

周波数領域BBSにおける近距離場モデルを用いた パーミュテーションの解法*

○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

残響下におけるブラインド音源分離(BSS: blind source separation)の手法の一つに周波数領域におけるICA(independent component analysis)がある。周波数領域での手法は、時間領域における畳み込み混合を複数の瞬時混合に置き替えて解くことにより、分離系を高速に求めることができるが、時間領域の信号を再構成するためには各周波数ごとに得られた解の並び方をそろえるという、いわゆるパーミュテーション問題を解く必要がある。

我々はこれまでに、信号の到来方向推定に基づく方法[1]と近傍の周波数間の相関による方法[2]を統合した解法を提案し[3]、さらに、倍音の関係にある周波数間の相関も利用することによって頑健にパーミュテーション問題を解決できることを示した[4]。

しかしながら信号が同じ方向から到来する場合には、到来方向推定に基づく方法が使えないため、相関による方法のみに頼らざるをえない。相関による方法は、1つの周波数での誤りが他の周波数に波及するため非常に不安定である。

本稿では、ICAによって得られた分離行列を近距離場モデルを用いて解釈することによって、音源が存在する球面を推定できることを示す。それによって得られた位置情報を到来方向推定の代わりに用いることにより、音源が同一方向にある場合でも安定的にパーミュテーションを解決でき、音源分離が可能であることを示す。これまでに、同一方向から到来する信号の分離を扱った例としては[5]などがあるが、我々の手法は、畳み込み混合を扱うこと、音源が必ずしも同一方向にある必要は無いこと、などに特徴がある。

2 周波数領域BSS

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 $y_k(t)$ ($k = 1, \dots, N$) とするとき、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モデルは周波数領域における瞬時混合モデル、 $\mathbf{X}(\omega, n) = \mathbf{H}(\omega)\mathbf{S}(\omega, n)$, $\mathbf{Y}(\omega, n) = \mathbf{W}(\omega)\mathbf{X}(\omega, n)$ 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(時刻 t に相当)、 $\mathbf{S}(\omega, n) = [S_1(\omega, n), \dots, S_N(\omega, n)]^T$ 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 $\mathbf{Y}(\omega, n)$ 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。

* A Solution for the Permutation Problem in Frequency Domain BSS using Near-Field Model, Ryo Mukai, Hiroshi Sawada, Shoko Araki, and Shoji Makino (NTT Communication Science Laboratories, NTT Corporation)

ICAの解にはスケー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) $\mathbf{P}(\omega)$ があったときに、 $\mathbf{P}(\omega)\mathbf{D}(\omega)\mathbf{W}(\omega)$ もまたICAの解になる。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。スケーリングについては、minimal distortion principle(MDP)[6]を用いた解 $\mathbf{D}(\omega) = \text{diag}(\mathbf{W}^{-1}(\omega))$ を採用することで、マイクで観測された音源信号 $((h_{ii} * s_i)(t))$ の良い推定値が復元できる。

3 提案手法

3.1 ICAの解におけるスケーリングに関する不变量
もし \mathbf{W} がスケーリングのあいまいさを除いて求められたとすると、 $\mathbf{D}(\omega)\mathbf{W}(\omega)\mathbf{H}(\omega) = \mathbf{I}$ が成立立つような \mathbf{D} が存在し、混合系の推定値は $\hat{\mathbf{H}}(\omega) = \mathbf{W}^{-1}\mathbf{D}^{-1}$ となる。MDPによって求めた \mathbf{D} は一般にはこれを満たさないため、ICAの解だけから混合系 \mathbf{H} そのものを推定することはできない。しかしながら、 $\hat{\mathbf{H}}$ の同じ列にある2つの要素の比 $\hat{H}_{ji}/\hat{H}_{j'i}$ は

$$\frac{\hat{H}_{ji}}{\hat{H}_{j'i}} = \frac{[\mathbf{W}^{-1}\mathbf{D}^{-1}]_{ji}}{[\mathbf{W}^{-1}\mathbf{D}^{-1}]_{j'i}} = \frac{[\mathbf{W}^{-1}]_{ji}}{[\mathbf{W}^{-1}]_{j'i}} \quad (1)$$

