

# 分散配置されたマイクロホンによる ブラインド音源分離の検討\*

小野拓磨 (東大・工), 宮部滋樹, 小野順貴, 嵯峨山茂樹 (東大・情報理工)

## 1 はじめに

元信号と混合過程が未知である条件からマイクロホンアレーを用いて音源を分離するブラインド音源分離 (BSS: Blind Source Separation) の手法はこれまで数多く提案されてきており、この中でも独立成分分析 (ICA: Independent Component Analysis) [1] に基づくものは確立されている。これら従来のマイクロホンアレーを用いた手法は、すべてのマイクロホンを近接して並べる場合が一般的であったが、これではマイクロホンから音源までの距離が遠い場合や到来方向が同程度の方向からの音源の分離は難しいものであった。これに対して各マイクロホンを空間的に適当に分散させることで、多様な音源配置に対応した音源分離が可能となると考えることができる。このような分散配置されたマイクロホンの枠組みは、音声認識やテレビ会議を視野に入れた研究 [2, 3] 等近年注目され重要性が高まっており、その中でブラインド音源分離の手法もいくつか提案されている [3, 4]。周波数領域 ICA を用いて音源分離を行うために必要なパーミュテーション解決の手掛かりとしては、到来時間差に基づくものが最も強力であるが、マイクロホンが分散配置されていてマイクロホン間の距離が大きくなる場合には、空間的エイリアシングの影響が大きくなり、到来時間差の利用が複雑になる。文献 [3, 4] では振幅差を元にパーミュテーションを解決しているが、マイクロホン間で大きな振幅差が表れない場合には分離が難しくなるという欠点がある。本報告では、マイクロホン数個を近接させたユニットを作り、このユニット複数個を分散配置させる音源分離の枠組みを提案する。ユニット単位では空間的エイリアシングの影響を受けずに到来時間差を推定できるため、パーミュテーション解決が単純化される。この有効性を検証するため、分散配置の応用例に基づいたワイヤレスマイクロホンでの実環境実験における分離性能の評価を報告する。

## 2 分散配置されたマイクロホンによる BSS

### 2.1 周波数領域 ICA

音源数を  $N$ 、マイクロホン数を  $M$  とすれば観測信号のベクトル列  $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]^T$  は

$$\mathbf{x}(t) = \sum_{u=0}^{\infty} \begin{bmatrix} a_{11}(u) & \cdots & a_{1N}(u) \\ \vdots & \ddots & \vdots \\ a_{M1}(u) & \cdots & a_{MN}(u) \end{bmatrix} \mathbf{s}(t-u) \quad (1)$$

と表せる。ここで、 $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^T$  は元信号のベクトル列、 $a_{mn}(t)$  は  $n$  番目の音源から  $m$  番目のマイクロホンへのインパルス応答である。これを短時間フーリエ変換して周波数領域で考えると  $\mathbf{X}(\tau, \omega) \simeq A(\omega)\mathbf{S}(\tau, \omega)$  のように表せる。ここで  $\tau$  はフレーム数、 $A(\omega)$  は混合行列である。周波数領域

ICA により、混合過程が未知の条件で観測信号から元信号を復元した信号  $\mathbf{Y}(\tau, \omega) = W(\omega)\mathbf{X}(\tau, \omega)$  を推定する。ここで  $W(\omega)$  は分離行列である。周波数領域 ICA の解はスケールの任意性とパーミュテーション問題が残るが、スケールの任意性については Projection Back [5] を用いて解決した。

### 2.2 主成分分析による次元削減

マイクロホンを分散配置させる利点の一つは、マイクロホンアレーを密集させた一つのシステムに比べ、容易にマイクロホンの数を増やすことができる点である。この利点を生かし、音源数より観測数の方が多い優決定音源分離をおこない、主成分分析により次元を削減した。直行基底  $[\gamma_1(\omega), \dots, \gamma_N(\omega)]$  を用いて残差の分散を最も小さくするような近似  $\mathbf{X}(\tau, \omega) \simeq \sum_{n=1}^N \tilde{X}_n(\tau, \omega)\gamma_n(\omega)$  を考え、観測信号  $\mathbf{X}(\tau, \omega)$  を低次元に射影する変換  $\tilde{\mathbf{X}}(\tau, \omega) = [\tilde{X}_1(\tau, \omega) \cdots \tilde{X}_N(\tau, \omega)]^T = W_{PCA}(\omega)\mathbf{X}(\tau, \omega)$  をおこなった。この処理後に周波数領域 ICA により求めた分離行列  $W_{ICA}(\omega)$  を用いて復元信号を  $\mathbf{Y}(\tau, \omega) = W_{ICA}(\omega)\tilde{\mathbf{X}}(\tau, \omega) = W_{ICA}(\omega)W_{PCA}(\omega)\mathbf{X}(\tau, \omega)$  と学習した。

