重み  $\mathbf{w}(f) = [w_1^*(f), \dots, w_M^*(f)]^H$  は、指向性のピークを向ける方向 (指向方向)  $\theta$  から到来する音波に対するアレーの出力  $y$  を一定に保つという拘束条件

$$\mathbf{d}(\theta, f)^H \mathbf{w}(f) = 1 \quad (8)$$

のもとで、アレーの出力のパワーを最小にするという規範に基づいて次式のように得られる。

$$\mathbf{w}(f) = \frac{\mathbf{R}(f)^{-1} \mathbf{d}(\theta, f)}{\mathbf{d}(\theta, f)^H \mathbf{R}(f)^{-1} \mathbf{d}(\theta, f)} \quad (9)$$

そして、この  $\mathbf{w}(f)$  を用いたときのアレーの出力パワー、すなわち、MV 法による指向方向  $\theta$  からの入射パワーの推定式は次式となる<sup>8)</sup>。

$$\begin{aligned} P_{\text{MV}}(\theta, f) &= \sum_{i=1}^K |\mathbf{w}(f)^H \mathbf{x}(k, f)|^2 \\ &= 1 / \mathbf{d}(\theta, f)^H \mathbf{R}(f)^{-1} \mathbf{d}(\theta, f) \quad (10) \end{aligned}$$

MV 法の指向性は指向方向に対する感度は 1 であるが、アレーの出力を最小にするために音源の方向に死角が形成される。従って、指向方向と音源の方向が十分に近づかないかぎり  $P_{\text{MV}}$  の値は小さく抑えられるので  $P_{\text{MV}}$  は  $P_{\text{DSA}}$  よりピークが鋭いと期待される。なお、 $\mathbf{R}(f)$  の固有値のばらつきが大きいと  $P_{\text{MV}}(\theta, f)$  に不要なピークが現れるので、本検討では極端に小さな固有値をなくすため  $\mathbf{R}(f)$  の対角要素を 1.1 倍した。

MUSIC 法の  $P_{\text{DSA}}$ ,  $P_{\text{MV}}$  に対応する関数は次式で与えられる<sup>8)</sup>。

$$P_{\text{MUSIC}}(\theta, f) = 1 / \mathbf{d}(\theta, f)^H \mathbf{R}(f)_n \mathbf{d}(\theta, f) \quad (11)$$

式 (10) と比べると分かるように、 $P_{\text{MV}}(\theta, f)$  での  $\mathbf{R}(f)^{-1}$  の代りに  $P_{\text{MUSIC}}(\theta, f)$  では  $\mathbf{R}(f)_n$  になっている。 $\mathbf{R}(f)_n$  は、 $\mathbf{R}(f)$  の小さい  $L$  個の固有値に対応する固有ベクトルを  $\mathbf{v}_1, \dots, \mathbf{v}_L$  と表したとき、

$$\mathbf{R}(f)_n = \sum_{i=1}^L \mathbf{v}_i \mathbf{v}_i^H \quad (12)$$

と定義される行列である。ただし、 $L$  = (マイクロホンの数) - (音の到来方向の個数) である。このように、

MUSIC では音の到来方向の個数が既知である必要がある。

雑音が音源と無相関でありマイクロホン間の雑音の相関行列が単位行列の定数倍であると仮定する。そして、 $N$  個の音の到来方向がそれぞれ  $\theta_i, i = 1, \dots, N$  であると仮定する。このとき、 $\mathbf{R}(f)$  の  $N$  個の大きな固有値に対応する固有空間は  $\mathbf{d}(\theta_i, f)$  が張る空間に等しい。固有ベクトルは互いに直交しているので、 $\mathbf{d}(\theta_i, f)$  は  $\mathbf{v}_1, \dots, \mathbf{v}_L$  に直交している。従って、式 (11) の分母は理想的には音の到来方向に一致する  $\mathbf{d}(\theta_i, f)$  で 0 となる。実際には 0 にはならないが  $P_{\text{MUSIC}}(\theta, f)$  は  $P_{\text{MV}}(\theta, f)$  より鋭いピークを持つと期待される。

さて、本論文における実験では音源の数は 1 個であるとし、信号としては広帯域信号である音声を用いた。広帯域音源の方向を推定するために、我々はまず、次のように中心周波数が  $f$  の狭帯域で入射パワー  $P(\theta, f)$  を求め、それを平均する方法をとった。

### 1. $R(f)$ の算出。

各マイクロホンで受信した信号を 300 Hz~3,400 Hz で帯域制限して 8 kHz サンプルングを行った。そして、128 サンプル長のハニング窓をかけた 128 点の 0 詰めを行って 256 点の FFT を行った。窓の位置は 1/4 ずつずらして FFT を繰り返すことにより各マイクロホンの受信信号に対応した周波数成分の時系列  $x_i(k, f), i = 1, 2, \dots, M$  を得た。次に各帯域でマイクロホン間の共分散行列  $R(f)$  を式 (5), (6) のように求めた。時間区間  $K$  は 3,000 点と長くし、音声の無音区間に影響されないようにした。

