周波数領域畳込みモデルに基づく高残響下での音源分離*

古屋 武志[†]・石橋 孝昭[‡]・白土 浩[†]・五反田 博[†]

Blind Source Deconvolution Based on Frequency Domain Convolution Model Under Highly Reverberant Environments*

Takeshi KOYA[†], Takaaki ISHIBASHI[‡], Hiroshi SHIRATSUCHI[†] and Hiromu GOTANDA[†]

In order to solve the blind source deconvolution, a convolved mixing process in the time domain is often transformed into an instantaneous mixing model in the frequency domain. However, the model is only an approximated one and thus does not work effectively under highly reverberant environments. By dividing the impulse response properly, Servière has precisely transformed the time-domain convolved mixture to a frequency-domain convolved mixture and has proposed a new FDICA approach available under high reverberation. In the approach, however, the permutation and scaling problems are unresolved as well as the distortion due to whitening. In the present paper, an improved approach without such problems is proposed and is confirmed to be valid in a highly reverberant real environment.

1. まえがき

近年, 独立成分分析 (ICA: Independent Component Analysis) に基づくブラインド信号分離 (BSS: Blind Source Separation) が注目を集めている [1–4]. ICA は, 原信号に統計的独立性を仮定して, 混合信号から原信号を分離する手法で, その応用技術の開発が様々な分野で試みられている. さらに, 実環境下のように, 原信号が畳み込んで観測されるとき BSS はブラインドソースディコンボリューション (BSD: Blind Source Deconvolution) とよばれ, 現在盛んに研究されている.

BSD には, 時間領域で分離を行う時間領域 ICA (TDICA: Time Domain ICA) と周波数領域で分離を行う周波数領域 ICA (FDICA: Frequency Domain ICA) の二つのアプローチがある. 一般に, TDICA は分離フィ

ルタの次数が大きくなるにつれて計算負荷が増大しアルゴリズムの収束性も悪くなることから, FDICA がよく用いられている. FDICA は収束性も良く, 低残響下では良好に機能するが, 高残響下では分離能力が著しく低下する [5, 6]. これは, 時間領域での畳込み混合過程を周波数領域に変換するとき, 離散時間フーリエ変換の窓幅を超えた残響部分を無視して, 初期反射部分のみを周波数領域の瞬時混合モデルとして近似したことによる [6].

残響問題について, Servière は, インパルス応答系列を複数のブロックに分けて個々のブロックを低次の狭帯域フィルタに分担させることにより, 時間領域での畳込み混合過程を周波数領域での畳込み混合モデル (以降, 分割畳込み混合モデルと略記) として厳密に定式化した. そして, これに対する分離モデル (以降, 分割畳込み分離モデルと略記) に対して TDICA アルゴリズムの概念を適用した新たな ICA (以降, ConvFDICA と略記) 法を提案し [6], 高残響下での音源分離の道を拓いた. しかし, ConvFDICA は, 周波数領域に変換した信号に対して TDICA でよく用いられる自然勾配 (NG: Natural Gradient) 法 [8] を単に適用しただけなので, TDICA の性質を継承したことによる白色化歪みの問題に加えて, FDICA アプローチでは避けられない成分置換やスケールの不定性の問題を未解決のまま抱えている.

残響問題に対して Araki らはサブバンドの考えを導入し, インパルス応答系列を低次の狭帯域フィルタに分担

* 原稿受付 2008 年 6 月 25 日

[†] 近畿大学 産業理工学部 電気通信工学科 Department of Electrical and Communication Engineering, Kinki University; Kayanomori, Iizuka-shi, Fukuoka 820–8555, JAPAN[‡] 熊本電波工業高等専門学校 情報通信工学科 Department of Information and Communication Engineering, Kumamoto National College of Technology; Suya, Koshi-shi, Kumamoto 861–1102, JAPAN**Key Words:** high riverberation-resistant BSD, frequency-domain ICA, segmentally convoluted mixing model, permutation, scaling, whitening distortion.

