

# 周波数領域ブラインド音源分離における permutation 問題の頑健な解法 \*

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音源信号を分離・抽出する技術である. 遅延・残響を伴う実環境での混合に対しては, 残響時間に適した長さの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求める周波数領域での手法は, ICA (independent component analysis) を瞬時混合モデルに適用できるため幅広く用いられているが, 周波数毎に得られた ICA の結果の並び順を周波数間で揃えるという permutation の問題を解決する必要がある. 我々は前回の発表 [1] で, 信号の到来方向による方法 [2] と分離信号の相関を利用する方法 [3] を統合した解法を提案した. しかし, 残響の影響がかなり強い場合は時々 permutation を間違えることを, その後の実験で確認した. 原因は, 低域での到来方向推定の困難さにある.

本稿では, 前回提案した解法を以下の 2 点に関して改善する. まず, 上記の問題に対応するため, 音声の調波構造を利用して低域での permutation を頑健に解決する手法を提案する. 次に, 以前は 2 音源にしか適用できなかった信号の到来方向の推定を 3 音源以上に拡張する. これらの改善により, 頑健さが増し, 適用範囲も広がった.

## 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,  $Q$  個のマイクで観測信号  $x_q(t) = \sum_{p=1}^P \sum_k h_{qp}(k) s_p(t-k)$  が得られたとする. 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. 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. この permutation の問題を解決した後,  $W_{rq}(f)$  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.

\*A robust method for the permutation problem of frequency-domain blind source separation, by H. Sawada, R. Mukai, S. Araki, S. Makino (NTT Communication Science Laboratories, NTT Corporation)

```
/* 到来方向の推定による permutation の決定 */
for (forall f) {
  Theta(f) = DOA(f, W(f)) /* 到来方向推定 */
  Theta_s(f) = sort(Theta(f))
}
m_Theta = <Theta_s(f)>_f /* 各周波数での推定方向を平均 */
F = empty /* permutation が決定した周波数の集合 */
for (forall f) {
  if (confident(Theta(f), m_Theta, W(f))) {
    pi_f = getPermutation(Theta(f))
    F = F union {f}
  }
}
/* 分離信号の相関による permutation の決定 */
while (exists f not in F) {
  for (forall f not in F) {
    c_f = max_pi sum_{|g-f| <= D, g in F} sum_p cor(v_{pi(p)}^f, v_{pi(g)}^g)
    pi_f = argmax_pi sum_{|g-f| <= D, g in F} sum_p cor(v_{pi(p)}^f, v_{pi(g)}^g)
  }
  i = argmax_f c_f
  F = F union {i}
  c_i = 0
}
```

図 1: 統合手法のアルゴリズム

## 3 統合手法

前回提案した permutation の解法は, 2 つの手法を統合し, 2 段階で構成されている (図 1). 前半では各周波数毎に信号の到来方向を推定して permutation ( $\pi$  で表現) を決定する. ただし, その方向推定が不確かであったり, ICA による分離が不十分と判定された周波数では決定しない. 後半では, 分離信号の相関に基づき, 未決定の周波数の permutation を決定していく. この仕組みにより, 前半で絶対的な基準で確実に permutation を決定し, 全体として大きく間違えることを防ぐ.

アルゴリズム前半では, まず,  $DOA(f, \mathbf{W}(f))$  により, 分離行列  $\mathbf{W}(f)$  の各行から信号の到来方向を推定して  $P$  個の方向  $\Theta(f)$  を求める. その後,  $confident(\Theta(f), \mathbf{m}_\Theta, \mathbf{W}(f))$  により, 推定方向  $\Theta(f)$  が平均値  $\mathbf{m}_\Theta$  と大きく異ならず, かつ  $\mathbf{W}(f)$  による分離が十分であるかを判定する. これを満足すれば到来方向に基づいて決定を行う:  $\pi_f = getPermutation(\Theta(f))$ . なお  $F$  は permutation が決定された周波数の集合である.