($[\cdot]_{ji}$ は、行列の j 行 i 列目の要素を表す) で与えられるため、 \mathbf{D} に対して不变な量として推定可能である。以前に我々はこの量を遠距離場モデルと比較することによって方向推定ができる事を示した[4]。今回は、これを近距離場モデルと比較することによって、音源が存在する球面が推定できることを示す。

3.2 近距離場モデルによる音源位置の推定

信号の減衰を考慮した近距離場モデルによると、混合系 $H_{ji}(\omega)$ は以下のように表現できる。

$$H_{ji}(\omega) = \frac{1}{\|\mathbf{q}_i - \mathbf{p}_j\|} e^{j\omega c^{-1}(\|\mathbf{q}_i - \mathbf{p}_j\|)} \quad (2)$$

ここで \mathbf{q}_i は音源 i の位置、 \mathbf{p}_j はマイクロホン j の位置、 c は音速である。マイクロホン j と j' (マイクロホンペア jj' と呼ぶ)について式(2)の比をとり、式(1)と比較して絶対値をとると、以下の式を得る。

$$\frac{\|\mathbf{q}_i - \mathbf{p}_{j'}\|}{\|\mathbf{q}_i - \mathbf{p}_j\|} = \left| \frac{[\mathbf{W}^{-1}]_{ji}}{[\mathbf{W}^{-1}]_{j'i}} \right| \quad (3)$$

この式の右辺の値を $r_{i,jj'}$ と書き、 \mathbf{q}_i について解くと、

$$\text{中心 } O_{i,jj'} = \mathbf{p}_j - \frac{1}{r_{i,jj'}^2 - 1} (\mathbf{p}_{j'} - \mathbf{p}_j) \quad (4)$$

$$\text{半径 } R_{i,jj'} = \left\| \frac{r_{i,jj'}}{r_{i,jj'}^2 - 1} (\mathbf{p}_{j'} - \mathbf{p}_j) \right\| \quad (5)$$

の球面が得られる。図1は、 $r_{i,jj'}$ をパラメータとしたときに式(3)によって決まる球面の様子を示したものである。

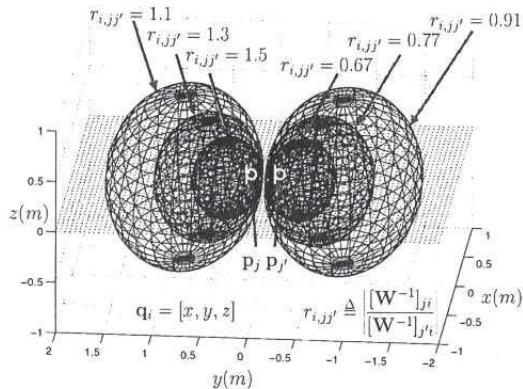


図 1: 式(3)によって与えられる球面の様子 ($p_j = [0, 0.15, 0]$, $p_{j'} = [0, -0.15, 0]$)

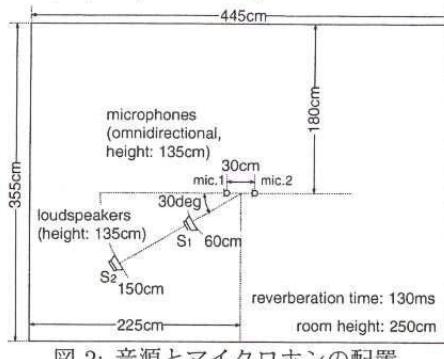


図 2: 音源とマイクロホンの配置

このようにして、ICA の解 \mathbf{W} とマイクロホンペア jj' の座標から q_i が存在する球面の推定値 $(O_{i,jj'}, R_{i,jj'})$ が得られる。マイクロホンペアが複数ある場合には、複数の情報を組み合わせたり、到來方向推定と組み合わせたりすることにより、音源位置をより詳細に推定することができる。この方法によれば、同一方向にある音源でも区別することができるため、推定した音源位置に従って出力信号を並び替えることによりパーミュテーションを解決することができる。

4 実験および考察

残響時間約 130ms の室内において測定したインパルス応答に話者 4 名(男声 2, 女声 2)の音声を畳み込んで作成した 12 通りの組合せの混合音声を用いて分離実験を行った。音源信号およびマイクロホンの配置を図 2 に、その他の実験条件を表 1 に示す。

