

# マイクロホンアレーを用いた CSP 法に基づく複数音源位置推定

西浦 敬信<sup>†</sup>

山田 武志<sup>†</sup>

中村 哲<sup>†</sup>

鹿野 清宏<sup>†</sup>

Localization of Multiple Sound Sources Based on CSP Analysis  
with a Microphone Array

Takanobu NISHIURA<sup>†</sup>, Takeshi YAMADA<sup>†</sup>, Satoshi NAKAMURA<sup>†</sup>,  
and Kiyohiro SHIKANO<sup>†</sup>

あらまし 本論文では、マイクロホンアレーを用いて複数の音声を高品質に収録するために、複数の音源位置を推定する方法について検討する。これまでに、白色化相互相関法（CSP 法：Cross-power Spectrum Phase Analysis）に基づく単一音源位置推定法が提案されている。この方法では、複数のマイクロホンのペアにより音源方向を推定し、その推定方向の交点から音源位置を推定する。しかし、この方法を用いて複数の音源位置推定を行う場合、異なる音源同士の相関によるクロス項の影響により音源方向推定精度が劣化するという問題がある。更に正確に音源方向を推定した場合でも、推定方向の交点から複数の音源位置を推定すると、異なる音源推定方向の交点の影響により、音源位置推定精度が劣化する問題がある。本論文では、この問題に対処するために、複数の CSP 係数を加算してクロス項の影響を抑える方法と、ビームフォーミングによる推定方向のクラスタリング法を提案する。提案法を評価するために、音響実験室において収録したマイクロホンアレーの信号を用いて位置推定実験を行った。その結果、クロス項の発生する 2 音源の場合で約 10%、3 音源の場合で約 25% 位置推定率が向上することが明らかとなった。

キーワード ハンズフリー、マイクロホンアレー、複数音源位置推定、CSP 係数

## 1. ま え が き

現在のテレビ会議システムや音声入力による機器制御システムでは、音声を収録するために、卓上型マイクロホンや接話マイクロホンなどが用いられている。これらのマイクロホンを用いることにより信号対雑音比（Signal to Noise Ratio：SNR）の高い音声を安定して受音できるが、マイクロホンを常に意識する必要があるためユーザにとっては非常に煩わしい。よって、自然で使い勝手の良い音声インタフェースを実現するためには、ユーザにマイクロホンを意識させずに、任意の位置での発話を許容する必要がある。しかし実環境において、マイクロホンから離れて発話された音声を受音する場合、背景雑音や部屋の残響により目的音声は著しくひずみ、音質が劣化してしまう。本研究の最終的な目的は、背景雑音や部屋の残響が存在する一般的な会議室において、任意の離れた場所で発話され

た音声を高品質に収録することである。以下、本論文の中では、話者から離れた位置にあるマイクロホンにより音声を高品質に受音することをハンズフリー受音と呼ぶ。

ハンズフリー受音を実現するためには雑音を抑圧することが重要となる。これまで雑音の問題に対して、スペクトルサブストラクション [1]、ウィナーフィルタリング [2]、など様々な手法が提案されているが、これらの手法は、雑音の性質や定常性に依存している。しかし、実環境に存在する雑音は多種多様であるので、限られた雑音環境下でしか効果が得られない。また残響を考慮しておらず、残響の影響に対処することができない。そこで近年、多チャネルの素子を利用して残響及び雑音抑圧を行うマイクロホンアレーを用いた方法が注目されている。

マイクロホンアレーとは、複数のマイクロホンを空間的に配置したものである。各マイクロホンの受信信号間には、マイクロホンと音源位置関係に応じた位相差や振幅差等が生じる。マイクロホンアレー信号処理による残響・雑音抑圧法では、この位相差の情報を利

<sup>†</sup> 奈良先端科学技術大学院大学情報科学研究科，生駒市  
Graduate School of Information Science, Nara Institute of  
Science and Technology, 8916-5 Takayama, Ikoma-shi, 630-  
0101 Japan

用している。また、先に挙げた手法と違って雑音の性質や定常性に依存しないので、様々な環境で効果的に残響・雑音抑圧を実行できる。したがって本研究では、ハンズフリー受音系を実現するためにマイクロホンアレーを用いる。

マイクロホンアレーを使用して高品質に音を受音する場合、音源位置に対して指向性を形成するために、音源位置を推定する必要がある。特に残響のある部屋や雑音源などが存在する環境下では、残響や雑音の影響を抑圧して目的音を高品質に受音する必要があるため、正確な音源の位置推定が必要となる。

音源位置の推定方法としては、ビームフォーミングによる音源方向推定法 [3], MUSIC (Multiple Signal Classification) 法による音源方向推定法 [4], 最小分散法 (MV 法: Minimum Variance method) による音源方向推定法 [5], [6], 同期乗算による音源位置推定法 [7], 及び白色化相互相関法 (CSP 法: Cross-Power Spectrum phase analysis) による音源方向推定法 [8]~[12] など様々な方法が提案されている。これらの方法を用いて、単一若しくは複数の音源位置を推定する研究が既に行われている。CSP 法は 2ch のマイクロホンで音源方向を推定できることから、他の手法よりも計算量が少なく大変有利な手法である。しかし、現在のところ複数の音源の位置推定には適用されていない。そこで本論文では、複数の音源が存在する環境において、CSP 法を用いて複数の音源位置を推定する手法を提案する。

実環境において複数音源の位置推定を行うにあたり、以下の問題点が挙げられる。

- (1) (相関のある) 複数音源
- (2) 残響

残響は同じ音が複数存在する (1) の特殊な場合としてとらえることができる。したがって、本論文では、(1) についての問題をまず解決する手法を確立することを目指す。その次のステップとして (2) の問題に適用することとする。よって本論文では残響の影響が小さい環境下を対象として、相関のある複数音源の位置推定を実現する方法を提案する。