### 2.3 時間差ベクトルのクラスタリングによる

#### パーミュテーション解決

次に、ユニットとして近接した 2 つのマイクロホンがペアとなっているとみなすことのできるステレオマイクロホンを分散させることを考え、ステレオマイクロホンのペア内のマイクロホンの時間差を使ったパーミュテーション解決法を提案する。以下ではステレオマイクロホンのペア数を  $M_p$  ( $2M_p = M$ ) とおく。混合行列の推定値  $\hat{A}(\omega)$  は

$$\hat{A}(\omega) = \underset{A(\omega)}{\operatorname{argmin}} E\{\|A(\omega)\mathbf{Y}(\tau, \omega) - \mathbf{X}(\tau, \omega)\|^2\} \quad (2)$$

を解けばよく

$$\hat{A}(\omega) = E\{\mathbf{X}(\tau, \omega)\mathbf{Y}^H\} (E\{\mathbf{Y}(\tau, \omega)\mathbf{Y}^H(\tau, \omega)\})^{-1} \quad (3)$$

と求まる。これは各マイクロホンのステアリングベクトルの推定を与え、ここからステレオマイクロホンペア  $i, j$  内での  $n$  音源目の時間差  $d_n^{ij}(\omega)$  は

$$d_n^{ij}(\omega) = \{\arg(\hat{A}_{in}(\omega)) - \arg(\hat{A}_{jn}(\omega))\} / \omega \quad (4)$$

と求まる。ここで  $\hat{A}_{in}(\omega)$  は  $\hat{A}(\omega)$  の  $(i, n)$  要素である。直接波の到来時間差は周波数によらず、また多くの帯域で直接波はドミナントになるため、 $k$  番目の角周波数での時間差  $d_n^{ij}(\omega_k)$  をステレオマイクロホンペア数だけ並べた  $M_p$  次元ベクトル  $\mathbf{d}_{kn}$  の推定値は、角周波数によらないのでこの時間差ベクトル  $\mathbf{d}_{kn}$  をクラスタリングして直接波の到来時間差を推定することによりパーミュテーションを推定することができ

\*Investigation of blind source separation using dispersively arranged microphones.  
by ONO Takuma, MIYABE Shigeki, ONO Nobutaka, SAGAYAMA Shigeki (The University of Tokyo)

Table 1 実環境実験の条件

sources	TIMIT database (female)
microphones	SANYO HM-W300
room size	6.3 m × 7.7 m × 2.7 m
reverberation time	300 ms
sampling rate	16 kHz
data length	5.0 s
frame length	8192 points(512 ms)
frame shift	2048 points
ICA algorithm	Infomax

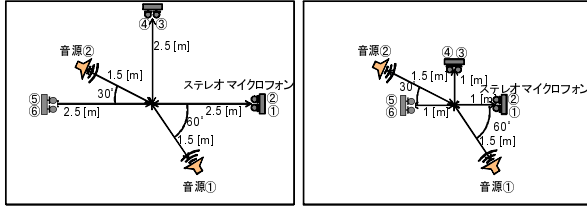


Fig. 1 音源とマイクロホン配置:設置型と会議型

る。ここで、到来時間差の推定はパワーが大きいものほど信頼性が高いため、周波数帯域・マイクロホンペア毎にパワーに応じた信頼度を持たせる。ここでは各マイクロホンペアの各周波数での振幅のフレーム平均を重み  $w_{kn}$  として与え、重み付き距離を最小化する  $k$ -means アルゴリズムによるクラスタリングをおこなった。以下のように  $1, \dots, N$  を並べ替えた順列  $p_k(n) (n = 1, \dots, N)$  とクラスタ中心  $\mathbf{m}_n$  を更新する。

$$p_k(n) \leftarrow \underset{p_k(n)}{\operatorname{argmin}} \sum_k \sum_{n=1}^N \sum_{j=1}^{M_p} w_{kp_k(n)}^j |r_{kn}^j|^2 \quad (5)$$

$$\mathbf{r}_{kn} = \mathbf{d}_{kp_k(n)} - \mathbf{m}_n$$

$$\mathbf{m}_n \leftarrow [m_n^1, \dots, m_n^{M_p}]^T$$

$$m_n^j = \frac{\sum_k w_{kp_k(n)}^j d_{kp_k(n)}^j}{\sum_k w_{kp_k(n)}^j} \quad (j = 1, \dots, M_p)$$