今回は 2 音源、2 マイクロホンでの実験のため、音源位置によるパーミュテーションの解決は、推定された球の半径 $R_{1,12}$ と $R_{2,12}$ を単純に比較し、 $R_{1,12} \leq R_{2,12}$ となるように解いた。

音源位置による方法、(隣接ビンおよび倍音関係のあるビン間の)相関による方法、両者を併用した方法を比較するために、しきい値パラメータ A_{th} を導入し、 $\max(R_{1,12}, R_{2,12}) \geq A_{th} \cdot \min(R_{1,12}, R_{2,12})$ を満足する周波数ビンは、音源位置による解決が信頼できると判断して音源位置による方法を適用し、それ以外の周波数ビンでは相関による方法を適用する。 $A_{th} = 1.0$ のときは全て音源位置による方法で、 $A_{th} = \infty$ のときは全て相関による方法で解くことになる。 A_{th} を変化させながら、分離性能(SIR: Signal to Interference Ratio)を音声の組合せごとにプロットした結果を図 3 に示す。

表 1: 実験条件

Sampling rate	8 kHz
Data length	2 s
Frame length	1024 point (128 ms)
Frame shift	256 point (32ms)
ICA algorithm	Infomax (complex valued)

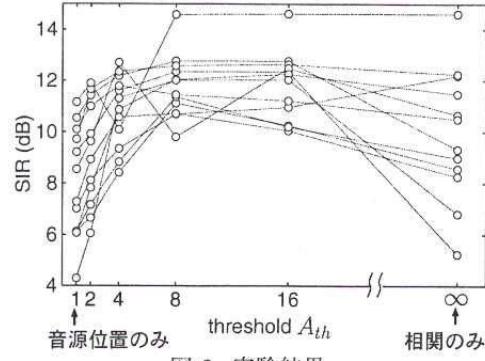


図 3: 実験結果

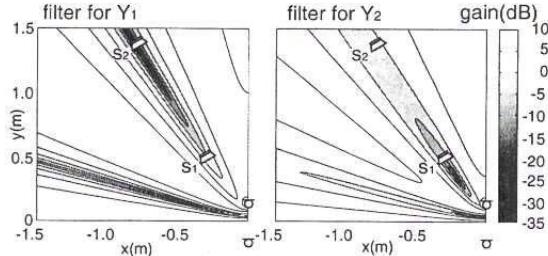


図 4: 分離系の空間的ゲイン特性 ($f = 1000\text{Hz}$)

この結果より、音源位置のみによる方法では、全体的に性能が悪く、相関のみによる方法では分離性能がばらつくことがわかる。両者を併用すると安定して高い性能が得られた。これは、到來方向推定による方法と相関による方法とを併用した場合の結果[4]と同様の傾向である。また、 A_{th} は比較的大きな値を用いる必要があることがわかった。良い性能が得られた $A_{th} = 8 \sim 16$ のときには、全体の $1/5 \sim 1/10$ 程度の周波数ビンを音源位置によって決定していた。

最後に、分離系のゲイン特性を調べた結果を示す。図 4 は 1 つの周波数ビン ($f=1000\text{Hz}$) におけるゲイン特性(マイク 1 で観測した信号をそのまま出力したときのゲインを基準とし、横軸、縦軸に空間の x 方向 y 方向をとり、 $z=0$ の平面で切った断面)をプロットしたものである。これによると、妨害音の位置にスポット的に null を形成して分離を実現していることがわかった。

参考文献

- S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in Proc. of ICASSP'00, 2000, pp. 3140-3143.
- S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," in Proc. of ICANN'98, 1998, pp. 761-766.
- 澤田 宏, 向井 良, 荒木 章子, 牧野 昭二, "周波数領域ブレインド音源分離における permutation 問題の解法," 音講論集, Sep. 2002, pp. 541-542.
- 澤田 宏, 向井 良, 荒木 章子, 牧野 昭二, "周波数領域ブレインド音源分離における permutation 問題の頑健な解法," 音講論集, Mar. 2003, pp. 777-778.
- M. Iwaki and A. Ando, "Selective microphone system using blind separation by block decorrelation of output signals," in Proc. Int'l. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003), pp. 1023-1028.
- K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in Proc. of Int'l. Workshop on Independent Component Analysis and Blind Signal Separation (ICA'01), 2001, pp. 722-727.