2. 各帯域について、式 (7), (10), (11) のいずれかを用いて、 $P(\theta, f)$  を求めた。 $\theta$  は  $1^\circ$  ごとに求めた。

3.  $P(\theta, f)$  は周波数について平均し、次式で与えられる  $Q(\theta)$  を求めた。

$$Q(\theta) = \int_{f_1}^{f_2} 10 \log(P(\theta, f)) df \quad (13)$$

$P(\theta, f)$  の対数を積分している理由は、音声の場合に低域に比べてパワーの真値が小さい高域成分も  $Q(\theta)$  の値に反映させて方向の推定に利用するためである。積分帯域は  $f_1 = 300$  Hz,  $f_2 = 3,400$  Hz とした。

4. 単一音源を仮定しているので  $Q(\theta)$  が最大値をとる  $\theta$  の値を音源方向の推定値とする。

室内音場においては、音源からの直接音以外に、壁、

床などからの反射音が、多数の方向から到来する。一方、MUSIC法では、音の到来方向の数が既知であり、その数はマイクロホンの数未満であることが動作条件とされている。従って、理論的には、室内音場においてMUSICを適用することはできない。一方、実用的な観点から、到来音波を直接音と少数の初期反射音及び残響音の和とモデル化して、残響音を雑音と見なすことで、直接音と初期反射音の到来方向を推定することはできる。しかし、MUSIC法で得られる  $Q(\theta)$  は到来音波のパワーを反映していないので、DSA法やMV法などのように、 $Q(\theta)$  が最大となる  $\theta$  が直接音の到来方向(音源方向)という論理を適用することはできない。そこで本論文では、すべての反射音を拡散雑音と見なして、音の到来方向は直接音のみの一方向と仮定してMUSIC法を適用した結果について検討を行った。

音源が  $-45^\circ$  の方向の無限遠にあり、反射のない自由空間において受信し、S/N=30 dB の場合の  $Q(\theta)$  の計算機シミュレーションの結果を Fig. 2 に示す。どの方法においても音源方向 ( $= -45^\circ$ ) は正しく推定されているが、DSA, MV, MUSIC の順でピークが鋭くなるのが分かる。

### 3. 音源方向推定実験

本章では音源が 1 個の場合に音源方向推定が良好に行える条件を明らかにするために行った実験結果について述べる。3.1 節において、反射音の量を定量的に表すための指標である D/R について説明し、3.2 節において、実験の諸条件を示す。3.3 節において結果を述べ、3.4 節において考察を行う。3.5 節では、人の場合の D/R の推定式を示し、3.6 節では方向推定が良好に行える条件の一例を示す。

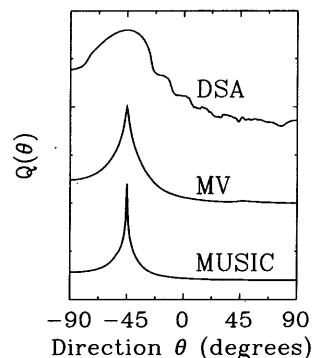


Fig. 2 Examples of analysis when there are no reflections. The sound source is located infinitely far in the  $-45^\circ$  direction. The number of microphones is 8. S/N is 30dB.

### 3.1 反射音の影響の評価パラメータ

室内の壁面などによる反射音の多くは、実際に音源が存在する方向とは大きく異なった方向から到来する。従って、これらの反射音は音源方向推定の妨害要因となり、反射音の量が増加すると音源方向推定誤差が増加することが予想される。そこで本論文では、音源方向の情報を持つ直接音と、妨害要因である反射音(間接音)の比(D/R)をパラメータとして方向指定誤差の評価実験を行った。

本論文での個々のD/Rの測定方法は、まず、音源とマイクロホンとの間のインパルス応答を計測した。次に、そのインパルス応答の波形から視察によって直接音と間接音の境界を決定して、直接音のエネルギーと、間接音のエネルギーの値を求め、それらの比として各マイクロホンごとのD/Rを算出した。そして、マイクロホンアレーにおけるD/Rとしてはアレーを構成する各マイクロホンでのD/Rの対数値の平均値を用いた。ただし、音源としては8 cm口径のスピーカを用い、マイクロホンは無指向性のマイクロホンを用いた。インパルス応答の計測には、M系列相関法を用いた。

D/Rを変化させるために、音源とアレー中心との距離、スピーカの向き、及び部屋の残響時間を変化させた。直接音のエネルギーは音源から遠ざかるにつれて小さくなるので、音源-アレー間距離に変化させれば直接音のエネルギーが変化する。また、スピーカは指向特性を持っており、スピーカの正面がアレーの方向からずれるに従って、直接音のエネルギーは小さくなることから、スピーカの向きを変化させると直接音のエネルギーが変化する。一方、剛壁面の実験室において、壁面に0.8 m×1.5 mのグラスウールの吸音板をランダムに付加することで、残響時間を制御したが、このことは反射音のエネルギーを変化させることになる。このようにして求めたD/Rの値と方向推定誤差の関係を求めた。