させる方法を提案している．その中で，成分置換がICAの初期値に依存することを指摘し，それを利用した成分置換解消法を提唱している[5]．しかし，TDICAにより分離を進めていく過程で起こるスケールの不定性については後処理が必要で，TDICAに固有の白色化歪みの問題は未解消である．

一方，TDICAにより起こる白色化歪みについて，Matsuokaらは最小歪み原理(MDP: Minimal Distortion Principle)に基づくTDICAを提案している[9]．これにより，TDICAに固有の白色化歪みは軽減できるが，分離フィルタ長が長くなると，収束性が悪くなる問題が生じるため，高残響下での使用は困難である．

実用的な使用を考えた場合，高残響下でも有効に機能するICA手法の開発が必要であるが，これまで決め手となる方法を実現するまでには至っていない．

本論文では，高残響下での音源分離を目的に，2.でServièreによる分割畳込みモデルとそのモデルに対するConvFDICAアルゴリズムを述べる．3.では，成分置換の初期値依存性を利用したArakiらのアプローチに準拠して，同様のことがConvFDICAの場合，簡略化した方法で実現できることを明らかにする．また，Matsuokaらの時間領域MDPを分割畳込みモデルの枠組みに拡張して得られる周波数領域MDPを導き，周波数領域MDPをConvFDICAに適用して，スケールの不定性や白色化歪みのない新たなConvFDICA(以降，改良型ConvFDICAと略記)法を提案する．4.では，提案法の有効性を高残響下での実験により検証し，マイク入力時の音声がほとんど歪みなく復元できることを示す．最後に5.で以上をまとめる．

2. 分割畳込み混合とConvFDICA

ここでは，これまでFDICAで用いられてきた瞬時混合モデルについて，残響が強くなると時間領域と周波数領域とで混合過程の整合性が悪くなることを述べる．そして，この不整合の問題が分割畳込みモデルを導入することで解決されることを説明した後，分割畳込みモデルに対するConvFDICAを概説する．

2.1 畳込み混合過程の周波数領域表現

混合信号 $x_j(t)$ は，音源 $s_i(t)$ がインパルス応答で畳み込まれてマイクで観測されたもので，時間領域では

$$x_j(t) = \sum_{i=1}^{L_s} \sum_{\tau=0}^{T-1} a_{ji}(\tau) s_i(t-\tau) \quad (1)$$

のように線形畳込み混合過程として表現される．ここに， i は音源の番号， j はマイクの番号， L_s は音源の数， $a_{ji}(\tau)$ は音源 i からマイク j までのインパルス応答， $\tau=0, 1, \dots, T-1$ は遅れ時間， T はインパルス応答長である．

線形畳込み混合過程は，周波数領域に変換すると

$$x_j(\omega_n, k) \approx \sum_{i=1}^{L_s} a_{ji}(\omega_n) s_i(\omega_n, k) \quad (2)$$

のように伝達関数 $a_{ji}(\omega_n)$ と原信号 $s_i(\omega_n, k)$ の積として表され，瞬時混合モデルとして近似できる．ここに， $x_j(\omega_n, k)$ ， $s_i(\omega_n, k)$ ， $a_{ji}(\omega_n)$ はそれぞれ $x_j(t)$ ， $s_i(t)$ ， $a_{ji}(\tau)$ を離散フーリエ変換(DFT)した値， ω_n は正規化角周波数， k はフレーム番号である．

逆に，周波数領域での混合モデル，つまり，(2)式の右边を逆離散フーリエ変換(IDFT)して時間領域に戻すと，巡回畳込み混合過程となる．巡回畳込み混合過程は，インパルス応答長が $T \ll N$ と窓幅 N より極めて短いとき，(1)式の線形畳込み混合過程とほぼ等しくなるが，それ以外では等しくならない．すなわち，残響がほとんどない場合，インパルス応答長 T は極めて小さいことから，通常のように窓幅 N を数十ms程度に設定すれば， $T \ll N$ なる条件は自然に満たされるので，(2)式の両辺は等価とみなせる．したがって，残響がほとんどない場合，時間領域での畳込み混合過程を周波数領域で瞬時混合過程と近似すれば，通常のFDICAで音源を良好に分離できる．

しかし，残響時間が長くなると，インパルス応答長 T が $T > N$ と窓幅 N より長くなるため，二つの誤差が生じる．一つは，(2)式の右边を時間領域に戻したものが，線形畳込み混合過程とはならず，巡回畳込み混合過程となることに起因する誤差である．もう一つは， $T > N$ となって必ずしもインパルス応答のすべてを窓内に収容できない場合，インパルス応答をDFTした値 $a_{ji}(\omega_n)$ が正確に求められないことに起因する誤差である．

