

周波数領域ブラインド音源分離における permutation 問題の解法*

○澤田 宏, 向井 良, 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

ブラインド音源分離 (BSS: blind source separation) は、観測された混合信号のみから混合前の音源信号を推定する技術であり、独立成分分析 (ICA: independent component analysis) [1] に基づく手法が盛んに研究されている。遅延・残響を伴う実環境での混合に対しては、ある程度の長さの FIR フィルタを用いて分離を行う必要がある。フィルタを直接求めずにその周波数特性を ICA で求める周波数領域での手法は、瞬時混合モデルに ICA を適用できるため幅広く用いられているが、周波数間で信号の並び順を揃えるという permutation の問題を解決する必要がある。これまでに、指向特性をプロットして妨害音の方向を推定する方法 [2] や、分離信号の相関を用いる方法 [3] などが提案されてきたが、常に十分に精度良く解決できるとは言い難い。

本稿では、上記の2種類の方法をうまく組み合わせ、安定的に高い性能で permutation を解く手法を提案する。また、その中の妨害音方向の推定に関しては、分離行列から解析的にその方向を求める方法を提案する。これにより、指向特性を逐一プロットすることが不要となり、高速な処理が可能となる。従来手法との比較を実験で行ったところ、提案手法の有効性が確認された。

2 周波数領域 BSS

P 個の音源 $s_p(t)$ が残響も含めて混合され Q 個のマイク $x_q(t) = \sum_{p=1}^P \sum_k h_{qp}(k) s_p(t-k)$ で観測されたとする。ここで、 $h_{qp}(k)$ は音源 p からマイク q へのインパルス応答である。BSS の目的は、 $s_p(t)$ や $h_{qp}(k)$ を知ることなしに、分離のための FIR フィルタ係数 $w_{rq}(k)$ と分離信号 $y_r(t) = \sum_{q=1}^Q \sum_k w_{rq}(k) x_q(t-k)$ を求めることにある。図1に $P=Q=2$ の場合の構成を示す。

周波数領域の手法では、ICA を用いてフィルタ $w_{rq}(k)$ の周波数応答 $W_{rq}(f)$ を求める。そのために、観測信号 $x_q(t)$ の時間周波数表現 $X_q(f, m)$ を短時間フーリエ変換により求め、各周波数 f で瞬時混合の ICA: $\mathbf{Y}(f, m) = \mathbf{W}(f) \mathbf{X}(f, m)$ を解く。ここで、 $\mathbf{W}(f)$ は要素が $W_{rq}(f)$ である $P \times Q$ の行列であり、 $\mathbf{X}(f, m) = [X_1(f, m), \dots, X_Q(f, m)]^T$, $\mathbf{Y}(f, m) = [Y_1(f, m), \dots, Y_P(f, m)]^T$ である。

ICA では、 $\mathbf{W}(f)$ の行が入れ替わっても $\mathbf{Y}(f, m)$ の独立性は保たれるため、任意の2つの周波数 f_1 と f_2 の間で、 $Y_r(f_1, m)$ と $Y_r(f_2, m)$ が同じ音源に対応しているとは限らない。従って、周波数間で $Y_r(f, m)$ が同じ音源に属するように、 $\mathbf{W}(f)$ の行を揃える必要がある。これが permutation の問題である。これを解決した後、 $W_{rq}(f)$ に逆フーリエ変換を施すことで、最終的にフィルタ $w_{rq}(k)$ が得られる。

*Solving the permutation problem in frequency-domain blind source separation, by H. Sawada, R. Mukai, S. Araki, S. Makino (NTT Communication Science Labs., NTT Corp.)

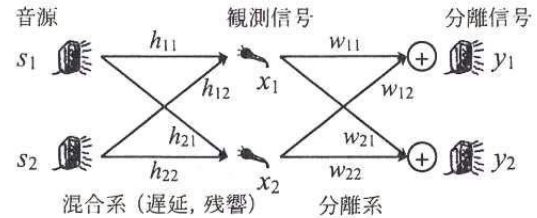


図 1: 2 音源 2 マイクの BSS

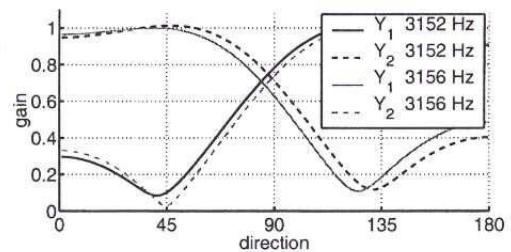


図 2: 指向特性と permutation

3 妨害音方向の推定による方法

マイク間隔が適度に狭ければ、BSS の分離フィルタ w_{rq} は多くの場合、妨害音方向からの音を消すように形成される。例えば図1の w_{11} と w_{12} は、 s_2 の方向からの音を消すように働く。従って、各周波数でどの方向の音を消しているかを解析することにより、permutation をある程度解決することができる。