実験においては、音源とアレー間距離は1~4 mと変化させ、スピーカの向きはアレーに対して正面、側面、背面向きと変化し、また、残響時間は300~800 msと変化してD/Rの値を約10~-30 dBの範囲で変化させた。そして、各D/Rの値と音源方向推定誤差との関係を調べた。

### 3.2 推定誤差要因と実験項目

音源方向の推定性能はD/Rの値以外にも様々な要因に影響されることが予想される。例えば、音源とアレーの配置、受信装置(アレーの規模、形状等)、推定手法(DSA, MV, MUSIC等)等々が考えられる。

本論文ではそのうち、次の条件については固定とした。

- アレーの形状は等間隔直線アレー(4 cm 間隔)とした。理由はマイクロホンの個数を一定にした場合には直線アレーが最も長く従ってビーム幅を鋭くできることと、信号方向推定の理論的解析の多くがこの形状を仮定していることである。
- 部屋の形状は直方体とした。理由は実際の多くの部屋が直方体であることと、計算機シミュレーションが簡単なことである。
- 室内でのアレーの配置はアレーの軸が最寄りの側壁及び床に並行になるように配置した。例えば、6.7 (D)×4.3 (W)×3.0 (H) mのサイズの部屋ではアレーをFig. 3に示すように配置した。

そして、以下の項目を変化させて実験を行った。

実験1 推定の手法、音源の方向

実験2 マイクロホン数

また、D/Rを変化させる要因である以下の項目について実験を行った。

実験3 部屋の容積

実験4 音源の指向性

四つの実験の条件をTable 1にまとめておく。

### 3.3 実験結果

#### 実験1 音源の方向、推定手法と推定誤差

ここではD/Rと推定誤差の関係が音源方向及び推定手法にどのように依存しているかを調べる。音源の方向による差について実験するのは直線アレーの指向性のビーム幅が方向 $\theta_s$ により変化し、それに伴って方向推定の分解能も精度も変化すると考えられるためである。音源の方向を $\theta_s = 0^\circ, 30^\circ, 60^\circ, 90^\circ$ と変化した場合のD/Rと推定誤差の関係をDSA, MV, MUSICの各手法について調べその結果をFig. 4(a)~(c)に示した。これらの図から、D/Rと誤差の関係は推定手法にはあまり依存しないことが分かる。また、推定誤差は音源がアレーの正面方向( $\theta_s = 0^\circ$ )では小さくなるが、 $\theta_s = 0^\circ$ の場合を除けば、D/Rが-5

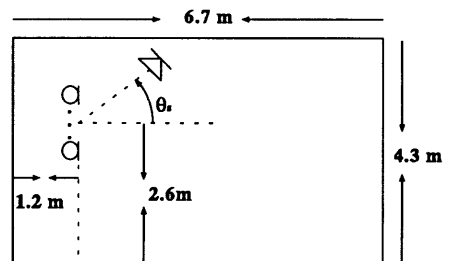
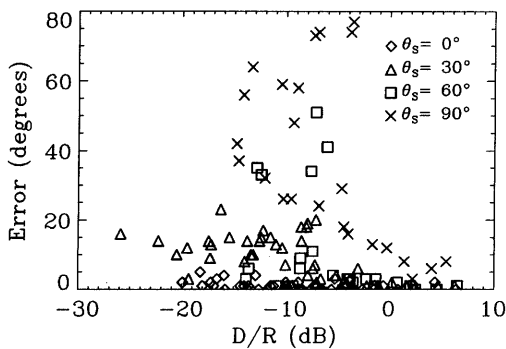


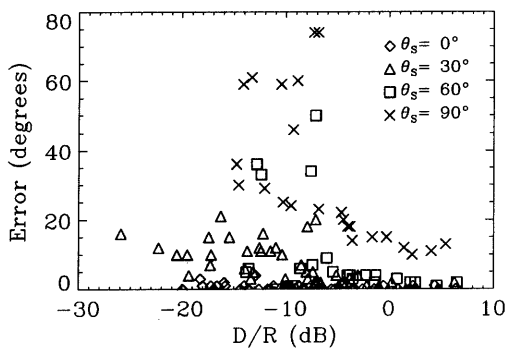
Fig. 3 Example of arrangement of microphone array and sound source. The height of the room is 3 m.

Table 1 Experimental conditions.