以上のことから，高残響下で良好な音源分離を期待するには， $T \ll N$ を満たすように窓幅 N をインパルス応答長 T よりかなり大きめに設定すればよいことになる．しかし，その場合，ICAを行ううえで統計的に意味ある十分な個数のデータが得られないことになる．

2.2 分割畳込み混合モデル

インパルス応答が窓内に収まらない場合，上述のように時間領域での畳込みは周波数領域で正確に表現できない．そのため，FDICAを高残響下で適用しても良好な分離結果は得られない．そこで，Servièreは，時間領域でのインパルス応答系列を複数のブロックに分割し，各ブロックをDFTして狭帯域のインパルス応答として周波数領域で表現することにより，時間領域での畳込み混合過程を周波数領域での畳込み混合モデル(分割畳込み混合モデル)として厳密に定式化した[6]．この分割畳込み混合モデルは，各周波数 ω_n での狭帯域インパルス応答長を M_0 と仮定した場合

$$\begin{aligned} h(\omega_n) \otimes x_j(\omega_n, k) \\ = h(\omega_n) \otimes \sum_{i=1}^{L_s} \sum_{m=0}^{M_0-1} s_i(\omega_n, k-m) a_{ji}(\omega_n, m) \end{aligned} \quad (3)$$

と表現される．ここに， $h(\omega_n)$ ($n=0,1,\dots,2N-1$) は前半 N 個が 0 の窓関数 $h=[0,\dots,0,h_0,h_1,\dots,h_{N-1}]^T$ を DFT したもので， $\otimes$ は巡回畳込み操作を表す．窓関数 $h(\omega_n)$ は周波数領域でのスペクトルの漏れを防ぐためであるが，理論解析上は無視できる．

また，その場合，分割畳込み混合モデルは

$$x(\omega_n, k) = \sum_{m=0}^{M_0-1} A(\omega_n, m) s(\omega_n, k-m) \quad (4)$$

のようにベクトル形式で表される．ここに， $x(\omega_n, k)=[x_1(\omega_n, k), \dots, x_{L_r}(\omega_n, k)]^T$ は混合信号， L_r はマイクの数， $s(\omega_n, k)=[s_1(\omega_n, k), \dots, s_{L_s}(\omega_n, k)]^T$ は原信号， $A(\omega_n, m)$ は以下で定義される狭帯域混合行列である．

$$A(\omega_n, m) = \begin{bmatrix} a_{11}(\omega_n, m) & \cdots & a_{1L_s}(\omega_n, m) \\ \vdots & \ddots & \vdots \\ a_{L_r1}(\omega_n, m) & \cdots & a_{L_rL_s}(\omega_n, m) \end{bmatrix}$$

以上のことから，各周波数で狭帯域インパルス応答長 M_0 と窓幅 N の積が時間領域インパルス応答長と等しく， $T=M_0N$ となるように M_0 を設定すれば，時間領域での畳込み混合過程が，分割畳込み混合モデルとして正確にモデル化できることがわかる．

2.3 ConvFDICA アルゴリズム

分割畳込み混合モデルに対して，Servière は分割畳込み分離モデルを以下のように定義した．

$$y(\omega_n, k) = W(\omega_n, z) x(\omega_n, k) \quad (5)$$

ここに， $y(\omega_n, k)=[y_1(\omega_n, k), \dots, y_{L_s}(\omega_n, k)]^T$ は分離信号， $W(\omega_n, z)$ は

$$W(\omega_n, z) = \sum_{m=0}^{M-1} W(\omega_n, m) z^{-m} \quad (6)$$

で定義される狭帯域分離フィルタ， $M(>M_0)$ は狭帯域分離フィルタのタップ長， $W(\omega_n, m)$ は下の行列である．

$$W(\omega_n, m) = \begin{bmatrix} w_{11}(\omega_n, m) & \cdots & w_{1L_r}(\omega_n, m) \\ \vdots & \ddots & \vdots \\ w_{L_s1}(\omega_n, m) & \cdots & w_{L_sL_r}(\omega_n, m) \end{bmatrix}$$

この分割畳込み分離モデルに TDICA の概念を適用して，Servière は ConvFDICA アルゴリズムを

$$\Delta W(\omega_n, m) = \eta \{ I \delta(m) - \phi(y(\omega_n, k)) y^H(\omega_n, k-m) \} W(\omega_n, z) \quad (7)$$