ここで右上の添え字  $j$  はベクトルの  $j$  行目の要素を表す。このように各周波数ごとに時間差をクラスタリングすることでパーミュテーション問題を解決した。

### 3 実環境実験と評価

#### 3.1 実験条件

マイクロホンを部屋の隅々に配置する場合や容易にマイクロホンの数を増やせる利点を生かした例としてワイヤレスマイクロホンを使用した実環境実験をおこなった。ワイヤレスシステムによりすべてのマイクロホン間の時間同期はとれている元で、部屋の壁に設置されたマイクロホンを想定したマイクロホンが音源の外側になる配置(以下‘設置型’)と会議録音を想定したマイクロホンが音源の内側になる配置(以下‘会議型’)の二種類について実験した。中心から 1.5 [m] の距離にある 2 つの音源に対して、ステレオマイクロホン 3 組 (モノラルマイクロホン 6 個)

Table 2 実験結果: SIR による評価

microphone number		1 - 2	3 - 4	5 - 6	
SIR [dB]	提案法	音源 1	4.79	7.36	6.68
		音源 2	4.92	0.61	3.48
	理想的	音源 1	9.57	9.99	10.23
		音源 2	8.69	7.30	9.47
SIR [dB]	提案法	音源 1	11.59	10.10	8.51
		音源 2	6.23	8.52	7.62
	理想的	音源 1	12.57	11.15	11.12
		音源 2	7.52	9.18	8.76

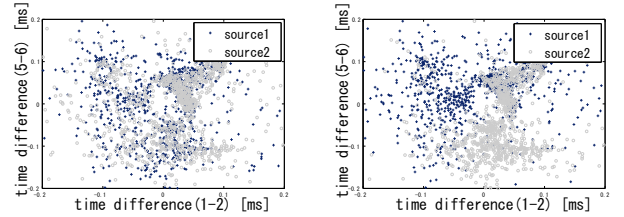


Fig. 2 時間差ベクトル: クラスタリングの前後

を設置型では中心から 2.5 [m] の位置に 3 か所、会議型では中心から 1.0 [m] の位置に 3 か所を配置した。

#### 3.2 実験結果と考察

復元信号は各マイクロホンの元信号から音源数だけ得られる。Table 2 に各ステレオマイクロホンの片方からの復元信号の SIR (Single to Interference Ratio) による評価を、提案手法を用いてパーミュテーション解決した場合と元信号との SN 比から理想的にパーミュテーション解決した場合について示す。また、会議型配置でのクラスタリングの結果をある二つのステレオマイクロホンのペア内時間差を横軸と縦軸にとり Fig. 2 に示す。実験結果よりマイクロホンが音源の内側にあるような会議型配置では理想的な場合に近い分離性能が得られており、提案したパーミュテーション解決の手法が有効であることを示している。このような会議型配置で分散されたマイクロホンでのブラインド音源分離が可能であると言える一方で、マイクロホンが音源の外側の設置型配置では分離性能が低いという結果を得た。これは、音源の指向性から音源の後方にあるマイクロホンでは到来時間差が正確に求まらず、パーミュテーション問題がうまく解けなかったためと考えられる。この解決方法について我々は、離れたマイクロホン間の振幅比を比較することでパーミュテーション問題を解決する手法を検討している。

#### 参考文献

- [1] A. Hyvärinen *et al.*, Independent Component Analysis, John Wiley, New York, 2001.
- [2] Yong Zhao *et al.*, *Proc. ICASSP*, pp. 4181–4184, 2009.
- [3] Jacek P. Dmochowski *et al.*, *Proc. ICASSP*, pp. 89–92, 2009.
- [4] E. Robledo-Arnuncio *et al.*, *Proc. ICASSP*, pp. 949–952, 2007.
- [5] N. Murata *et al.*, *Neurocomputing*, vol. 41, no. 1–4, pp. 1–24, 2001.