複数の音源位置を正確に推定することができれば、任意の部屋・位置で発話された音声に対して、雑音や残響の影響を抑圧して高音質に音声を収録するハンズフリー受音を実現できることから、テレビ会議システムなどに応用可能である。また、音の異常と場所を検出できることから、セキュリティシステムなどにも応

用可能である。

しかし、CSP 法を用いて複数音源の位置推定を行う場合、CSP 法は音の相関をもとにして音源位置を推定するため、異なる音源同士の相関により正確に位置を推定できないという問題がある。そこで、複数の CSP 係数を加算して異なる音源同士の相関の影響を削減する方法を提案する。

提案法では、複数の音源が存在するときの CSP 法による方向推定精度の劣化を防ぐため、CSP 係数の加算により複数の音源方向を推定する。そして、異なる音源を示す推定方向同士の交点の発生を防ぐため、ビームフォーミングにより音を抽出して各音波の相関値をもとに推定方向をクラスタリングすることにより、精度良く複数の音源位置を推定する。本論文では提案法の有効性を、音響実験室における実験により評価する。

本論文の構成は、まず 2. で従来の音源位置推定法とその問題点について説明する。3. では提案システムの構成について説明する。4. では提案法の性能評価実験について述べ、最後に 5. において本論文の結論と今後の課題について述べる。

## 2. 従来の音源位置推定法とその問題点

マイクロホンアレーを用いて音源位置を推定する基本的なシステムは、音の受音、音源方向推定、音源位置推定の三つの要素から構成される。以下従来法の各処理について詳細に説明する。

### 2.1 CSP 法による音源方向推定

図 1 に示すようにマイクロホン  $i, j$  で信号  $s_i(n)$ ,  $s_j(n)$  を受音すると、CSP 法により 2ch の受音信号の CSP 係数及び到来時間差 (DOA: Delay Of Arrivals) を推定できる。CSP 係数及び到来時間差の推定式を

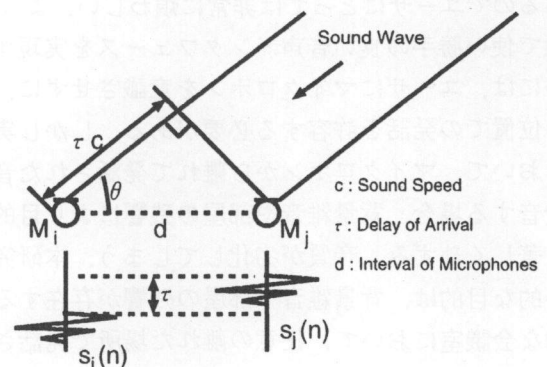


図 1 CSP 法による到来時間差の推定  
Fig.1 Estimation of DOA with CSP method.

式 (1), (2) に示す。

式 (1) では, 2ch の受信信号をフーリエ変換して, 振幅で正規化を行う。そして位相差を求めて逆フーリエ変換を行い CSP 係数を求めている。更に式 (2) により CSP 係数の大きくなる時間差 (相関の強い時間差) を求めることにより, 到来時間差の推定を行う。

$$\text{CSP}_{i,j}(k) = \text{DFT}^{-1} \left[ \frac{\text{DFT}[s_i(n)]\text{DFT}[s_j(n)]^*}{|\text{DFT}[s_i(n)]||\text{DFT}[s_j(n)]|} \right] \quad (1)$$

$$\tau = \underset{k}{\operatorname{argmax}}(\text{CSP}_{ij}(k)) \quad (2)$$

音源が一つしか存在しない環境において音源方向を推定するときは, 式 (1), (2) により  $\tau$  を求めて式 (3) により音源方向を推定すればよい。なお式中の  $c$  は音速,  $d$  はマイクロホン間隔,  $F_s$  はサンプリング周波数を示す。

$$\theta = \cos^{-1} \left( \frac{c \cdot \tau / F_s}{d} \right) \quad (3)$$

しかし, 音源が複数存在する環境において複数の音源位置を推定するとき, 真の音源方向以外の CSP 係数が大きくなり, 誤った音源方向を推定することがある。例えば二つの音源からの信号  $x(t)$ ,  $y(t)$  が到来する場合, 受信信号  $s_i(n)$ ,  $s_j(n)$  は式 (4), (5) のように表せる。

$$s_i(n) = a_i x(m + \varphi_i) + b_i y(m + \xi_i) \quad (4)$$

$$s_j(n) = a_j x(m + \varphi_j) + b_j y(m + \xi_j) \quad (5)$$

ここで信号  $x(t)$ ,  $y(t)$  に対して, 対応する  $m$  は時刻,  $\varphi$ ,  $\xi$  は到来時間,  $a$ ,  $b$  は振幅,  $X(w)$ ,  $Y(w)$  は信号  $x(t)$ ,  $y(t)$  の周波数表現を表したものである。式 (4), (5) より式 (1) の  $\text{DFT}^{-1}[\cdot]$  の分子は式 (6) と表せる。

$$\begin{aligned} \text{DFT}(s_i(n))\text{DFT}(s_j(n))^* &= a_i a_j X(w)^2 e^{-jw(\varphi_i - \varphi_j)} \\ &+ b_i b_j Y(w)^2 e^{-jw(\xi_i - \xi_j)} \\ &+ X(w)Y(w) (a_i b_j e^{-jw(\varphi_i - \xi_j)} \\ &+ a_j b_i e^{-jw(\xi_i - \varphi_j)}) \end{aligned} \quad (6)$$