次に後半では, 分離信号  $Y_r(f, m)$  の絶対値 (いわゆる envelope)  $v_r^f(m) = |Y_r(f, m)|$  が同じ音源に対応する場合, 特に近傍の周波数で高い相関  $cor(x, y) = \langle x \cdot y \rangle - \langle x \rangle \cdot \langle y \rangle / (\sigma_x \cdot \sigma_y)$  を持つことを利用して permutation を決定する. 具体的には, permutation が決定していない周波数  $f \notin F$  に対して, 周波数の差  $D$  以下の近傍で permutation が決定した周波数との相関の和  $\sum_{|g-f| \leq D, g \in F} \sum_p cor(v_{\pi(p)}^f, v_{\pi(g)}^g)$  を最大化するように決定していく.

```

for ( $\forall f \notin F$ ) {
   $H_a = \text{setOfHarmonics}(f)$ ;
   $c_f = \max_{\pi} \sum_{g \in H_a \cap F} \sum_p \text{cor}(v_{\pi(p)}^f, v_{\pi_g(p)}^g)$ 
  if ( $c_f \geq \text{threshold}$ ) {
     $\pi_f = \text{argmax}_{\pi} \sum_{g \in H_a \cap F} \sum_p \text{cor}(v_{\pi(p)}^f, v_{\pi_g(p)}^g)$ 
     $F = F \cup \{f\}$ 
  }
}

```

図 2: 調波構造を利用するアルゴリズム

## 4 提案手法

### 4.1 調波構造の利用

まず、音声の調波構造を利用して統合手法を改善する。一般に低周波数では、マイク間での位相差が小さくなるため、到来方向の推定が難しい。残響の影響が強い場合はさらに難しくなる。その結果、ある周波数より低域では、統合手法の前半では permutation が全く決まらず、後半の相関による方法だけに頼ることもある。この場合、どこかで間違えるとそれより低域では全て間違ってしまうことになるため、頑健な方法とは言えない。

この問題に対応するため、本稿では音声の調波構造を利用する。良く知られているように、音声の envelope は基本周波数  $f$  とその倍音  $2f, 3f, \dots$  で高い相関を持つ。従って、例えば周波数  $f$  が未決定で  $2f, 3f$  が既決定の場合、 $f$  が基本周波数に近ければ、

$$c_f = \max_{\pi} \sum_{g=2f, 3f} \sum_p \text{cor}(v_{\pi(p)}^f, v_{\pi_g(p)}^g)$$

は十分に大きくなり、高い確信度で  $f$  の permutation を決定できる。図 2 に調波構造を利用する部分のアルゴリズムを示す。 $\text{setOfHarmonics}(f)$  は周波数  $f$  の倍音の集合を与え、 $\text{threshold}$  は確信度を調節する。

### 4.2 信号の到来方向の推定方法

次に、信号の到来方向の推定方法を改善する。文献 [2] では、 $W(f)$  によって形成される指向特性をプロットし、その最小値としての null を求めることで妨害音の到来方向を推定している。我々は前回の発表 [1] で、その最小値を解析的に求める式を提案している。しかし、いずれも、3 音源以上への適用は難しい。

本稿では 3 音源以上にも適用できる新たな方法を提案する。ここでは、null の角度 (妨害音方向) を求める代わりに、分離行列の逆行列  $H(f) = W^{-1}(f)$  により混合系の周波数特性を推定し、直接、信号の到来方向  $\Theta(f)$  を求める。行列  $H(f)$  の要素  $H_{qp}(f)$  を実数の減衰度  $A_{qp}$  と原点での位相差  $e^{j\varphi_p}$  を用いて  $H_{qp}(f) = A_{qp} e^{j\varphi_p} e^{j2\pi f c^{-1} d_q \cos \theta_p}$  とモデル化する。ここで  $\theta_p$  は音源  $p$  の到来方向、 $d_q$  はマイク  $q$  の位置、 $c$  は音速である。ただし ICA の性質から  $H(f)$  の列の順序と大きさには任意性が残る。同じ音源に対応する別の要素  $H_{q'p}(f)$  も考慮してそれらの比を求めると、 $H(f)$  の列の大きさの任意性は取り除かれる:

$$H_{qp}/H_{q'p} = A_{qp}/A_{q'p} e^{j2\pi f c^{-1} (d_q - d_{q'}) \cos \theta_p}$$

