3—7—3

実環境におけるブラインド音源分離と残響除去性能に関する検討*

○向井 良　　荒木 章子　　牧野 昭二
(NTT コミュニケーション科学基礎研究所)

1 はじめに

音源信号の特性や音響系に関する情報を用いずに，観測された混合信号のみから複数の音源信号を分離・抽出する処理をブラインド音源分離 (Blind Source Separation, BSS) といい，信号の独立性を仮定する独立成分分析 (Independent Component Analysis, ICA) をはじめ多くの手法が提案されている [1]．瞬時混合した音源信号に対して BSS はかなり有効であり，また実環境における混合信号に適用した例も報告されているが [2]，一方で残響の多い環境においては十分な性能が得られていないことも指摘されている [3]．

本稿では，残響の多い環境において観測された信号を ICA による周波数領域 BSS を用いて分離した場合の分離性能，残響除去性能について，(1) 目的音の直接音，(2) 目的音の残響音，(3) 妨害音の直接音と残響音，のそれぞれに着目して評価し，妨害音については，直接音だけでなく初期反射音や残響音まで含めてある程度は除去できていること，目的音の残響音はあまり除去できていないことを示す．

2 評価手法

音源信号を $s_i(1 \le i \le N)$，マイク j で観測される信号を $x_j(1 \le j \le M)$，分離信号 $y_i(1 \le i \le N)$ とすると，

$$x_j = \sum_{i=1}^{N} h_{ji} * s_i, \quad y_i = \sum_{j=1}^{M} w_{ij} * x_j \qquad (1)$$

となる．ここで，h_{ji} は音源 i からマイク j へのインパルス応答，w_{ij} は BSS によって計算された分離システムを時間領域の FIR フィルタとして見たときの係数，$*$ は畳み込みを表す．

本稿では簡単のために，$N = M = 2$ とする (図 1)．なお，BSS では信号の入れ替わりを許すが，以下の議論では，s_1 は y_1 に s_2 は y_2 に分離されるものとする．

2.1 残響および妨害音の抑圧量

s_1 に単位インパルス $\delta(t)$，s_2 に 0 を入力したのときの観測信号 x_1 を x_{1s1}[図 2(a)]，分離信号 y_1 を y_{1s1}[図 2(b)]，逆に，$s_1 = 0, s_2 = \delta(t)$ のときの x_1，y_1 を x_{1s2}，y_{1s2}[図 2(c), x_{1s2} は省略] と呼ぶことにする．s_1 を目的音，s_2 を妨害音としたとき，y_{1s1} は抽出された目的音の直接音および残響音，y_{1s2} は妨害音の消し残りと考えることができ，これらは式 (1) から，

$$x_{1s1} = h_{11}, \quad x_{1s2} = h_{12} \qquad (2)$$

$$y_{1s1} = w_{11} * h_{11} + w_{12} * h_{21} \qquad (3)$$

$$y_{1s2} = w_{11} * h_{12} + w_{12} * h_{22} \qquad (4)$$

*Blind source separation and removal of reverberation in the real environment, Ryo Mukai, Shoko Araki and Shoji Makino (NTT Communication Science Laboratories)

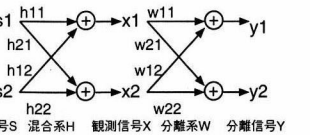

図 1: 系の構成

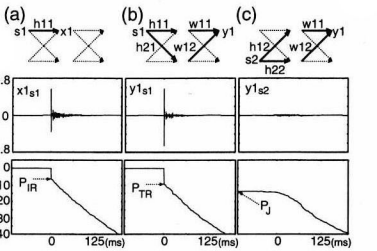

図 2: 各信号の波形と大きさ

で計算できる．x_{1s1} および x_{1s2} の大きさが 0dB になるように H を正規化し，x_{1s1} の残響音の大きさを P_{IR}，y_{1s1} の残響音の大きさを P_{TR}，y_{1s2} の大きさを P_J とする．これらを用いて，目的音の残響抑圧量 R_T と妨害音の抑圧量 R_J を，

$$R_T = -(P_{TR} - P_{IR}) \qquad (5)$$

$$R_J = -P_J \qquad (6)$$

と定義する．

3 周波数領域 BSS

実環境における遅れや畳み込みの影響を受けながらの混合は，周波数領域での瞬時混合に変換できるため，周波数領域での BSS が有効である．式 (1) を周波数領域で表現すると，

$$\mathrm{X}(\omega, m) = \mathrm{H}(\omega)\mathrm{S}(\omega, m) \qquad (7)$$

$$\mathrm{Y}(\omega, m) = \mathrm{W}(\omega)\mathrm{X}(\omega, m) \qquad (8)$$

となる．今回は評価対象のアルゴリズムとして，ICA による周波数領域 BSS アルゴリズム [4] を使用し，以下の学習規則によって，周波数ごと独立に Kullback-Leibler divergence を最小にするような W を推定する．

$$W_{i+1} = W_i + \eta[\mathrm{diag}(\langle\Phi(Y)Y^H\rangle) - \langle\Phi(Y)Y^H\rangle]W_i \qquad (9)$$