その解析を行う一般的な方法は、指向特性をプロットし、最もゲインが小さい角度 (null の角度) を求めることである [2]。そのためにまず、混合系のインパルス応答を直接音のみで近似し、さらに平面波を仮定する。音源 s_p の到来方向を $0^\circ \leq \theta_p \leq 180^\circ$ (マイクの並びと垂直な方向が 90°)、マイク q の位置を d_q とすると、混合系の周波数応答は $H_{qp}(f) = e^{j2\pi f c^{-1} d_q \cos \theta_p}$ と表現できる (c は音速)。すると、角度 θ_p にある音源 s_p から分離信号 y_r への周波数応答 $\sum_q W_{rq}(f) \cdot H_{qp}(f)$ として指向特性 $B(f, \theta_p) = \sum_q W_{rq}(f) \cdot e^{j2\pi f c^{-1} d_q \cos \theta_p}$ が求まる。図2は、隣り合う2つの周波数に関して、指向特性のゲインをプロットしたものである。ゲインが最小となっている角度が null である。図2では、2つの周波数の間で null の位置が互い違いになっているため、どちらか一方の周波数で Y_1 と Y_2 を入れ替えることで permutation が解決される。

さて、null の位置を求めるためには、指向特性のゲインの最小値を与える角度を求めなければならない。最も単純な方法は、 θ_p を少しずつ変えて指向特性をプロットして最小値を求めることであるが、この処理を周波数毎に行うと非常に時間がかかる。そこで本稿では、解析的

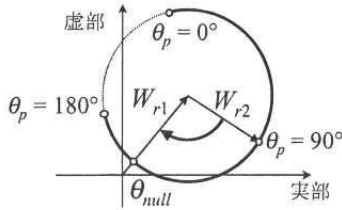


図 3: 複素平面上での指向特性の解釈

に null の位置を求める手法を提案する (ただし 2 音源 2 マイクを仮定する). 座標系を $d_1 = 0, d_2 = d$ と設定すると, $B(f, \theta_p) = W_{r1} + W_{r2} \cdot e^{j2\pi f c^{-1} d \cos \theta_p}$ と書ける. 図 3 はその複素平面上での解釈である. W_{r1} を中心として, W_{r2} が $e^{j2\pi f c^{-1} d \cos \theta_p}$ に従って回転し, 原点に最も近づく時点が null を与える角度 θ_{null} となる. その際には, W_{r2} を回転させたものが W_{r1} とは逆の方向を向くことになるため, $\text{angle}(W_{r2}) + 2\pi f c^{-1} d \cos \theta_{null} = \text{angle}(W_{r1}) \pm \pi$ が成り立つ. 従って,

$$\theta_{null} = \cos^{-1} \frac{\text{angle}(W_{r1}) - \text{angle}(W_{r2}) \pm \pi}{2\pi f c^{-1} d}$$

で求めることができる. なお, $\pm \pi$ の符号は, 分子が $-\pi$ から π の間に収まるように選ぶ. f の値が小さい低周波数では, $\cos^{-1}$ の引数の絶対値が 1 を越え θ_{null} が虚数になる場合があるが, これは null が存在しないことを意味する.

本提案手法により, 高速に null の角度を求めて permutation を解決することが可能となる. しかし, 全ての周波数で理想的な null が形成されるわけではなく, 特に低周波数では形成されないことが多い. 従って, より高精度に permutation を解決するためには, 次章で述べる相関による方法など他の手法に頼る必要がある.

4 周波数間の信号の相関による方法

2 つの信号 $x(t)$ と $y(t)$ の相関を $\text{cor}(x, y) = [\langle x \cdot y \rangle - \langle x \rangle \cdot \langle y \rangle] / (\sigma_x \cdot \sigma_y)$ と定義する. ここで, $\langle \cdot \rangle$ は時間平均, σ は標準偏差である. $\text{cor}(x, x) = 1$ であり, x と y が無相関ならば $\text{cor}(x, y) = 0$ である.