この両辺の偏角を取ると  $\theta_p$  を推定する式が求まる:

$$\theta_p = \cos^{-1} \frac{\text{angle}(H_{qp}/H_{q'p})}{2\pi f c^{-1} (d_q - d_{q'})}$$

列の順序の任意性により、 $\theta_p$  は別の音源に対応するかもしれない。なお、2 音源の場合、本方法は null を求める解析式 [1] と同じ結果になることが証明できる。

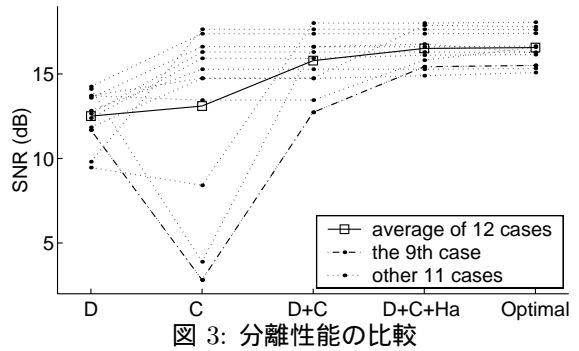


図 3: 分離性能の比較

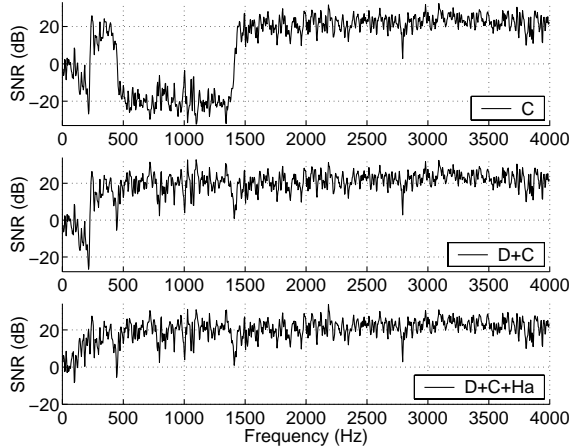


図 4: 9 組目の各周波数での SN 比

## 5 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 BSS を適用して分離した。permutation の解法は以下の 5 種類を用いた。

- D 到来方向の推定 (図 1 の前半) のみ
- C 信号の相関 (図 1 の後半) のみ
- D+C 前回の統合手法 (図 1 全体)
- D+C+Ha 今回の提案手法 (図 1 図 2 図 1 の後半)
- Optimal 最適 (各周波数で SN 比を計算して決定)

図 3 に全体の結果を示す。“D” は安定的に解決しているが性能が不十分である。“C” は非常に良い場合もあるが安定性に欠ける。“D+C” は安定して良い性能となっている。調波構造も考慮した “D+C+Ha” は、更に性能が良くなり、ほぼ “Optimal” に近づく。

提案手法の効果を示すため、9 組目の状況を詳しく見る。図 4 に各周波数での SN 比を示す。“C” では、広い範囲 (450 ~ 1400 Hz) で間違っている。これは、相対的な基準を用いる相関による方法だけにに基づいているからである。“D+C” では、到来方向に基づく方法がその範囲のいくつかの permutation を決定したため、その間違いは修正されている。しかし、250 Hz 以下では到来方向が推定できずに間違いが起こっている。“D+C+Ha” では、調波構造を利用してこの範囲の permutation を解決して修正している。

### 参考文献

- [1] 澤田, 向井, 荒木, 牧野. 周波数領域ブラインド音源分離における permutation 問題の解法. 音講論集, pp. 541–542, Sep. 2002.
- [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. ICASSP 2000*, pp. 3140–3143, June 2000.
- [3] S. Ikeda and N. Murata. An approach to blind source separation of speech signals. In *Proc. ICANN '98*, pp. 761–766, Sep. 1998.