式 (6) には  $X(w)^2$ ,  $Y(w)^2$  の項以外にも  $X(w)$  と  $Y(w)$  がかわるクロス項がある。ここで  $X(w)$  と  $Y(w)$  が完全に無相関であれば問題はない。しかし,

実際には実環境においてどのような信号を受音するのかわからないことから, このクロス項を無視できない場合がある。ここで  $X(w)$  と  $Y(w)$  に強い相関がある場合を考える。図 2(a) におけるペア  $M_A(M_{A1}, M_{A2})$  (以下ペア  $M_A$  とする) の CSP 係数 ( $\text{CSP}_A(\theta)$ ) が図 2(b) である。ただし CSP 係数を式 (7) により時間軸 ( $\text{CSP}_A(k)$ ) から方向軸 ( $\text{CSP}_A(\theta)$ ) に変換している。

$$\theta = \cos^{-1} \left( \frac{c \cdot k / F_s}{d} \right) \quad (7)$$

図 2(b) より真の音源方向 ( $90^\circ$ ,  $130^\circ$ ) 以外にも CSP 係数が大きい方向があることがわかる。これがクロス項の影響である。特にサンプリング周波数が低い場合, サンプル点とサンプル点の間で高い相関値をもつことがあるため, 真の音源方向よりも, クロス項の方が CSP 係数が高くなる場合がたびたび起こる。CSP 法を用いて複数音源の位置推定を行うためにはサンプリング周波数を上げるなどにより, この問題に対処する必要がある。

また,  $X(w)$  と  $Y(w)$  の相関が弱い場合でも, 複数の音源方向推定が困難な場合がある。図 3(a) のようにマイクロホンに三つの音源から信号  $x(t)$ ,  $y(t)$ ,  $z(t)$  が到来する場合を考える。このとき, 音源  $Y$  の音圧が小さいときのペア  $M_A$  の CSP 係数は図 3(b) のようになり, 音源  $Y$  を示す CSP 係数よりも, その他の相関によるピーク  $Y'$  などの方が CSP 係数が大きくなっていることがわかる。

これより音圧が小さい音に対して精度良く方向を推定するためには, 真の音源方向の CSP 係数を強調することなどを検討する必要があることがわかる。

## 2.2 音源位置推定法

音源の方向が推定できれば, 各推定方向同士の交点

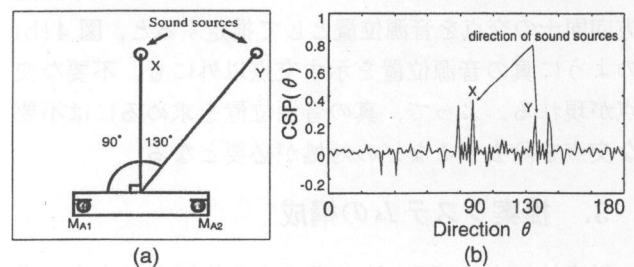


図2 音源間の相関が強い場合のペア  $M_A$  の CSP 係数  
Fig.2 CSP coefficients derived from microphone pairs  $M_{A1}$  and  $M_{A2}$  when two signals are correlated.



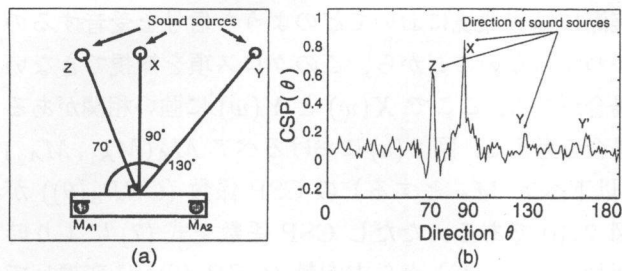


図3 音源 Y の音圧が小さい場合のペア  $M_A$  の CSP 係数  
Fig. 3 CSP coefficients of microphone pairs  $M_{A1}$  and  $M_{A2}$  when the magnitude of the sound source  $Y(w)$  is lower than that of the other sound sources.

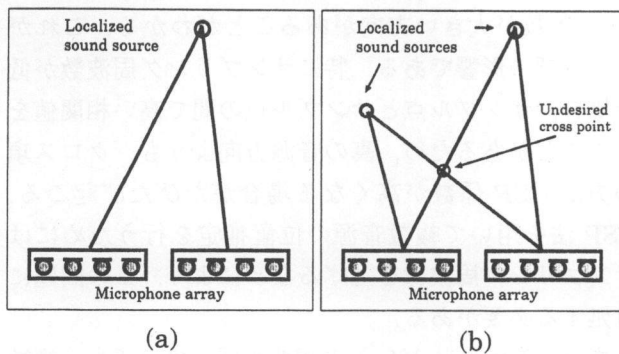


図4 推定方向による音源位置推定  
Fig. 4 Source localization of single sound source and multiple sound sources by finding cross points of the DOAs.

などを用いることにより音源位置を推定できる。以下その方法と問題点について述べる。

### 2.2.1 音源が単一の場合

音源が単一の場合、各マイクロホンアレーによる推定方向同士の交点を音源位置として推定することが可能である。音源位置は、図 4(a) に示す交点により求めることができる。

### 2.2.2 音源が複数の場合

音源が複数の場合、各マイクロホンアレーによる推定方向同士の交点を音源位置として推定すると、図 4(b) のように真の音源位置を示す交点以外にも、不要な交点が現れる。よって、真の音源位置を求めるには不要な交点を除去するなどの対処が必要となる。

## 3. 提案システムの構成

従来法である CSP 法は単一音源位置を推定する場合は有効であったが、複数の音源推定を行う場合、様々な相関の影響により複数の音源方向を推定することは困難であった。そこで本論文では、CSP 法を用いて複

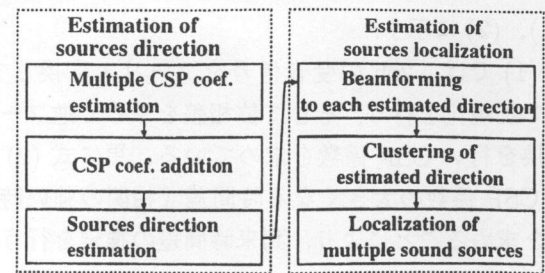


図5 提案法の処理手順  
Fig. 5 An overview of the proposed algorithm.