$Y_r(f, m)$ の絶対値の移動平均 (いわゆる envelope)

$$v_r^f(t) = \frac{1}{2M+1} \sum_{|m-l| \leq M} |Y_r(f, m)|$$

は, 同じ音源 s_p に対応する場合, 特に近傍の周波数で高い相関を持つ. 従ってこれらの相関を求めることで permutation (本章以降では π で表現) を決定できる.

まず考えられる方法は, 周波数の差 D 以下の近傍で相関の和が最も大きくなるように

$$\pi_f = \arg\max_{\pi} \sum_{|k-f| \leq D} \sum_p \text{cor}(v_{\pi[p]}^f, v_{\pi_k[p]}^k) \quad (1)$$

に基づき周波数 f での permutation を求めていくものである. しかしこの方法では, ある周波数で permutation を間違えるとその先すべての周波数でも間違えることになり, 安定性に欠ける. そのため文献 [3] では, ICA による分離性能が良いと思われる周波数から順に

$$\pi_f = \arg\max_{\pi} \sum_p \text{cor}(v_{\pi[p]}^f, \sum_{k \in F} v_{\pi_k[p]}^k) \quad (2)$$

に従って permutation を決定していく. ここで F は既に permutation が決定した周波数の集合である. しかしこの方法では, 近傍ではない周波数間でも envelope の相関が強いことを仮定するため, 実際にその仮定を満たさない音源に対しては不具合が起こる.

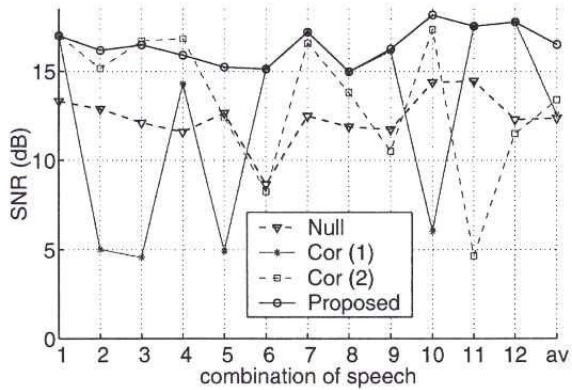


図 4: 実験結果

5 統合手法

これまでの議論を踏まえ本研究では, null による方法と相関による方法を統合して permutation を解決する. まず, null による方法を用いて, 図 2 のような理想的な null が形成されている周波数のみにおいて permutation を決定する. 次に, permutation が未定の周波数 f に対しては, 周波数の差が D 以下の近傍で既に permutation が決定した周波数との相関の和を最大化するように π_f を決定していく.

$$\pi_f = \arg\max_{\pi} \sum_{|k-f| \leq D, k \in F} \sum_p \text{cor}(v_{\pi[p]}^f, v_{\pi_k[p]}^k)$$

6 実験結果および考察

実験では, 12 組の音声データにインパルス応答を畳み込んで混合し, 周波数領域の ICA [4] を適用して分離した. 図 4 に実験結果を示す. “av” は 12 組の平均である. 比較のため permutation の解決は以下の 4 つを用いた. “Null” は null による方法, “Cor (1)” は式 (1), “Cor (2)” は式 (2) による方法, “Proposed” は提案する統合手法である. “Null” は安定的に解決しているが性能が不十分である. “Cor (1)” や “Cor (2)” は非常に良い場合もあるが安定性に欠ける. “Proposed” は常に良い性能となっている.

null による方法では方向という絶対的な基準で permutation を解決するため, 精度にはやや欠けるが, 大きく間違えることが少ない. また, 3 章で提案した解析的手法により高速に処理できる. 一方, 相関による方法では, 高い精度で permutation を解決できるが, どこかで間違った時の被害が大きい. 提案手法は, これら 2 種類の利点を活かして統合しているため, 安定的に高い精度で permutation を解決できるものとする.

参考文献

- [1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.
- [2] 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. ICASSP2000*, 2000, pp. 3140–3143.
- [3] S. Ikeda and N. Murata, “An approach to blind source separation of speech signals,” in *Proc. ICANN '98*, Sept. 1998, pp. 761–766.
- [4] H. Sawada, R. Mukai, S. Araki, and S. Makino, “Polar coordinate based nonlinear function for frequency-domain blind source separation,” in *Proc. ICASSP2002*, May 2002, pp. 1001–1004.