と導いた [6]．ここに， $\phi(y(\omega_n, k))=[\phi_1(y_1(\omega_n, k)), \dots, \phi_{L_s}(y_{L_s}(\omega_n, k))]^T$ で，その要素は $\phi_i(y_i(\omega_n, k))=\tanh\{\text{Re}(y_i(\omega_n, k))\}+j\tanh\{\text{Im}(y_i(\omega_n, k))\}$ と定義される非線形関数， H は共役転置， $\Delta W(\omega_n, m)$ は

$$\Delta W(\omega_n, m) = \begin{bmatrix} \Delta w_{11}(\omega_n, m) & \cdots & \Delta w_{1L_r}(\omega_n, m) \\ \vdots & \ddots & \vdots \\ \Delta w_{L_s1}(\omega_n, m) & \cdots & \Delta w_{L_sL_r}(\omega_n, m) \end{bmatrix}$$

のように定義される行列， η は探索ステップ幅である．

3. 改良型 ConvFDICA アプローチ

前節では，分割畳込み混合モデルと，それに対する分離モデル，さらに，その狭帯域分離フィルタを求める ConvFDICA を紹介した．この ConvFDICA はインパルス応答の長い高残響下であっても機能する可能性を秘めている．しかし，ConvFDICA は，TDICA の性質を継承しているため，白色化歪みの問題が未解決である．また，FDICA アプローチでは不可避の成分置換やスケールの不定性の問題に対する Servière の施策 [7] は効果的とはいえず，ConvFDICA のもつ高い潜在能力を活かすには至っていない．本節では，これらの問題に対する解決指針を示して，改良型 ConvFDICA を提案する．

3.1 成分置換の是正と収束性の改善について

ICA 固有のスケールの不定性と成分置換の問題は，周波数領域でのアプローチにとって不可欠の解決課題である．つまり，各周波数で振幅と位相のスケールがそろわず，成分が置換されたままの狭帯域信号を時間領域に戻すだけでは，原信号は復元されない．

Araki らは，成分置換が TDICA アルゴリズムの初期値に依存して決まることを利用して，成分置換を是正する方法を提唱している [5]．これは，ICA の分離荷重が近似的に妨害音に対して死角を向ける性質があることを利用した方法で，成分置換の是正を行うサブバンド分析部と TDICA アルゴリズムから成るサブバンド合成部に分かれる．まず，サブバンド分析部では，Null ビームフォームにより音源の到来方向 $\theta(\omega_n)$ を推定し，その方向をもとに妨害音に対して死角を形成する荷重 $\hat{W}(\omega_n)$ を求める．つぎに，サブバンド合成部では，その荷重 $\hat{W}(\omega_n)$ を基底帯域（ベースバンド）に移動し， $\hat{W}(\omega_n)$ とその共役 $\hat{W}^*(\omega_n)$ の二つの複素スペクトルを時間領域に戻して，得られる値を初期値として TDICA アルゴリズムを駆動させる．これにより，成分置換のない解が得られると同時に収束性が向上することが示されている [5]．

上述のことを ConvFDICA に焼き直した場合，以下のような成分置換解消手順が得られる．まず，通常の FDICA により成分置換を是正した分離荷重 $\hat{W}(\omega_n)$ を求める．このとき，成分置換問題は，たとえば，Null ビーム法 [5]，伝達関数法 [10] 等を適用する．この $\hat{W}(\omega_n)$ を (6) 式の初期値とする．この処理によって，成分置換のない ConvFDICA が実現できる．この手順では周波数領域で処理を行うため，サブバンド合成が不要である．

なお，ConvFDICA に与える初期値は Null ビーム法で求めた荷重でもよいが，Null ビーム法は干渉音の直接波しか除去できず，反射波は除去できない．これに対

して, FDICA はある程度の残響成分は除去できるので, FDICA により求めた分離荷重を ConvFDICA の初期値として与える方がより効果的と考えられる.

3.2 白色化歪みとスケールの不定性の解法指針

本節では, ConvFDICA のスケールの不定性と白色化歪みの問題を検討し, その原因を究明する. そして, ConvFDICA アルゴリズムの白色化歪みやスケールの不定性の問題に対する解法指針を示す.