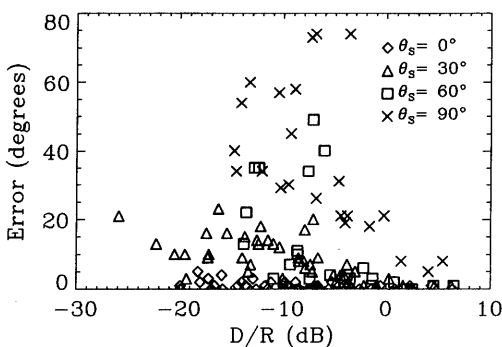
	音源方向	推定法	マイクロホン数	部屋の容積 (m <sup>3</sup> )	音源	音源の向き (指向性)
実験 1	0°, 30°	DSA, MV	8	86	スピーカ	正面, 後ろ, 上
	60°, 90°	MUSIC				
実験 2	30°	DSA	4, 8, 16	86	スピーカ	正面, 後ろ, 上
実験 3	30°	DSA	8	44, 86, 640	点音源	無指向性
実験 4	30°	DSA	8	86	スピーカ	正面, 後ろ, 上



(a) DSA



(b) MV



(c) MUSIC

Fig. 4 Estimation error vs. D/R for three estimation methods and various directions of sound sources.

～10 dB 以下では誤差が増加することが分かる。

以下の実験では、この D/R の減少に伴って推定誤差が増加する現象に着目する。推定手法は DSA 法、音源の方向は 30° として以下の実験を行った。

#### 実験 2 マイクロホンの数による影響

この実験ではアレーの長さによる D/R と推定誤差の関係性を調べる。この実験を行う理由はマイクロホン間隔を一定 (4 cm) として、マイクロホン数を変化させるとアレー長が変化しその結果、アレーの指向性のビーム幅が変化するので、方向推定の分解能、精度が変化するためである。Fig. 5 (a)～(c) にマイクロホンの数  $M$  をそれぞれ 4, 8, 16 としたときの実験結果を示す。

これらの図からマイクロホンの数によらず、実験 1 と同様の傾向で、D/R が小さくなると誤差が増加することが分かる。D/R が -5 dB 以上での誤差はマイクロホンの数により異なるが、約 -5 ～ -10 dB 以下ではマイクロホン数によらず誤差が大きく増加することが分かる。

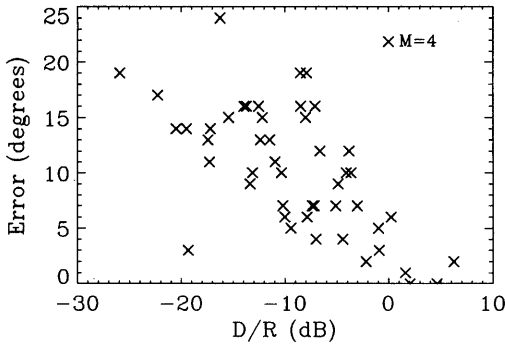
#### 実験 3 部屋の容積の影響

部屋のサイズの影響を見るため、実験 1, 2 で使用した部屋 (medium room,  $6.7 \times 4.3 \times 3.0 = 86 \text{ m}^3$ ) とその約 8 倍 (large room,  $8.8 \times 11.0 \times 6.6 = 640 \text{ m}^3$ )、及び、1/2 (small room,  $3.4 \times 5.4 \times 2.4 = 44 \text{ m}^3$ ) の容積の部屋におけるインパルス応答を鏡像法によりシミュレーション<sup>9)</sup>して、方向推定誤差を比較した。壁での吸音率はすべての壁で等しく、容積  $86 \text{ m}^3$  の部屋で 0.08～0.35 (残響時間にして 280 ms～800 ms)、容積  $640 \text{ m}^3$  の部屋で 0.07～0.35 (残響時間にして 310 ms～1,570 ms)、容積  $44 \text{ m}^3$  の部屋で 0.05～0.2 (残響時間にして 230～1,020 ms) とした。

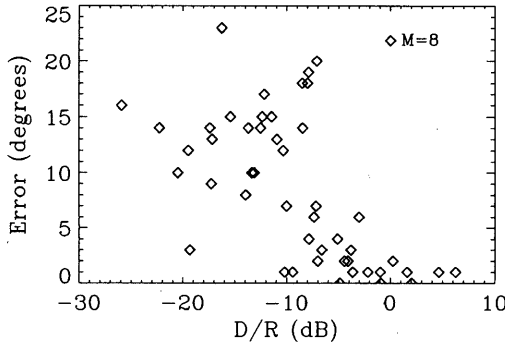
実験結果を Fig. 6 に示した。Fig. 6 から、部屋の容積に関係なく D/R が -5 ～ -10 dB 以下になると推定誤差が急激に増加することが分かる。

#### 実験 4 音源の指向性による影響

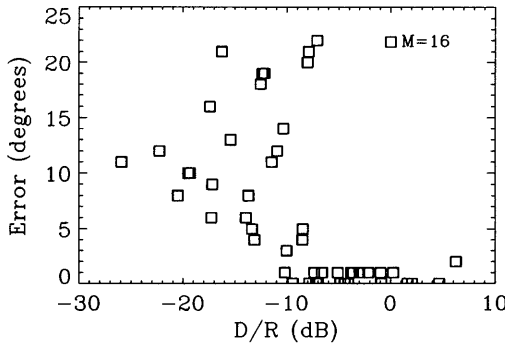
音源の指向性は実験 1, 2 では区別していなかったが、ここでは音源の指向性が D/R と推定誤差の関係に



(a) Four microphones



(b) Eight microphones



(c) Sixteen microphones

Fig. 5 Estimation error vs. D/R for various numbers of microphones.