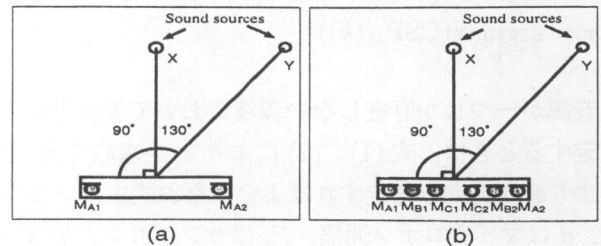


図6 マイクロホンの配置 (CSP 係数の加算法)  
Fig. 6 Position of microphones. (a) the conventional method. (b) the proposed method.

数音源の方向を推定するために、異なるマイクペアによる CSP 係数を加算して、様々な相関の影響を抑えて真の音源方向を強調する方法を提案する。また、推定方向から音源位置を推定するために、推定方向に対してクラスタリングを行うことにより不要な交点を除去する方法も併せて提案する。提案法の処理の流れを図 5 に示す。以下提案法の各処理について詳細に説明する。

### 3.1 CSP 係数の加算法

CSP 法による複数の音源方向の推定は、各音源から各マイクロホンへの信号の到来時間差を用いて推定するためマイクロホンの位置が変われば、真の音源方向及びクロス項による音源方向とも変化する。しかし、図 6(a) のペア  $M_A$  の内側に図 6(b) のようにペア  $M_B(M_{B1}, M_{B2})$ ,  $M_C(M_{C1}, M_{C2})$  (以下ペア  $M_B$ ,  $M_C$  とする) を各ペアの中心を合わせて配置すると、真の音源方向の CSP 係数は常に大きくなるが、クロス項の CSP 係数はペアにより係数の大きくなる方向が異なる。ペア  $M_A$ ,  $M_B$ ,  $M_C$  の CSP 係数を図 7(a)(b)(c) に示す。

図 7(a)(b)(c) より、ペアの位置が変われば CSP 係数は変化していることがわかる。だが、ペアの中心座標を合わせることで、真の音源方向 ( $90^\circ$ ,  $130^\circ$ ) の CSP 係数は常に大きくなる。よって、この CSP 係

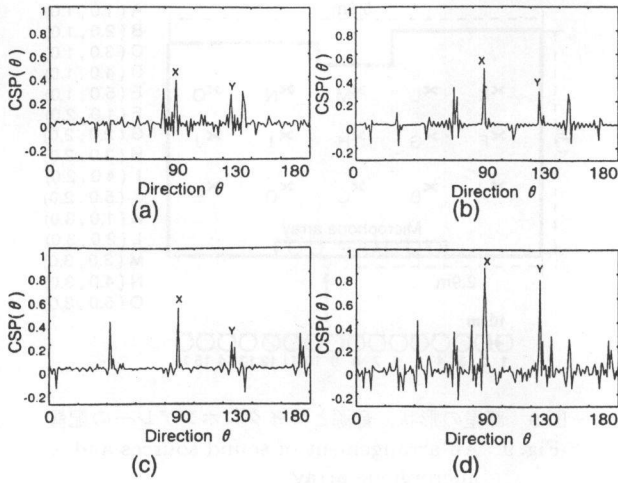


図7 CSP 係数

Fig.7 CSP coefficients derived from microphone pairs  $M_A, M_B$  and  $M_C$ , respectively.

数を加算すれば、真の音源方向の CSP 係数を強調できるはずである。図 7(d) はペア  $M_A, M_B, M_C$  の CSP 係数の加算結果である。加算においては、最も方向分解能が密な最外郭ペア（ペア  $M_A$ ）を基準にして、方向軸 ( $\theta$ ) においてペア  $M_B, M_C$  の CSP 係数 ( $CSP(\theta)$ ) を加算した。その結果、図 7(d) より真の音源方向 ( $90^\circ, 130^\circ$ ) の CSP 係数が強調されていることがわかる。この方法を用いれば、音源間の相関が強い場合はもちろんのこと、ある音源の音圧が小さい場合でも CSP 係数を加算することにより、その音源に対する CSP 係数が強調されるため、音源方向の推定精度が向上することが期待される。

今、図 6 のようなマイクロホンの配置で、音源から信号  $x(t), y(t)$  が到来する場合、受信信号  $s_{A1}(n), s_{A2}(n)$  及び  $s_{B1}(n), s_{B2}(n)$  は式 (8)~(11) のように表せる。

$$s_{A1}(n) = a_{A1}x(m + \varphi_{A1}) + b_{A1}y(m + \xi_{A1}) \quad (8)$$

$$s_{A2}(n) = a_{A2}x(m + \varphi_{A2}) + b_{A2}y(m + \xi_{A2}) \quad (9)$$

$$s_{B1}(n) = a_{B1}x(m + \varphi_{B1}) + b_{B1}y(m + \xi_{B1}) \quad (10)$$

$$s_{B2}(n) = a_{B2}x(m + \varphi_{B2}) + b_{B2}y(m + \xi_{B2}) \quad (11)$$

ここで信号  $x(t), y(t)$  に対して、対応する  $m$  は時刻、 $\varphi, \xi$  は到来時間、 $a, b$  は振幅、 $X(w), Y(w)$  は