まず, ConvFDICA の (7) 式の対角項は

$$\phi_i(y_i(\omega_n, k))y_i^*(\omega_n, k-m) = \delta(m) \quad (8)$$

なる制約がある. これは, $\phi_i(y_i(\omega_n, k)) \approx y_i(\omega_n, k)$ と近似して, フレーム番号 k に対して平均すると

$$E[y_i(\omega_n, k)y_i^*(\omega_n, k-m)] = \delta(m) \quad (9)$$

となって, 近似的に $y_i(\omega_n, k)$ の自己相関関数を表していると考えられる. ここに, k は時間フレーム, m は時間フレーム k における遅延 (つまり, 狭帯域信号としての遅延), $*$ は共役, $E[\cdot]$ は期待値を表す.

時間領域における音声信号 $s_i(t)$ のように信号自身が相関をもつ場合, その自己相関関数 $r_{s_i}(\tau) = E[s_i(t)s_i(t-\tau)]$ は $\tau \neq 0$ でも $r_{s_i}(\tau) \neq 0$ と相関をもつ. したがって, 周波数領域における音声信号 $s_i(\omega_n, k)$ の自己相関関数 $r_{s_i}(\omega_n, m)$ も相関をもつ. 以上のことから, ConvFDICA では分離信号が強制的に白色化されることに起因して, 白色化歪みやスケールの不定性問題が起こる.

つぎに, 上述の検討結果を基に ConvFDICA の改良指針について検討する. (7) 式の $\Delta W(\omega_n, m)$ は非対角成分と対角成分に分けて, 以下のように表現できる.

$$\Delta W(\omega_n, m) = \eta\{\Delta W_1(\omega_n, m) + \xi\Delta W_2(\omega_n, m)\} \quad (10)$$

ここに, $\Delta W_1(\omega_n, m)$ と $\Delta W_2(\omega_n, m)$ は, それぞれ

$$\begin{aligned} \Delta W_1(\omega_n, m) &= -\text{offdiag}\{\phi(y(\omega_n, k)) \\ &\quad y^H(\omega_n, k-m)\}W(\omega_n, z) \\ \Delta W_2(\omega_n, m) &= \text{diag}\{\delta(m) \\ &\quad -\phi(y(\omega_n, k))y^H(\omega_n, k-m)\}W(\omega_n, z) \end{aligned}$$

で, ξ は $\Delta W_1(\omega_n, m)$ に対する $\Delta W_2(\omega_n, m)$ の比率で, (7) 式では $\xi = 1$ である.

(10) 式の非対角成分 $\Delta W_1(\omega_n, m)$ は信号の空間分離に寄与し, 対角成分 $\Delta W_2(\omega_n, m)$ はスケールに寄与していると解釈される. したがって, 白色化歪みを回避するには $\Delta W_2(\omega_n, m)$ を削除すればよいことになる. しかしその場合, スケールを制約する成分がなくなるので, 分離信号が無数の自由度をもつことになって, スケールが不定になると同時に収束が不安定になる.

以上のことから, (10) 式の対角成分をスケールの不定性が起きないように更新できれば, 白色化歪みの問題は解消されるという指針が得られる.

3.3 MDP の分割畳込みモデルへの適用

ここでは, Matsuoka らが提唱している MDP について, その概要を説明するとともに, MDP を TDICA に適用して, 白色化歪みとスケールの不定性を解消する.

MPD は窓幅を無限長としたフーリエ変換を前提に, 時間領域での混合信号と分離信号の平均二乗誤差 $E_t[\|y(t) - x(t)\|^2]$ が周波数領域での連続スペクトルの平均と等しくなることを利用して導かれたものである. そのため, 窓幅の短い短時間離散フーリエ変換を前提に定式化された分割畳込みモデルの場合, MDP をそのまま適用できる保証はない.

そこで, 時間領域の平均二乗誤差 $E_t[\|y(t) - x(t)\|^2]$ が周波数領域での混合信号と分離信号の短時間離散スペクトルの平均二乗誤差と一致することを示し, MDP が分割畳込みモデルに適用できることを保証する.