与える影響を検討した。スピーカを上向きにした場合、アレーの正面に向けた場合、アレーに後向きにした場合の D/R と誤差の関係を Fig. 7 に示した。この図より、スピーカをどの向きにしても、D/R を小さくしていくと、誤差が急激に大きくなる D/R は -5 ~ -10 dB の間である。一方、実験 3 のシミュレーションでは無指向性音源を仮定しているが同様の結果が得られている。従って、D/R と推定誤差のこの関係は、音源の指向性には大きく影響されないと結論される。

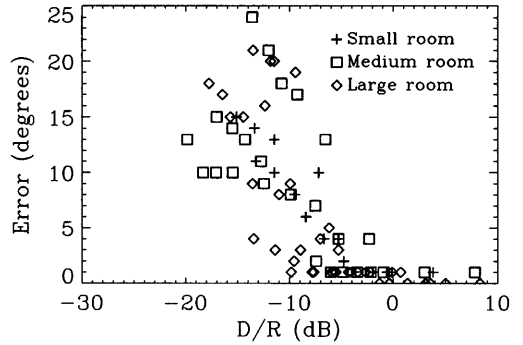


Fig. 6 Estimation error vs. D/R for various room in volumes.

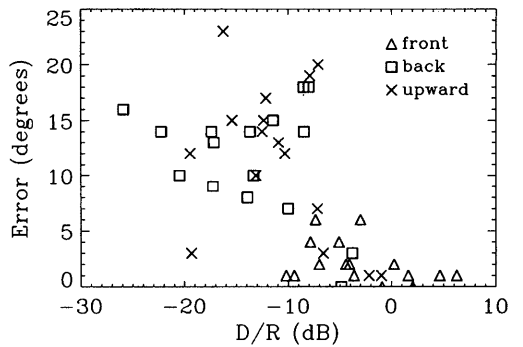
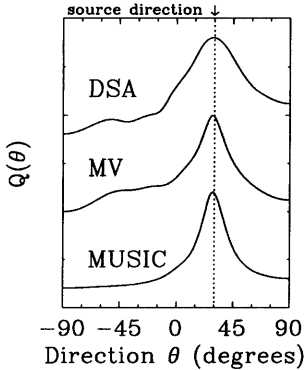


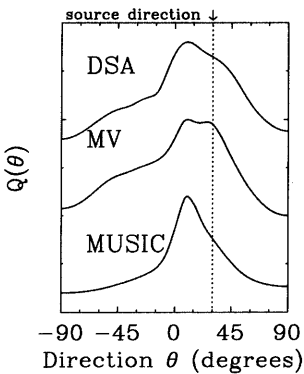
Fig. 7 Estimation error vs. D/R for various loud-speaker directions.

### 3.4 実験結果の考察

推定誤差が小さい場合と大きい場合について、各手法で得られた  $Q(\theta)$  を Fig. 8 (a), (b) に示した。音源の方向は  $30^\circ$  で、実験条件は室容積が  $86 \text{ m}^3$ 、音源とアレーの距離が  $1 \text{ m}$ 、マイクロホンの数が  $8$  個、スピーカは天井を向いているというのが共通した条件であり、誤差が小さい Fig. 8 (a) の場合には残響時間が  $500 \text{ ms}$ 、 $D/R = -2.2 \text{ dB}$  で、誤差が大きい Fig. 8 (b) の場合には残響時間が  $800 \text{ ms}$ 、 $D/R = -7.1 \text{ dB}$  であった。Fig. 8 (a) より、残響の少ないときには DSA, MV, MUSIC 法はピーク幅に差はあるが、どの方法においても音源方向を正しく推定できている。Fig. 8 (b) の実験条件は、Fig. 8 (a) の条件に比べて残響時間を  $500 \text{ ms}$  から  $800 \text{ ms}$  とし、全体の反射音の量が増加するようにしたものである。この条件では、音源の方向 ( $30^\circ$ ) に加えて、約  $10^\circ$  の方向からの到来音を検出されていることが分かる。DSA 法に比べて MV 法はビーム幅が鋭いため、これらの方向は、より明確な二つのピークとして検出されている。そして、約  $10^\circ$  方向に見られたピークの値が  $30^\circ$  方向のピーク値より大



(a) Reverberation time is 500 ms. (D/R=-2.2dB)



(b) Reverberation time is 800 ms. (D/R=-7.1dB)

Fig. 8  $Q(\theta)$  for three analysis methods. Room volume is  $86 \text{ m}^3$ . Distance between the sound source and the center of the array is 1 m.