ここで，$\langle\cdot\rangle$ は期待値演算，i は更新回数，η はステップサイズであり，Φ として，

$$\Phi(Y)=\frac{1}{1+e^{-\mathrm{Re}(Y)}}+j\frac{1}{1+e^{-\mathrm{Im}(Y)}} \qquad (10)$$

を用いた．なお，今回の実験では周波数ごとの信号の入れ替わりの影響は小さかったため，その解決は行わず，式 (9) によって得られた W について，出力の大きさを正規化するためのスケーリングのみ行っている．

4 実験

周波数領域 BSS では，フレーム長によって分離性能が変化することが確認されている．このとき，何が分離され，何が雑音として残っているのかを調べるための実験を行った．

4.1 実験方法

図 3 に示した環境でインパルスを実測し，これを h_{ji} として用いる．残響時間は 300ms，直接音の寄与分は，h_{11} と h_{21} が 6.6dB，h_{12} と h_{22} が 5.7dB であった．BSS の入力としては，ASJ 研究用音声コーパス中の 2 文 (サンプリング周波数 8kHz，長さ約 8 秒) に h_{ji} を計算機上で畳み込んだものを与えた．

DFT のフレーム長を T，フレームシフトを S とし，T を 32 から 4096 まで変化させ，S は $T/2$ および $T/32$(2 倍および 32 倍のオーバーサンプリングに相当) とした．式 (9) による更新回数は 100 回としたが，$S=T/2$ の場合に過学習による性能の悪化が確認されたため，T =1024，2048，4096 のときにそれぞれ 70 回，30 回，20 回で学習を止めている．また，比較対象として，音源の位置を陽に与えた死角制御型ビームフォーマ (Null BeamFormer, NBF) を用いた評価も行った．

4.2 実験結果および考察

実験結果を図 4 に示す．横軸にフレーム長，縦軸に目的音の残響の抑圧量 (s_1 を目的音としたときの R_{T1}，s_2 を目的音としたときの R_{T2}) と妨害音の抑圧量 (R_{J1}, R_{J2}) を示している．

(a),(b) はそれぞれ $S=T/2$，$T/32$ として上記の ICA を用いた BSS による結果，(c) は NBF による結果である．

まず R_J について見てみると，$T\leq 128$ では ICA も NBF と同程度の抑圧量であるが，$256\leq T\leq 2048$ では T を大きくするに従い性能が向上している．$T=2048$，$S=T/32$ の場合には $R_{J1}=16.6$dB，$R_{J2}=19.5$dB と，直接音寄与分の 5.7dB，6.6dB を大きく上まわっており，直接音だけでなく初期反射音や残響音まで抑圧できていることがわかる．

一方，R_T については，フレーム長 T，フレームシフト S に関わらずほぼ一定の値であり，ICA の方が NBF より若干良い程度である．目的音に起因する初期反射音や残響音の除去はほとんど行われていないことがわかる．

以上のことから，W は混合系 H の逆システムを近似するものではなく，妨害音のみを除去するフィルタを形成していると言える．

5 まとめ

ICA による周波数領域 BSS の性能がフレーム長によって変化する要因，および NBF 以上の性能を持つ要因として，妨害音の初期反射音や残響音の除去効果が影響していることを実験によって確認した．また，目的音の残響音の除去に関しては NBF を若干上まわる程度であることがわかった．

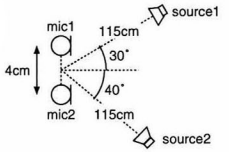

図 3: インパルス測定環境の配置

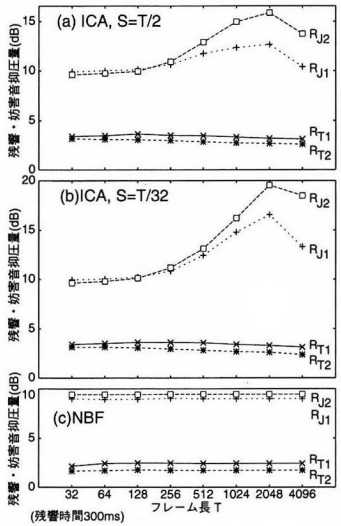

図 4: フレーム長と抑圧量

参考文献

[1] A.J.Bell and T.J.Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[2] T.W.Lee, A.J.Bell, and R.Orglmeister, "Blind source separation of real world signals," *Neural Networks*, vol. 4, pp. 2129–2134, 1997.

[3] M.Z.Ikram and D.R.Morgan, "Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment," in *ICASSP'00*, 2000, pp. 1041–1044.

[4] S.Kurita, H.Saruwatari, S.Kajita, K.Takeda, and F.Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in *ICASSP'00*, 2000, pp. 3140–3143.