TDICA により分離を行った場合, 成分置換が解決されたとしても, 分離信号 $y(t)$ は原信号 $s(t)$ にスケールの不定性を伴って

$$y(t) = \sum_{\tau=0}^{T-1} D(\tau)s(t-\tau) = D(z)s(t) \quad (11)$$

と得られる. ここに, $y(t) = [y_1(t), y_2(t), \dots, y_{L_s}(t)]^T$, $s(t) = [s_1(t), s_2(t), \dots, s_{L_s}(t)]^T$, $D(z)$ はスケールの不定性を表す対角行列 $D(z)$ で

$$D(z) = \sum_{\tau=0}^{T-1} D(\tau)z^{-\tau} \quad (12)$$

と定義される. このとき, $D(z) = \text{diag}[d_1(z), d_2(z), \dots, d_{L_s}(z)]$ で, その要素 $d_i(z)$ はスケールの不定性による任意の係数となる. 一般に, 添字 i が異なれば $d_i(z)$ ($i = 1, 2, \dots, L_s$) も異なるため, スケールは不定となり, さらに白色化歪みの問題へと発展する.

Matsuoka らは, この白色化歪みを, 分離信号と混合信号の平均二乗誤差 $E_t[\|y(t) - x(t)\|^2]$ で評価し, これを連続スペクトルで表現して

$$\begin{aligned} E_t[\|y(t) - x(t)\|^2] &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr}[\{D(\omega_n) - A(\omega_n)\} \\ &\quad \Phi_s(\omega_n)\{D(\omega_n) - A(\omega_n)\}^H] d\omega_n \end{aligned} \quad (13)$$

となる関係を導いている [9]. ここに, $D(\omega_n)$ は $D(z)$ の周波数領域表現, $A(\omega_n)$ は混合行列 $A(z)$ の周波数領域表現, $\Phi_s(\omega_n) = s(\omega_n)s^H(\omega_n)$, E_t は時間平均, $\text{tr}[\cdot]$ はトレースである. そして, この関係より, 白色化歪み, つまり, (13) 式の左辺が最小になるとき, つぎの関係式が成り立つことを示している.

$$D(\omega_n) = \text{diag} A(\omega_n) \quad (14)$$

この関係式は, 時間領域では $D(z) = \text{diag} A(z)$ と表現され, 分離信号 $y_i(t)$ が

$$y_i(t) = a_{ii}(z)s_i(t) \quad (15)$$

のように原信号 $s_i(t)$ を $a_{ii}(z)$ (混合行列 $A(z)$ の対角成分) で畳み込んだ形で復元されることを意味している。つまり、 i 番目の音源のみを駆動して i 番目のマイクで観測した信号と同一になることを意味している。そして、このことは、(15) 式を周波数領域に直して $y_i(\omega_n) = a_{ii}(\omega_n)s_i(\omega_n)$ と表現すれば、分割スペクトルが伝達関数 $a_{ii}(\omega_n)$ と音源 $s_i(\omega_n)$ の積で表されて [10]、スケールの不定性が解消されることと一致する。

以上のことから、時間領域では、(13) 式の左辺、つまり、 $E_t[\|y(t) - x(t)\|^2]$ を最小化することにより、スケールの不定性が解消されるとともに、白色化歪みが最小となることがわかる。しかし、(13) 式の右辺は窓幅を無限長とするフーリエ変換を前提とした連続スペクトルであるのに対して、ConvFDICA で必要となるスペクトルは、(7) 式に示すように、有限長の窓幅でフーリエ変換して得られる短時間離散スペクトルである。そのため、(14) 式と同様の関係が短時間離散スペクトルについてもそのまま成り立つとは限らない。そこで、(14) 式と同様の関係が短時間スペクトルのときでも成り立つことを述べる。

(13) 式の左辺を短時間離散スペクトルで表現すると

$$\begin{aligned} E_t[\|y(t) - x(t)\|^2] \\ = \frac{1}{N} \sum_{n=0}^{N-1} E_k[\|y(\omega_n, k) - x(\omega_n, k)\|^2] \end{aligned} \quad (16)$$