きな値となり、音源方向を約  $10^\circ$  の方向と誤推定して、大きな推定誤差が発生することが分かる。一方、MUSIC 法では、音の到来方向は一方向と仮定されているので、一つのピークのみが推定されている。この約  $10^\circ$  方向のピークは、その方向から、パワーの大きな反射音（単一又は複数の）が到来していることが原因と予想される。

各種推定法の大きな違いは、そのビーム幅 ( $Q(\theta)$ ) のピークの鋭さの違いにある。しかし、ある特定方向からの反射音が増加することが原因で発生する推定誤差は、Fig. 8 (b) から分かるように、ビーム幅の鋭さには大きく依存しない。このことから、大きな推定誤差を発生する主要因は DSA 法、MV 法、MUSIC 法などの推定法の差ではなく、反射音の量を表す D/R にあるものと考えられる。また、同様の理由から、実験 2 のようにマイクロホンの数を増加させて鋭いビームを形成することも、推定誤差が大きく変化する要因とは

ならないと考えられる。

本論文では、アレーの形状を直線アレーとし、直方体の部屋の壁と床に並行に配置されているものと仮定した。このような状況は最も自然なものであると考えられるが、音源方向の推定という観点からは幾つか問題点がある。一つは、直線アレーの指向性は軸対象であって仰角の区別ができないので、平面や立体的にマイクロホン配置した仰角の区別がある指向性を持つアレーより反射音を多く受音してしまうという問題である。もう一つは、直方体の部屋で直線アレーを壁、床に並行に配置したことで、対抗壁、床と天井の間に立つ定在波がどちらもアレーの正面方向にある音源に見えるという問題である。従って、平面アレーを用いたり直線アレーを非並行に配置したりした場合には、今回の結果よりも小さな D/R であっても方向推定が行える可能性がある。言い換えると、今回の結果は直線アレーを用いた場合の最悪条件に対する結果を示したものと考えられるが、これらの事柄に関する実証、検討は今後の重要な課題である。

### 3.5 音源が人の場合の D/R の求め方

以上の実験結果では方向推定を良好に行うための条件が D/R により与えられた。従って、推定が良好に行えるかどうかを判定するには D/R を求める必要がある。音源がスピーカの場合にはインパルス応答から D/R を計算することができた。しかし、音源が人の場合には、インパルス応答の測定が困難なので、他の方法で求める必要がある。ここでは、インパルス応答以外の測定可能な物理量から D/R を推定する方法を示す。

#### 3.5.1 D/R の予測式

拡散音場を仮定すれば、直接音のエネルギー密度  $E_d$  と反射音（拡散音）のエネルギー密度  $E_r$  が理論的に与えられる<sup>10)</sup>。

$$E_d = \frac{QW}{4\pi r^2 c} \quad (14)$$

$$E_r = \frac{4W}{cS\bar{\alpha}}(1-\bar{\alpha}) \quad (15)$$

ここで、 $W$  は音源の出力、 $r$  は音源と受音点の距離 (m)、 $c$  は音速 (m/s)、 $S$  は部屋の全表面積 ( $\text{m}^2$ )、 $\bar{\alpha}$  は壁の平均吸音率、 $Q$  は受音点に対する音源の指向性係数を表す。 $Q$  は無指向性音源で 1 であり、無指向性音源の場合を基準とした直接音のエネルギー密度として、次式で与えられる。ただし、対象とする音源の出力、受音点との距離は無指向性音源と同じである。

$$Q = \frac{\text{直接音のエネルギー密度}}{W/(4\pi r^2 c)} \quad (16)$$

式 (14), (15) から D/R は

$$\begin{aligned} D/R &= E_d/E_r \\ &= \frac{QS\bar{\alpha}}{16\pi r^2(1-\bar{\alpha})} \end{aligned} \quad (17)$$

となる。一方, Sabine の残響式により, 残響時間  $T_{60}$  (s) が次式で与えられる<sup>10)</sup>。

$$T_{60} = \frac{55.3}{c} \frac{V}{S\bar{\alpha}} \quad (18)$$

ただし,  $V$  は部屋の容積 ( $\text{m}^3$ ) である。この式 (18) を (17) に代入し,  $S\bar{\alpha}$  を消去し,  $c=340$  (m/s) とすると, 次式となる。

$$\begin{aligned} D/R &= \frac{1}{16\pi} \frac{55.3}{c} \frac{QV}{r^2 T_{60}(1-\bar{\alpha})} \\ &= 10 \log_{10}(V) - 10 \log_{10}(r^2 T_{60}) \\ &\quad - 10 \log_{10}(1-\bar{\alpha}) - 25 + F \text{ (dB)} \\ F &= 10 \log_{10}(Q) \end{aligned} \quad (19)$$

実験では残響時間は約 300~1,000 ms であり, これは  $1-\bar{\alpha}=0.65\sim 0.95$  に対応する。従って, 実験の範囲では,  $|10 \log_{10}(1-\bar{\alpha})|$  の最大値は  $|10 \log_{10}(0.65)| = 1.9$  dB である。この値は他の項に比べて小さいので無視し, 次式を D/R の予測式とする。

$$\begin{aligned} D/R &= 10 \log_{10}(V) - 10 \log_{10}(r^2 T_{60}) \\ &\quad - 25 + F \text{ (dB)} \end{aligned} \quad (20)$$

