

狭間隔・広間隔の複数マイクロホン対を用いた 周波数領域ブラインド音源分離*

○向井 良 澤田 宏 荒木 章子 牧野 昭二

(日本電信電話株式会社, NTT コミュニケーション科学基礎研究所)

1 はじめに

独立成分分析 (ICA: Independent Component Analysis) を用いたブラインド音源分離 (BSS: Blind Source Separation) では、原理的には音源数以上の観測信号があれば分離可能とされているが、残響下において実際に多数の音源が存在する場合を扱った例は少ない。その理由として、時間領域での手法では計算量が膨大になり、現実的な処理時間で分離フィルタを計算するのが困難であること [1]、周波数領域での手法では、分離された信号を全ての周波数ビンで揃えるというパーミュテーションの問題の解決が難しいことなどが考えられる。

我々はこれまでに、音源位置 (方向や距離) 推定による方法と信号の相関による方法を併用してパーミュテーションを安定的に解く手法を提案した [2][3]。しかしながら、音源数が多く、さまざまな方向に配置されているような場合には、単に線形アレイの素子数を増やしただけでは音源位置推定に関するいくつかの問題が発生する。本稿では、それらの問題点について説明し、軸方向の異なる複数の狭間隔マイクロホン対と広間隔マイクロホン対を組合せて用いることにより、パーミュテーションの解決に必要な音源位置情報を頑健に推定する手法を提案する [4]。実験では 8 マイクロホン 6 音源 (うち 2 つは同一方向に配置) の分離に成功した。

2 周波数領域 BSS

周波数領域 BSS の処理の流れを図 1 に示す。音源信号を $s_i(t)$ ($i = 1, \dots, N$)、観測信号を $x_j(t)$ ($j = 1, \dots, M$)、分離信号を $y_k(t)$ ($k = 1, \dots, N$) とすると、BSS のモデルは $x_j(t) = \sum_{i=1}^N (h_{ji} * s_i)(t)$, $y_k(t) = \sum_{j=1}^M (w_{kj} * x_j)(t)$ と書ける。ここで、 h_{ji} は音源 i からマイクロホン j へのインパルス応答、 w_{kj} は分離フィルタである。短時間フーリエ変換を用いることにより、時間領域における畳み込み混合は周波数領域における瞬時混合、 $\mathbf{X}(\omega, n) = \mathbf{H}(\omega)\mathbf{S}(\omega, n)$, $\mathbf{Y}(\omega, n) = \mathbf{W}(\omega)\mathbf{X}(\omega, n)$ として扱うことができる。ここで、 ω は角周波数、 n はフレーム番号、 $\mathbf{S}(\omega, n) = [S_1(\omega, n), \dots, S_N(\omega, n)]^T$ は周波数ビン ω の音源信号、 $\mathbf{X}(\omega, n) = [X_1(\omega, n), \dots, X_M(\omega, n)]^T$ は観測信号、 $\mathbf{Y}(\omega, n) = [Y_1(\omega, n), \dots, Y_N(\omega, n)]^T$ は分離信号、 $\mathbf{W}(\omega)$ は分離行列である。 $\mathbf{W}(\omega)$ は $\mathbf{Y}(\omega, n)$ の各要素が互いに独立となるようにして求められる ICA の解である。ICA の解にはスケールリングとパーミュテーションのあいまいさがある。すなわち、 $\mathbf{W}(\omega)$ が ICA の解だとすると、任意の複素対角行列 $\mathbf{D}(\omega)$ 、任意のパーミュテーション行列 (左から掛けて行を入れ替える行列) $\mathbf{P}(\omega)$ があつたときに、 $\mathbf{D}(\omega)\mathbf{P}(\omega)\mathbf{W}(\omega)$ もまた解になる。それゆえ、時間領域における分離信号を再構成するためには $\mathbf{D}(\omega)$ と $\mathbf{P}(\omega)$ を全ての ω について正しく求める必要がある。本稿では、スケールリングについては minimal distortion principle (MDP) による解 $\mathbf{D}(\omega) = \text{diag}(\mathbf{W}^{-1}(\omega))$ を採用し、以下ではパーミュテーションの解決方法について議論する。

*Frequency Domain BSS using Large and Small Spacing Microphone Pairs, Ryo Mukai, Hiroshi Sawada, Shoko Araki, and Shoji Makino (NTT Communication Science Laboratories, NTT Corporation)