信号  $x(t), y(t)$  の周波数表現を表したものである。式 (8), (9), 式 (10), (11) より式 (1) の  $DFT^{-1}[\cdot]$  の分子はそれぞれ式 (12), (13) と表せる。

$$\begin{aligned} DFT(s_{A1}(n))DFT(s_{A2}(n))^* &= a_{A1}a_{A2}X(w)^2e^{-jw(\varphi_{A1}-\varphi_{A2})} \\ &+ b_{A1}b_{A2}Y(w)^2e^{-jw(\xi_{A1}-\xi_{A2})} \\ &+ X(w)Y(w)(a_{A1}b_{A2}e^{-jw(\varphi_{A1}-\xi_{A2})} \\ &+ a_{A2}b_{A1}e^{-jw(\xi_{A1}-\varphi_{A2})}) \end{aligned} \quad (12)$$

$$\begin{aligned} DFT(s_{B1}(n))DFT(s_{B2}(n))^* &= a_{B1}a_{B2}X(w)^2e^{-jw(\varphi_{B1}-\varphi_{B2})} \\ &+ b_{B1}b_{B2}Y(w)^2e^{-jw(\xi_{B1}-\xi_{B2})} \\ &+ X(w)Y(w)(a_{B1}b_{B2}e^{-jw(\varphi_{B1}-\xi_{B2})} \\ &+ a_{B2}b_{B1}e^{-jw(\xi_{B1}-\varphi_{B2})}) \end{aligned} \quad (13)$$

ここで式 (12), (13) の位相項に着目する。ペア  $M_A$  と  $M_B$  の中心位置は等しいことから真の音源方向も等しくなる。よって、式 (3) より  $X(w)^2$  の項と  $Y(w)^2$  の項に対して式 (14), (15) が成り立つ。

$$\frac{\varphi_{A1} - \varphi_{A2}}{d_{A1,A2}} = \frac{\varphi_{B1} - \varphi_{B2}}{d_{B1,B2}} \quad (14)$$

$$\frac{\xi_{A1} - \xi_{A2}}{d_{A1,A2}} = \frac{\xi_{B1} - \xi_{B2}}{d_{B1,B2}} \quad (15)$$

ただし、 $d_{A1,A2}, d_{B1,B2}$  はマイクペアの間隔を示す。ここで、 $d_{A1,A2} = \alpha d_{B1,B2}$  とするとクロス項  $X(w)Y(w)$  に対して式 (16), (17) が成り立つ。

$$\begin{aligned} \frac{\varphi_{A1} - \xi_{A2}}{d_{A1,A2}} &= \frac{\varphi_{B1} - \xi_{B2}}{d_{B1,B2}} + \frac{\varphi_{A2} - \xi_{A1} + \alpha(\xi_{B1} - \varphi_{B2})}{\alpha d_{B1,B2}} \end{aligned} \quad (16)$$

$$\begin{aligned} \frac{\xi_{A1} - \varphi_{A2}}{d_{A1,A2}} &= \frac{\xi_{B1} - \varphi_{B2}}{d_{B1,B2}} + \frac{\xi_{A2} - \varphi_{A1} + \alpha(\varphi_{B1} - \xi_{B2})}{\alpha d_{B1,B2}} \end{aligned} \quad (17)$$

式 (16), (17) を見ると式 (18), (19) を満たしていればマイクペアの位置を変化させることにより、クロス項の発生方向が異なることがわかる。

$$\varphi_{A2} - \xi_{A1} + \alpha(\xi_{B1} - \varphi_{B2}) \neq 0 \quad (18)$$

$$\xi_{A2} - \varphi_{A1} + \alpha(\varphi_{B1} - \xi_{B2}) \neq 0 \quad (19)$$

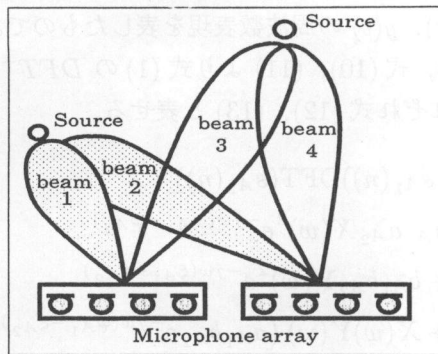


図8 ビームフォーミングによるクラスタリング  
Fig.8 DOA clustering with beam-forming.

これより式 (18), (19) を満たしていれば, CSP 係数の加算によりクロス項は削減されることがわかる.

### 3.2 ビームフォーミングを用いた複数音源方向のクラスタリング

方向を推定することができれば, マイクペアの推定方向同士の交点により音源位置を推定できる. しかし音源が複数存在する場合, 図4のように音源を示す交点以外にも不要な交点が現れる. この問題に対して, 不要な交点を削除する方法としてベクトル量子化 (VQ) によるクラスタリング法が提案されている [14]. しかし, この方法は各音源が離れて存在する場合有効であるが, 音源同士の間の距離がアレー長よりも短い場合, VQ を用いてクラスタリングするのは困難であった. そこでビームフォーミングを用いた新しいクラスタリング法を提案する.

推定方向に対して図8のように各々のアレーで遅延和法 (例えば文献 [3]) により音を抽出する. このとき抽出した音に対して, 同じ音源を示す音は相関値が高いことに着目し, 各々の抽出した音の相関を計算する. その相関値をもとに相関の高い方向から順にクラスタリングを行う. 例えば図8中の, beam1-2と3-4がクラスタリングできる. その結果, 図4中の不要な交点が除去できる. 本論文では相関値が最も高い方向同士の交点を音源位置としている.