次に複数の実験データに対して, この予測式をあてはめ検証した。実験データは 3.3 節で示した実験 1, 実験 3 で用いたデータを利用した。各実験における部屋の容積  $V$ , 距離  $r$ , 残響時間  $T_{60}$ , 及び指向性係数  $F$  を式 (20) に代入して D/R の予測値を求め, 対応する実験条件で計測 (計算) されたインパルス応答から求めた D/R と比較した。なお, 実験 1 で用いたデータ数は 144, 実験 3 で用いたデータ数は 150 で,  $V$ ,  $r$ ,  $T_{60}$  の値は 3.1, 3.3 節で述べたように変化させた。また, 指向性係数  $F$  の値は, 実験 3 の計算機シミュレーションでは無指向性音源を仮定しているため 0 dB とした。実験 1 のスピーカに対しては, まず式 (20) を  $F$  に関して解いたものに実験データの D/R,  $V$ ,  $r$ ,  $T_{60}$  を代入し得られた値の平均値として用いた。

このように得られた D/R の推定値とインパルス応答から求めた D/R との誤差の標準偏差は実験 1 に対しては  $(1.38 \text{ dB})^2$ , 実験 3 のデータに対しては  $(1.53 \text{ dB})^2$  であった。この結果から 1~2 dB の誤差範囲で,

この予測式が有効であることが確認された。

### 3.5.2 音源が人の場合の予測式

音源が人である場合に式 (20) を適用した結果について述べる。まず, 人の指向性係数を知るために, 人が音声を発生する場合の指向特性を測定した。測定は, 水平面内で人体から 1 m の距離に,  $0^\circ$  (正面) から,  $180^\circ$  まで  $30^\circ$  間隔でマイクロホンを設置して音声を録音した。測定した音声は周波数分割した後, 各角度でのパワーの値を正面方向のパワーで正規化して周波数ごとの指向特性を求め, その後, 周波数平均をとった。指向特性の測定結果の例を Table 2 に示した。次に人の指向特性が頭の前後を通る軸に関して軸対称であると仮定して, この指向特性の値から, 正面方向の指向性係数  $F(0^\circ)$  を次式により計算した。

$$F(0^\circ) = 10 \log_{10} \left( 4\pi \sum_{i=0}^6 E(\theta_i) \right), \quad (21)$$

$$\begin{aligned} \theta_i &= 30^\circ \times i \\ \text{for } i &= 0, 6 \end{aligned}$$

$$E(\theta_i) = \Omega(15^\circ) D(\theta_i) \quad (22)$$

$$\text{for } i = 1, 2, \dots, 5$$

$$E(\theta_i) = (\Omega(\theta_i + 15^\circ) - \Omega(\theta_i - 15^\circ)) D(\theta_i)$$

ただし,  $\Omega(\theta)$  は立体角であり次式で与えられる。

$$\Omega(\theta) = 2\pi(1 - \cos\theta) \quad (23)$$

計算結果は, 約 3 dB となった。人がマイクロホンに対して横方向 ( $\theta_i = 90^\circ$ ) や後ろを向いている場合 ( $\theta_i = 180^\circ$ ) の指向性係数  $F(\theta_i)$  は, この  $F(0^\circ)$  の値に Table 2 の  $D(\theta_i)$  の値を加算して,  $F(\theta_i) = 3 + D(\theta_i)$  dB となる。これより人の D/R の予測式は

$$\begin{aligned} D/R \text{ (dB)} &= 10 \log_{10}(V) - 10 \log_{10}(r^2 T_{60}) \\ &\quad - 22 + D(\theta_i) \end{aligned} \quad (24)$$

式 (24) を用いて予測した D/R と人が発声した音声の方向推定誤差をまとめると Fig. 9 となる。実験条件は室容積  $86 \text{ m}^3$ , 方向  $30^\circ$ , 残響時間は 400, 700 ms である。また, マイクロホンは 8 個, 分析は DSA 法である。Fig. 9 においても, D/R が  $-5 \sim -10$  dB 以下になると推定誤差が大きくなる傾向が確認されるので, このことに関しては式 (24) の D/R の予測式が有効であることが分かる。

### 3.6 方向推定可能な範囲の一例

実験の結果から, 音源がアレーの正面にある場合を

Table 2 Example of a man's directivity (dB).

