実環境における3音源以上のブラインド分離 *

○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

観測された混合信号のみから音源信号を分離・抽出するブラインド音源分離 (BSS: blind source separation) の研究が盛んに行われている．しかし，残響を伴う実環境での混合を考えた場合，これまでの研究のほとんどは2音源の分離に関するものであり，現実的な状況で3音源以上の分離に成功した例を見つけることは難しい．我々は最近，周波数領域での手法により，実環境での3音源あるいは4音源の分離に成功したので，本稿でこれを報告する．特に，3音源の場合はリアルタイム処理が可能であり，音源の位置が変化しても，その変化に追従して数秒後には元の分離性能を回復できる．

2 畳み込み混合に対する2つの手法

N 個の音源 $s_p(t)$ が実環境で混合され，M 個のマイクで観測信号 $x_q(t) = \sum_{p=1}^{N} \sum_k h_{qp}(k) s_p(t-k)$ 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．このような畳み込み混合に対しては，FIR フィルタ $w_{rq}(k)$ を用いて分離信号 $y_r(t) = \sum_{q=1}^{M} \sum_{k=0}^{L-1} w_{rq}(k) x_q(t-k)$ を得ることが一般的である．音源の位置や無音区間を知らずにフィルタ係数 $w_{rq}(k)$ を求めるため，我々は，独立成分分析 (ICA: independent component analysis) [1, 2] を用いる．畳み込み混合に ICA を適用する手法は，大きく2つに分類できる．一方は畳み込み混合のまま ICA を適用する時間領域 BSS，他方は周波数毎に瞬時混合の ICA を個別に適用する周波数領域 BSS である．

時間領域 BSS の畳み込み混合 ICA では，本来求めるべき分離信号の独立性を評価しながらフィルタ係数を学習していくため，ICA の解に収束すれば高い分離性能が得られる．しかし，最終的な解から遠く離れた解を初期値として学習を始めると，収束までに非常に多くの時間を要する．これは，数千タップにも及ぶフィルタ係数が互いに依存し合っているためである．従って，何らかの手段で得た良い初期解から始めない限り，現実的な方法とは言えない．

周波数領域 BSS では周波数毎に瞬時混合の ICA を解く．個々の ICA の学習は，周波数間での関連が無いため非常に高速に収束する．その代償として，以下に述べる2つの問題が発生する．第一は，各周波数での分離信号が同じ音源に対応するように分離信号を並べ変えなければならないという permutation の問題である．第二は，離散周波数表現の巡回性により発生する問題である．これらの問題に適切に対処することで，我々は，実環境で3音源以上の分離に成功した．

*Blind separation of more than two sources in a real room environment, by H. Sawada, R. Mukai, S. Araki, S. Makino (NTT Communication Science Labs., NTT Corporation)

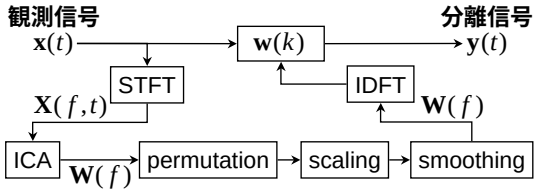

図 1: 周波数領域 BSS の処理の流れ

3 周波数領域 BSS

周波数領域 BSS の処理の流れ（図 1）と2つの問題への対処法を説明する．本方式では分離のための周波数特性 $W_{rq}(f)$ を逆フーリエ変換 (IDFT) してフィルタ $w_{rq}(k)$ を得る．そのためにまず，観測信号 $x_q(t)$ に L 点の短時間フーリエ変換 (STFT) を適用して周波数 f 毎の時間系列 $X_q(f,t)$ を求める（時刻 t はフレームシフト間隔で間引く）．次に，各周波数で瞬時混合の ICA: $\mathbf{Y}(f,t) = \mathbf{W}(f) \cdot \mathbf{X}(f,t)$ を解く．$\mathbf{W}(f)$ は要素が $W_{rq}(f)$ である $N \times M$ の分離行列であり，$\mathbf{X}(f,t) = [X_1(f,t), \ldots, X_M(f,t)]^T$ は観測信号，$\mathbf{Y}(f,t) = [Y_1(f,t), \ldots, Y_N(f,t)]^T$ は分離信号である．