## 4. 提案法の性能評価実験

本論文において提案法を評価するために音響実験室において実環境データの収録を行い, CSP 係数の加算による複数音源位置推定法の有効性の評価を行った.

### 4.1 音響実験室における実環境データの収録

簡易音響実験室においてデータの収録を行った. 収録に使用した部屋は図9に示す構造で, 小会議室程度

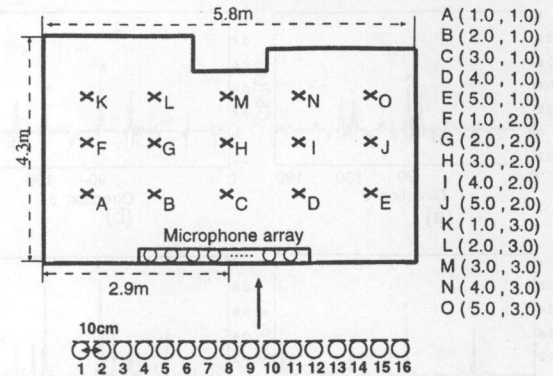


図9 部屋の形状, 音源とマイクロホンアレーの配置  
Fig.9 An arrangement of sound sources and a microphone array.

の大きさである. 室内残響時間 ( $T_{[60]}$ ) は, 0.18s である. マイクホンは無指向性マイクホンで 16 素子使用し, 間隔は 10 cm で配置した. また音源には話者 3 名の音声 (単語) を使用した. 収録は D/A と A/D を同期させて行った. A/D のサンプリング周波数は 48 kHz でローパスフィルタのカットオフは 16 kHz とした. 音源位置は図9に示す 15 箇所とした. また, 収録時の暗騒音レベルは 37.5 dBA, 室内温度は 18.0°C であった.

### 4.2 音源位置推定の実験条件

提案法では音源位置を推定するにあたり, 二つ以上のマイクロホンアレーが必要となる. そこで収録したマイクロホンの中から, array1, array2 の二つのマイクロホンアレーを見立てて実験を行った. 本実験ではマイクロホンアレーの構成が, 10 素子 × 2 組, 14 素子 × 2 組, の 2 通りを実験した. 表1に, 10 素子 × 2 組と 14 素子 × 2 組のときの選択したマイクロホンの番号を示す. また図10に素子選択の1例として 10 素子 × 2 組のときの 1 組目の使用するマイクロホン素子の様子を示す. 方向推定における CSP 係数は, 加算を行わないときは各マイクロホンアレーの最外殻ペアのみを使用する. また, 加算を行う場合は, 各マイクロホンアレーの最外殻ペア及び順次その内側のペアを使用しており, 加算回数 5 回まで実験した. また音源の数は, クロス項の影響がある 2, 3 音源を実験対象とした.

### 4.3 実験結果及び考察

10 素子 × 2 組と 14 素子 × 2 組のマイクロホンアレーを使用したときの複数音源位置推定の実験結果に対する位置推定率 ( $P_{cor}$ ) を図11, 図12に示す. 位



表1 マイクロホンアレーの構成  
Table 1 A configuration of microphone arrays.

素子数	array1		array2	
	No.	pair	No.	pair
10	1 ~ 10	1-10	7 ~ 16	7-16
		2-9		8-15
		3-8		9-14
		4-7		10-13
		5-6		11-12
14	1 ~ 14	1-14	3 ~ 16	3-16
		2-13		4-15
		3-12		5-14
		4-11		6-13
		5-10		7-12

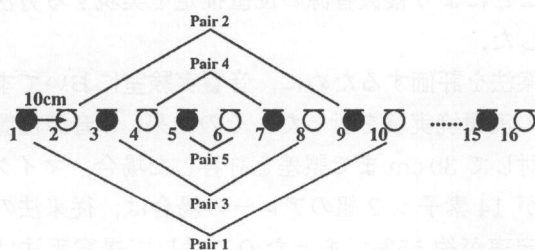


図10 マイクロホンの素子選択  
Fig. 10 A selection of microphones.

位置推定率 ( $P_{cor}$ ) は、実験結果 ( $Q_{est}$ ) と真の音源位置 ( $Q_{tru}$ ) の距離の差をもとに式 (20) により計算している。式 (21) 中の  $Er$  は誤差の許容範囲を示し、図 11, 図 12 の横軸の値を示す。また  $N$  はサンプル数を示す。

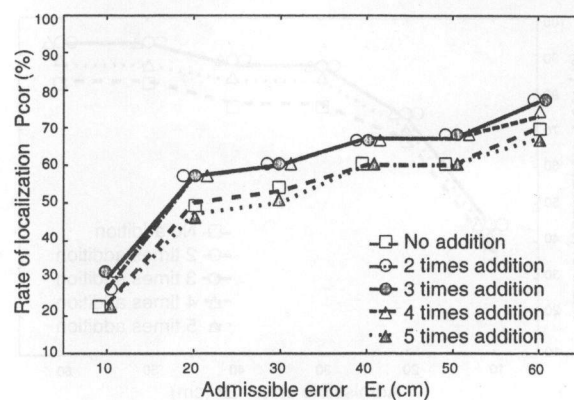
$$P_{cor} = \frac{\sum_{n=0}^N I_{cor}[n]}{N} \quad (20)$$

$$I_{cor}[n] = \begin{cases} 1 & \|Q_{est} - Q_{tru}\| \leq Er \\ 0 & \|Q_{est} - Q_{tru}\| > Er \end{cases} \quad (21)$$