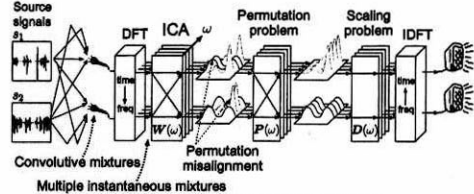


図 1: 周波数領域 BSS の流れ

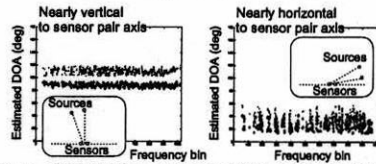


図 2: 音源位置による方向推定の精度の違い

3 パーミュテーション解決のための音源位置情報

3.1 ICA の解を用いた音源位置推定

ICA の解 $\mathbf{W}(\omega)$ から、パーミュテーション解決の手がかりとなる音源位置情報を推定することができる。狭間隔マイクロホン対を用い、 $\mathbf{W}^{-1}(\omega)$ を遠距離場モデルにあてはめると、音源 i のマイクロホン対 jj' に対する相対的な方向 $\theta_{i,jj'}$ の推定値は以下の式で求められる [2]。

$$\hat{\theta}_{i,jj'} = \cos^{-1} \frac{\arg([\mathbf{W}^{-1}]_{ji}/[\mathbf{W}^{-1}]_{j'i})}{\omega c^{-1} \|(\mathbf{p}_j - \mathbf{p}_{j'})\|} \quad (1)$$

ここで、 $[\cdot]_{ji}$ は行列の j 行 i 列目の要素、 c は音速、 $\mathbf{p}_j$ はマイクロホン j の位置を表すベクトルである。

一方、広間隔マイクロホン対を用い、近距離場モデルをあてはめると、音源 i が存在する球面の中心および半径の推定値は以下の式で求められる [3]。

$$\text{中心 } \hat{\mathbf{O}}_{i,jj'} = \mathbf{p}_j - \frac{1}{r_{i,jj'}^2 - 1} (\mathbf{p}_{j'} - \mathbf{p}_j) \quad (2)$$

$$\text{半径 } \hat{R}_{i,jj'} = \frac{r_{i,jj'}}{r_{i,jj'}^2 - 1} \|\mathbf{p}_{j'} - \mathbf{p}_j\| \quad (3)$$

ここで $r_{i,jj'} = \|[\mathbf{W}^{-1}]_{ji}/[\mathbf{W}^{-1}]_{j'i}\|$ である。

3.2 方向推定の問題点と解決方法

図 2 は、音源がマイクロホン対の軸と垂直方向にある場合、軸に沿った方向にある場合のそれぞれについて、(1) による方向推定の結果を示した例である。これを見ると、音源が軸の方向に近いときに推定誤差が大きい。この理由は次のように説明できる。

ICA の解から求められる値 $\arg([\mathbf{W}^{-1}]_{ji}/[\mathbf{W}^{-1}]_{j'i})$ に含まれる誤差を $\Delta \arg(\hat{H})$ 、 $\hat{\theta}_{i,jj'}$ に含まれる誤差を $\Delta \hat{\theta}$ と書くと、それらの比は (1) を偏微分した以下の式で近似できる。

$$\frac{\Delta \hat{\theta}}{\Delta \arg(\hat{H})} \approx \frac{1}{\omega c^{-1} \|\mathbf{p}_j - \mathbf{p}_{j'}\| \sin(\hat{\theta}_{i,jj'})} \quad (4)$$

図 3 は、いくつかの周波数においてこの比をプロッ

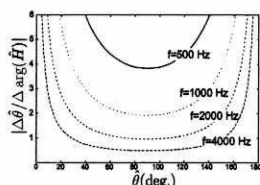


図 3: ICA 解の誤差と方向推定誤差との関係

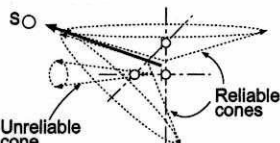


図 4: 複数のマイクロホン対を用いた方向推定

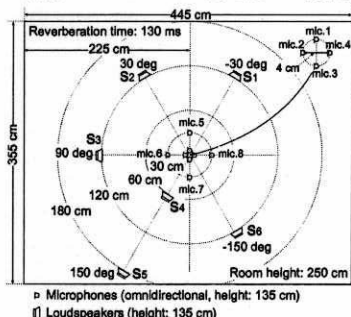


図 5: 音源とマイクロホンの配置

表 1: 実験条件

Sampling rate	8 kHz
Data length	6 s
Frame length	2048 point (256 ms)
Frame shift	512 point (64 ms)
ICA algorithm	Infomax (complex valued)

トしたものである。これを見ると、音源がマイクロホン対の軸方向に近いときには ICA 解の誤差が大きく拡大されて方向推定の誤差に影響することがわかる。

方向推定に関するもう一つの問題に、単一のマイクロホン対や線形アレイでは、軸に対して対称な位置 (3 次元的には円錐面上) にある音源を区別できない、ということがある。このことは、音源数が多く、さまざまな方向に存在する場合に特に問題となる。