Angle	$0^\circ$	$30^\circ$	$60^\circ$	$90^\circ$	$120^\circ$	$150^\circ$	$180^\circ$	$F$ (dB)
$D(\theta)$	0	-0.39	-1.32	-2.78	-5.1	-6.3	-5.1	2.7



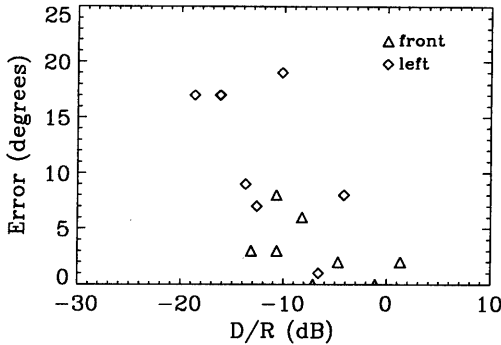


Fig. 9 Estimation error vs. D/R when sound source is a human.

除けば、音源の方向、推定手法、アレーの規模、部屋の容積、音源の指向性によらず、D/R が約  $-5$  dB 以下にすると推定誤差が急激に増加することが分かった。従って、この D/R が音源推定方向が可能な限界になる。また、3.5 節より音源が人の場合においても向きに応じて D/R を予測できる。例えば、部屋の容積が  $86 \text{ m}^3$ 、残響時間が  $500 \text{ ms}$  ならば、人が横を向いた状態で  $D/R = -5 \text{ dB}$  となる人とアレーとの距離は  $1.2 \text{ m}$  である。この距離以内に人がいる場合には発声者が横を向いていてもその方向を数度の誤差範囲で推定することが可能である。

#### 4. まとめ

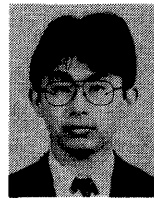
直方体の室内で、側壁と床、天井に並行に配置した直線アレーを用いて、単一音源の方向を推定する際の限界を実験的に検討した。室内反射音の多少を表す指標として直接音と間接音のパワーの比である D/R を採用した。音源の方向、方向推定手法、マイクロホン数、部屋の容積、音源の指向性を変え実験を行った。その結果、ほぼ実験条件に関わらず D/R の値が  $-10 \sim -5 \text{ dB}$  以下になると推定誤差が急激に大きくなり推定が良好に行えなくなることが分かった。また、音源が人間である場合などのように、インパルス応答を測定することができず、直接 D/R を計算できない場合のために、D/R をアレーと音源との距離、残響時間、部屋の容積と音源の指向性から推定する式を示した。

#### 文 献

1) J.L. Flanagan, J.D. Johnston, R. Zahn and J.W. Elko, "Computer-steered microphone arrays for sound

transduction in large rooms," J. Acoust. Soc. Am. **78**, 1508-1518 (1985).

- 2) 永田仁史, 安部正人, 城戸健一, "多数センサによる音源波形の推定," 音響学会誌 **47**, 268-273 (1991).
- 3) H.F. Silverman, "An algorithm for determining talker location using a linear microphone array and optimal hyperbolic fit," Proc. Speech and Natural Language Workshop, 151-156 June (1990).
- 4) I. Yamada and N. Hayashi, "Improvement of the performance of cross correlation method for identifying aircraft noise with pre-whitening of signals," J. Acoust. Soc. Jpn. (E) **13**, 241-252 (1992).
- 5) Y. Yamasaki and T. Itow, "Measurement of spatial information in sound fields by closely located four point microphone method," J. Acoust. Soc. Jpn. (E) **10**, 101-110 (1989).
- 6) Y. Kaneda, "Sound source localization for wide-band signals under a reverberant condition," J. Acoust. Soc. Jpn. (E) **12**, 47-48 (1993).
- 7) 福留公利, "球バップル埋め込みマイクロホンとその回折情報を利用した音源の方向及びスペクトル推定," 音響学会誌 **44**, 272-281 (1988).
- 8) S.U. Pillai, *Array Signal Processing* (Springer-Verlag, New York, 1989), pp. 7, 8, 28-33.
- 9) J.B. Allen and D.A. Berkly, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943-950 (1979).
- 10) 前川純一, 建築・環境音響学 (共立出版, 東京, 1990), pp. 51-56.



田中 雅史

昭和63年北大・工卒。平成2年同大学院修士課程了。同年NTT入社。以来、マイクロホンアレー、音響エコーキャンセラにおける信号処理の研究に従事。電子情報通信学会会員。



金田 豊

昭和50名大・工・電気卒。昭和52同大学院修士課程了。同年日本電信電話公社(現NTT)入社。以来、NTT研究所において、マイクロホンアレー信号処理、音場制御、音響エコーキャンセラ、音源方向検出などの音響信号処理の研究に従事。現在、NTTヒューマンインタフェース研究所音声情報研究部主幹研究員。工博。日本音響学会、米国音響学会、電子情報通信学会、IEEE各会員。



小島 順治

昭和48年早大・理工・電子通信卒。同年日本電信電話公社(現NTT)入社。以来、NTT研究所において拡声電話機、音声入出力装置、ISDN端末装置、及び音響信号処理の研究に従事。現在NTTヒューマンインタフェース研究所音声情報研究部主幹研究員。日本音響学会、電子情報通信学会、情報処理学会各会員。