ICA で得られる分離行列 $\mathbf{W}(f)$ の各行には，順序 (permutation) と大きさ (scaling) の任意性があるため，共に適切に調節する必要がある．第一の問題 permutation は，音源方向を推定した後に分離信号の相関を取ることで解決する [3]．文献 [3] で提案した方向推定方法が3音源以上に適応可能であるため，3音源以上の分離が現実的なものとなった．scaling は，MDP (Minimal Distortion Principle) [4] に従い，$\mathbf{W}(f) \leftarrow \mathrm{diag}[\mathbf{W}(f)^{-1}] \cdot \mathbf{W}(f)$ という操作により解決する．

第二の問題は，離散的周波数表現の巡回性（L 点でサンプリングされた周波数特性が L の周期を持つ時間信号を表現すること）による影響である．分離のために必要なフィルタ $w_{rq}(k)$ の長さが L に収まる場合は問題にならないが，L を超える場合は分離フィルタが別の周期と重なりを持ってしまう．2音源分離では必要なフィルタの長さが短いため問題にならなかったが，3音源以上の分離では必要なフィルタが長くなり，図3の上段に示すような現象が起こる．

この巡回性の問題に対し我々は，IDFT 後に得られるフィルタ $w_{rq}(k)$ の両端が 0 に収束するように周波数特性を平滑化 (smoothing) する．例えば，ハニング窓をフィルタに掛けて両端を 0 に収束させることは，周波数特性を $\mathbf{W}(f) \leftarrow [\mathbf{W}(f-\Delta f) + 2\mathbf{W}(f) + \mathbf{W}(f+\Delta f)]/4$ と smoothing することに相当する．しかし，単に smoothing を行うと，ICA で求めた分離のための周波数特性が変化するため，分離性能が劣化する可能性がある．そこで，smoothing 後の分離行列と ICA の解の誤差が最小化されるように，前もって scaling を調整する．詳しくは [5] を参照されたい．

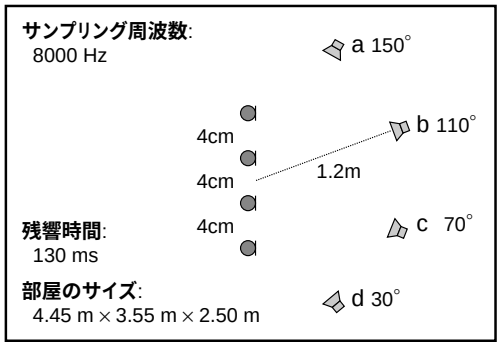

図 2: 部屋の特性とスピーカ／マイクの配置

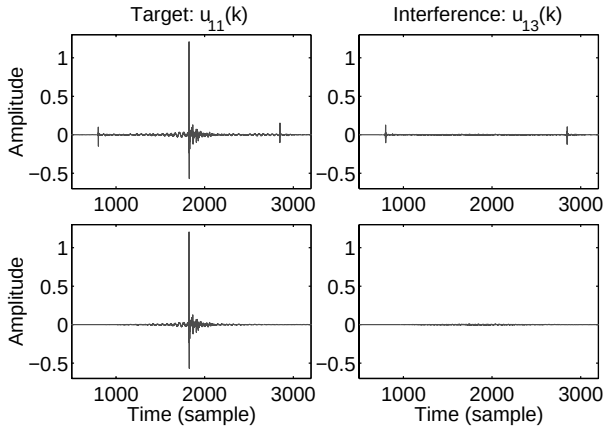

図 3: 3 音源分離における離散周波数表現の巡回性による影響（上段）と smoothing の効果（下段）

4 実験結果および考察

図 2 に示す条件で音源分離の実験を行った．図 3 は，3 音源の場合の音源 $s_p(t)$ から分離信号 $y_r(t)$ へのインパルス応答 $u_{rp}(k) = \sum_{q=1}^{M} \sum_{\tau=0}^{L-1} w_{rq}(\tau) h_{qp}(k-\tau)$ である．左側が目的音の抽出 $u_{11}(k)$，右側が干渉音の抑圧 $u_{13}(k)$ に対応する．巡回性の影響に対して何も処理を施さないと上段のように目的音の歪みおよび分離性能の劣化を引き起こすが，smoothing を行うことで下段のようにその影響が除去される．