これらの問題を解決するために、軸方向の異なる複数のマイクロホン対を利用することを提案する。(1) を用いて推定される方向 (3 次元的には円錐) を、複数のマイクロホン対について組み合わせることにより、音源方向を一意に推定することができる。その際、推定した方向によって信頼度に重み付けをすること (もしくは、推定した方向が軸に近い場合には単に棄却する等の処理) によって、音源方向にかかわらず頑健な方向推定が可能となる (図 4)。さらに、広間隔マイクロホン対を併用すると、(2)、(3) によって音源が存在する球面が推定でき、いくつかの音源が同一方向に存在する場合にパーミューテーションを解決するための手がかりとして利用できる。

4 実験および結果

室内で実測したインパルス応答を畳み込んで作成した混合音声を用い、6 音源 8 マイクロホンという条件で分離実験を行った。音源の配置を図 5 に、その他の主な実験条件を表 1 に示す。なお、音源数 $N = 6$ は既知とする。実験は以下の手順で行った。

まず、周波数領域の ICA によって分離行列 $\mathbf{W}(\omega)$ を求める。次に、 $\mathbf{W}(\omega)$ の各行のうち、マイクロホ

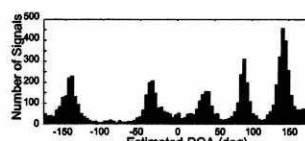


図 6: 方向推定結果のヒストグラム

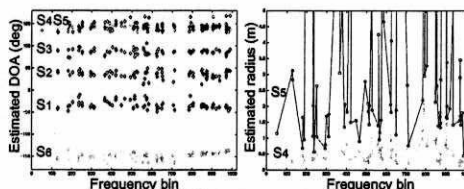


図 7: 位置情報を用いた信号の分類

表 2: 実験結果 (dB)

	SIR ₁	SIR ₂	SIR ₃	SIR ₄	SIR ₅	SIR ₆	ave.
入力 SIR	-8.3	-6.8	-7.8	-7.7	-6.7	-5.2	-7.1
C	4.4	2.6	4.0	9.2	3.6	-2.0	3.7
D+S+C	12.3	5.6	14.5	7.6	8.9	10.8	10.0

ン対 (1-3, 2-4, 1-2, 2-3) に相当する行を用いて音源方向推定を行う。推定した方向のヒストグラム (図 6) には 5 つのクラスタがあり、そのうち 1 つが他の約 2 倍の大きさであるため、6 つの音源のうち 2 つが同一方向 (150° 付近) から到来していると推測される。4 つの音源については推定した方向をもとにパーミューテーションを解決し (図 7(左))、残りの 2 つについては広間隔マイクロホン対 (7-5, 7-8, 6-5, 6-8) を用いて推定した球面の半径によって信号を区別した。例として、マイクロホン対 7-5 を用いて推定した球面の半径を図 7(右) に示す。

残響や推定誤差の影響により、位置情報だけではパーミューテーションを完全に解決することはできない。そのため、推定した位置情報によって信号が矛盾なく分類できた周波数ビンはその情報にもとづいて解決し、残りの周波数ビンは相關にもとづく方法を用いて解決した。最後に、時間領域の分離フィルタを求める際には、スペクトルの平滑化を行った [5]。

分離性能 (SIR) の評価結果を表 2 に示す。"C" は相關法のみによってパーミューテーションを解決したもの、"D+S+C" は方向推定と球面推定による解決の後に相關法を用いたもの (提案手法) である。提案手法では 6 音源全ての分離に成功し、SIR の改善は入力 SIR から平均 17.1 dB となった。実験結果のサンプル音声は [6] で聴くことができる。

参考文献

- [1] K. Matsuoka, Y. Ohba, Y. Toyota, and S. Nakashima, "Blind separation for convolutive mixture of many voices," in *Proc. IWAENC 2003*, 2003, pp. 279-282.
- [2] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," in *Proc. Intl. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, 2003, pp. 505-510.
- [3] 向井 良, 澤田 宏, 荒木 章子, 牧野 昭二, "周波数領域 BSS における近距離場モデルを用いたパーミューテーションの解法," 音講論集, Sep. 2003, pp. 589-590.
- [4] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Frequency domain blind source separation using small and large spacing sensor pairs," in *Proc. of ISCAS 2004*, 2004, accepted.
- [5] H. Sawada, R. Mukai, S. de la Kethulle de Ryhove, S. Araki, and S. Makino, "Spectral smoothing for frequency-domain blind source separation," in *Proc. IWAENC 2003*, 2003, pp. 311-314.
- [6] <http://www.kecl.ntt.co.jp/icl/signal/mukai/demo/iwaenc2003/>