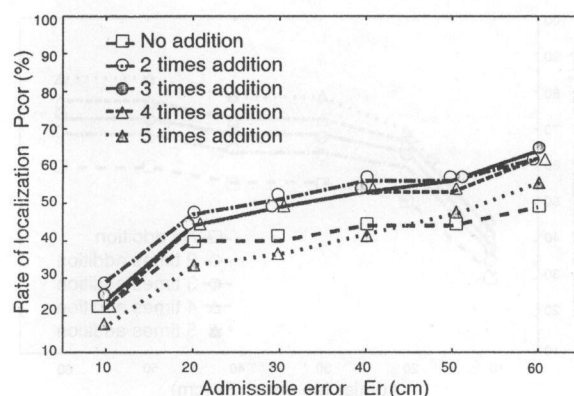
以下に考察を述べる。

- 図 11, 図 12 より CSP 係数の加算を行った方が、位置推定率が向上していることがわかる。クロス項の影響のある 2, 3 音源では加算を行うと著しく位置推定率が向上している。この結果から、CSP 係数の加算の有効性が確認できる。

- 加算の回数を増やすと位置推定率が向上することがわかる。これは加算を増やすことにより、真の音源方向が強調されるからである。しかし、特に素子数の少ないアレーの場合、ある回数を境にして、加算の回数を増やすと位置推定率が劣化するという結果を得た。これは、加算を行うほど加算する CSP 係数を求



(a) 2 sound sources



(b) 3 sound sources

図 11 位置推定率 (10 素子 × 2 組)

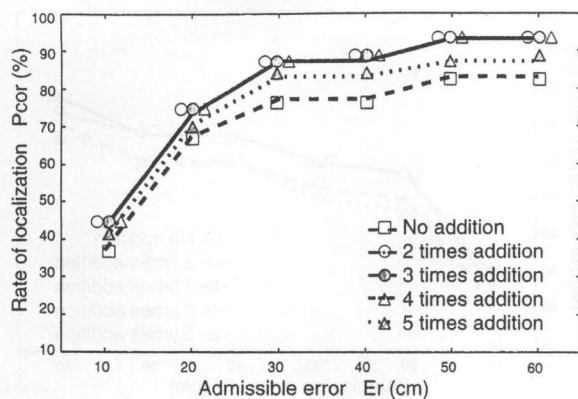
Fig. 11 Accuracy of sound source localization by 10 microphones × 2 sets.

めるマイクペアの間隔が小さくなるため、角度分解能が劣化する現象が起こる。その結果誤った係数を加算する場合があるため、位置推定率が劣化していると考えられる。

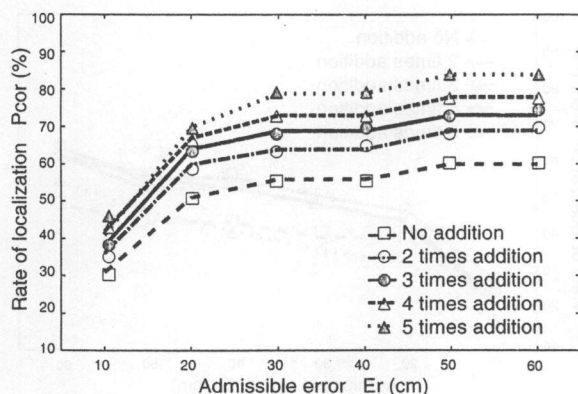
- 許容誤差 10 cm のときに比べて、許容誤差を 20 cm, 30 cm にすると位置推定率が飛躍的に向上している。これは、スピーカが点音源でないこと、及びマイクの指向性、スピーカ位置、マイク位置の計測誤差の影響などにより結果の位置推定精度が劣化していることが考えられる。

- また、マイクロホン素子数が多い (アレー長が長い) ほど推定位置と真の音源位置との距離の差が小さいことがわかった。これは、アレー長が長くなることにより、方向推定における CSP 係数を推定する際のマイクペアの間隔が広がるため、CSP 係数の角度分解能が向上する。その結果、推定位置が真の音源位置に近づくと考えられる。

- 更にアレーに対して側面方向に位置する音源に



(a) 2 sound sources



(b) 3 sound sources

図 12 位置推定率 (14 素子  $\times$  2 組)Fig. 12 Accuracy of sound source localization by 14 microphones  $\times$  2 sets.

対して、推定位置と真の音源位置との距離の差が大きくなった。これは、アレーに対して側面方向は CSP 係数の角度分解能が小さくなることが原因であると考えられる。また、推定位置の分布状況は、アレーの中心と真の音源位置を結ぶ直線上に分布する傾向が強い。これは位置推定に使用する二つのマイクロホンアレーの間隔に依存するものであると考えられる。

本提案法は実環境においても高い位置推定精度を保つことができるので、複数の音源位置推定に対して有効であると考えている。

## 5. む す び

マイクロホンアレーを使用すれば、話者から離れた位置にあるマイクロホンでも話者の音声を高品質に受音できる。しかし、マイクロホンアレーを用いて高品質に音声を受音するためには、話者の位置情報が必要となる。また、テレビ会議システムなどにおいては話者は複数存在し、同時に発話する状況が考えられる。

更に話者以外の音源も他数存在する。そこで複数の音源位置を推定してその中から話者位置を推定することが必要となる。音源位置を推定する一般的な方法は、まず音源の方向を推定し、その推定方向の交点により音源の位置を推定する。しかし、複数の音源方向及び位置を正確に推定することは困難である。そこで、本論文では、複数の音源方向を推定する方法として、CSP 係数の加算法を提案した。また、推定した音源方向から複数の音源位置を推定する方法として、ビームフォーミングを用いて推定音源方向をクラスタリングすることにより複数音源の位置推定を実現する方法を提案した。