音源数 2 から 4 の場合の 7 秒の音声に対するバッチ処理の結果を表 1 にまとめる．分離性能 SIR (Signal-to-Interference Ratio) は，出力信号を $y_r(t) = tar_r(t) + int_r(t)$ と分割し，それらのパワー比として計算した．ここで，$tar_r(t) = \sum_k u_{rr}(k) s_r(t-k)$ は目的音成分，$int_r(t) = \sum_{p \neq r} \sum_k u_{rp}(k) s_p(t-k)$ は干渉音成分である．次に，SDR (Signal-to-Distortion Ratio) は，目的音成分を $tar_r(t) = \alpha_r \cdot ref_r(t) + e_r(t)$ と参照音 $ref_r(t)$ の定数倍と歪み $e_r(t)$ に分割し，それらのパワー比として計算した．ここで，α_r は歪み $e_r(t)$ を最小にする定数であり，参照音としては MDP [4] に基づきマイク r での音源 r 成分 $ref_r(t) = \sum_k h_{rr}(k) s_r(t-k)$ を選んだ．表 1 により，smoothing を施すことで，SIR, SDR 共に改善されていることがわかる．また，分離フィルタ長としては $L = 2048$，ICA アルゴリズムとしては FastICA [1] で得た解をループ 50 回の Infomax + Natural gradient 型の ICA [2] でさらに改善するものを用いた．現実的な処理時間で高い性能の音源分離が行えていることが分かる．

表 1: バッチ処理の結果

音源数 / 位置	2 / a b		3 / a b d		4 / a b c d	
smoothing	なし	あり	なし	あり	なし	あり
平均 SIR (dB)	19.3	20.3	13.7	16.9	9.3	13.2
平均 SDR (dB)	18.0	19.3	13.9	15.7	10.8	11.3
実行時間 (s)	9.9		18.7		28.3	

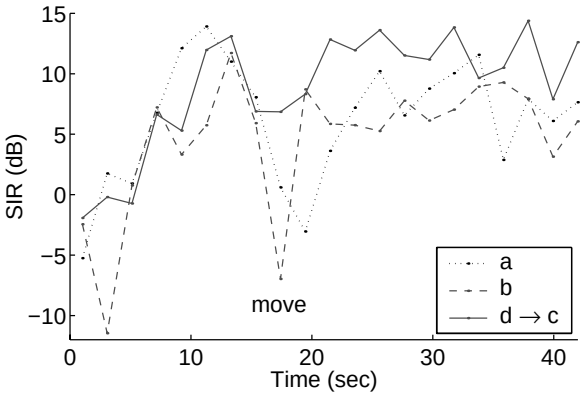

図 4: 音源移動に対するリアルタイム処理の分離性能

前回我々は 2 音源に対するリアルタイム処理 [6] を発表したが，分離フィルタ長を $L = 1024$，ループ回数を 35 とすることで，3 音源に対してもリアルタイム処理が可能となった．図 4 にその結果を示す．観測信号を 2 秒毎に区切って BSS の処理を行っているので，音源の移動に対しても追従が可能である．この例では，FastICA は用いずに Natural gradient だけで 2 秒毎に分離フィルタを更新していき，約 15 秒で位置 d の音源を位置 c に移動させた．数秒で以前の分離性能にまでほぼ回復していることが分かる．また，移動させた音源自身の分離性能がそれほど劣化しないことは，2 音源の場合 [6] と同様の現象として観測された．

5 まとめ

周波数領域 BSS により，実環境で 3 音源以上の分離に成功した．技術のポイントは，3 音源以上の方向推定が可能になったことにより permutation の問題が解決しやすくなったこと [3] にある．さらに，3 音源以上の分離により，離散的周波数表現の巡回性による影響という新たな問題が認識された．我々は周波数特性の smoothing によりこの問題を解決し [5]，SIR および SDR の改善を達成した．

参考文献

[1] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis.* John Wiley & Sons, 2001.

[2] S. Haykin, editor. *Unsupervised Adaptive Filtering (Volume I: Blind Source Separation).* John Wiley & Sons, 2000.

[3] 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 問題の頑健な解法. 音講論集, pp. 777–778, Mar. 2003.

[4] K. Matsuoka and S. Nakashima. Minimal distortion principle for blind source separation. In *Proc. ICA 2001*, pp. 722–727, Dec. 2001.

[5] H. Sawada et al. Spectral smoothing for frequency-domain blind source separation. In *Proc. IWAENC 2003*, Sep. 2003.

[6] 向井, 澤田, 荒木, 牧野. 移動音源の低遅延実時間ブラインド分離. 音講論集, pp. 779–780, Mar. 2003.