提案法を評価するために、音響実験室において本提案法の実環境実験を行った。その結果、3 音源位置推定に対して 30 cm まで誤差を許容した場合、マイクロホンが 14 素子  $\times$  2 組のアレーの場合は、従来法の位置推定率が約 55%であったのに対して提案手法は約 80%となり、約 25%位置推定率を改善した。

今後は、まず位置推定誤りの分析を行う必要がある。更に、本実験では直線型アレーを使用した。全方向に均一な角度分解能をもつ円形アレーなど、他の形状のアレーにおいても検討する必要があると思われる。本研究では、音源位置のみを推定したが、複数の音源位置を推定した後にどの音源が話者であるのか推定することも必要である。また、本実験では残響時間が比較的短い音響実験室 ( $T_{60} = 0.18$  s) にて提案手法の有効性を示したが、今後、残響時間が長い場合すなわち残響の影響を無視できない環境における音源位置推定も検討する必要がある。

将来的には、話者の位置だけでなくその話者に対する低次の反射音源位置も推定して、直接音だけでなく反射音も有効利用するマルチビームフォーミング [13]~[15] を使用することにより、より高品質に音声抽出することを検討している。

## 文 献

- [1] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-27, no.2, pp.113-120, 1979.
- [2] J.S. Lim and A.V. Oppenheim, "All-pole modeling of degraded speech," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-26, no.3, pp.197-210, 1978.
- [3] 安倍正人, "多数センサによる音源推定," 音響誌, vol.51, no.5, pp.384-389, 1995.
- [4] S.U. Pillai, Array Signal Processing, Springer-Verlag,



New York, 1989.

- [5] 大賀寿郎, 山崎芳男, 金田 豊, 音響システムとデジタル処理, コロナ社, pp.173-218, 1995.
- [6] 田中雅史, 金田 豊, “近距離音源の方向探索における平面波近似の有効性について,” 音講論, no.2-7-13, pp.445-446, March 1991.
- [7] 小林和則, 島田正治, 穂刈治英, “複数マイク自由配置による複数話者位置推定,” 信学論 (A), vol.J82-A, no.2, pp.193-200, Feb. 1999.
- [8] C.H. Knapp and G.C. Carter, “The generalized correlation method for estimation of time delay,” IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-24, no.4, pp.320-327, 1976.
- [9] M. Omologo and P. Svaizer, “Acoustic event localization using a crosspower-spectrum phase based technique,” Proc. ICASSP94, pp.273-276, 1994.
- [10] M. Brandstein, J. Adcock, and H. Silverman, “A closed-form method for finding source locations from microphone-array time-delay estimates,” Proc. ICASSP95, pp.3019-3022, 1995.
- [11] M. Omologo and P. Svaizer, “Acoustic source location in noisy and reverberant environment using CSP analysis,” Proc. ICASSP96, pp.921-924, 1996.
- [12] P. Svaizer, M. Matassoni, and M. Omologo, “Acoustic source location three-dimensional space using crosspower spectrum phase,” ICASSP97, pp.231-234, 1997.
- [13] J.L. Flanagan, A.C. Surendran, and E.E. Jan, “Spatially selective sound capture for speech and audio processing,” Speech Communication, vol.13, pp.207-222, 1993.
- [14] 西岡良典, 山田武志, 中村 哲, 鹿野清宏, “方向同定を備えたマルチビームマイクロホンアレー,” 信学技報, EA97-98, 1998.
- [15] 西浦敬信, 中村 哲, 鹿野清宏, “実環境における反射音を利用したマルチビームフォーミングの検討,” 信学技報, SP99-133, 2000.

(平成 11 年 7 月 30 日受付, 12 年 1 月 31 日再受付)



西浦 敬信 (学生員)

平 9 奈良高専・専攻科・電子情報卒. 平 11 奈良先端大・情報科学研究科・博士前期課程了. 現在, 同博士後期課程在学中. マイクロホンアレーを用いた音環境認識の研究に従事. IEEE, 日本音響学会各会員.



山田 武志 (正員)

平 6 阪市大・工・情報卒. 平 8 奈良先端大・情報科学研究科・博士前期課程了. 平 11 同博士後期課程了. 同年筑波大学電子・情報工学系講師. 博士 (工学). マイクロホンアレー, 音声・音環境の認識・理解の研究に従事. IEEE, 情報処理学会, 日本

音響学会各会員.



中村 哲 (正員)

昭 56 京都工繊大・工芸・電子卒. 昭 56 ~ 平 6 シャープ (株) 中央研究所及び情報技術研究所に勤務. 昭 61 ~ 平 1 ATR 自動翻訳電話研究所に出向. 平 6 年 4 月より奈良先端科学技術大学院大学情報科学研究科助教授. 平 8 年 3 月 ~ 8 月 Rutgers University・CAIP Center 客員教授. 音声情報処理, 主として音声認識の研究に従事. 京都大学博士 (工学). 平 4 日本音響学会栗屋学術奨励賞受賞. IEEE, 情報処理学会, 日本音響学会, 人工知能学会各会員.



鹿野 清宏 (正員)

昭 45 名大・工・電気卒. 昭 47 同大大学院工学研究科修士課程了. 同年電電公社武蔵野電気通信研究所入所. 昭 59 ~ 昭 61 カーネギーメロン大客員研究員. 昭 61 ~ 平 2 ATR 自動翻訳電話研究所音声情報処理研究室長. 平 4 NTT ヒューマンインタフェース研究所主席研究員. 平 6 年 4 月より奈良先端科学技術大学院大学情報科学研究科教授. 音情報処理学講座を担当, 主として音声・音情報処理の研究及び研究指導に従事. 工博. 昭 50 本会米沢賞, 平 3 IEEE SP 1990 Senior Award, 平 6 日本音響学会技術開発賞各受賞. IEEE, 情報処理学会, 日本音響学会各会員.