なる関係が得られる (付録参照)。ここに、 E_k は時間フレーム k に対する平均を表す。このとき、個々の周波数 ω_n で $E_k[\|y(\omega_n, k) - x(\omega_n, k)\|^2]$ を最小化すれば、(16) 式の左辺、つまり、白色化歪みは最小となることがわかる。また、個々の周波数 ω_n で $E_k[\|y(\omega_n, k) - x(\omega_n, k)\|^2]$ を最小化することにより、(14) 式と等価な関係式、つまり

$$D(\omega_n, z) = \text{diag} A(\omega_n, z) \quad (17)$$

が得られる。このことは、(17) 式は周波数領域でのスケールの不定性が解消されることを意味している。

3.4 改良型 ConvFDICA アルゴリズム

上述では、(16) 式の $E_k[\|y(\omega_n, k) - x(\omega_n, k)\|^2]$ を個々の周波数 ω_n で最小化することで、スケールの不定性と白色化歪みが解消できることを述べた。また、3.2 では、(10) 式の ConvFDICA アルゴリズムの対角成分 $\Delta W_2(\omega_n, m)$ にスケールの不定性がないようにすれば、白色化歪みの問題が解消されることを示した。ここでは、これらのことを考慮して、MDP を分割畳込みモデルに適用し、新たな ConvFDICA アルゴリズムを提案する。

(16) 式右辺の $E_k[\|y(\omega_n, k) - x(\omega_n, k)\|^2]$ の自然勾配を求めると、以下のように表現できる (付録参照)。

$$\Delta W_3(\omega_n, m) = -2\{y(\omega_n, k) - x(\omega_n, k)\}$$

$$\sum_{q=0}^{M-1} y^H(\omega_n, k-m+q) W(\omega_n, q) \quad (18)$$

そこで、(10) 式の $\Delta W_2(\omega_n, m)$ と $\Delta W_3(\omega_n, m)$ を入れ替えると、次のアルゴリズムが導かれる。これを改良型 ConvFDICA アルゴリズムとよぶ。

$$\begin{aligned} \Delta W(\omega_n, m) = \eta \left[-\text{offdiag} \left\{ \phi(y(\omega_n, k-M+1)) \right. \right. \\ \left. \left. \sum_{q=0}^{M-1} y^T(\omega_n, k-M+1-m+q) \right\} - \gamma \psi(\omega_n, k-M+1) \right. \\ \left. \sum_{q=0}^{M-1} y^T(\omega_n, k-M+1-m+q) \right] W(\omega_n, q) \end{aligned} \quad (19)$$

ここに、 $\gamma = 2\xi$ で、 $\psi(\omega_n, k)$ と $\psi_i(\omega_n, k)$ は、それぞれ、以下のように表現される。

$$\begin{aligned} \psi(\omega_n, k) &= [\psi_1(\omega_n, k), \dots, \psi_{L_s}(\omega_n, k)]^T \\ \psi_i(\omega_n, k) &= y_i(\omega_n, k) - x_i(\omega_n, k) \end{aligned}$$

音場ではよくあるケースであるが、混合過程が非最小位相系の場合、その分離過程 (逆システム) は非因果的フィルタとなる。そこで、時間領域では、時刻 0 より前の非因果的な部分が時刻 0 より後にくるように移動 (遅延) させて得られる因果的フィルタを非因果的フィルタの代わりに用いる。

これに対して、周波数領域では、理論上、時刻 0 よりも前 (システムに信号を入力する前) の非因果的な部分も取り扱うことができるので、非因果的フィルタを実現できる。そのため、周波数領域で MDP を適用した場合、混合過程が最小位相系であっても非最小位相系であっても、(20) 式のように、混合信号は分離信号 $y_i(\omega_n, k)$ と同一時刻の $x_i(\omega_n, k)$ が用いられるので、時間領域でのように遅れを考慮する必要はない。

4. 検証実験

4.1 実験条件

残響時間が 300ms の環境下で、Fig. 1 のように、二つの音源 ($L_s = 2$) が二つのマイク ($L_r = 2$) で観測される場合について検証実験を行った。混合信号は、新聞記事読み上げ音声コーパス [11] から選んだ男女各二名の発話音声 (8 秒) に、RWCP 実環境音声・音響データベース [12] の残響時間 300ms のインパルス応答を畳み込んで 12 パターンを作成した。標準化周波数は 8kHz、DFT については、窓幅 $N = 128$ [points]、シフト幅は $N/4$ 、窓関数は Hamming 窓とした。FDICA としては、NG[2] を用い、初期荷重は複素乱数、最大の更新回数は 100 回、探索ステップ幅 $\eta = 0.1$ とした。

ConvFDICA と提案法 (改良型 ConvFDICA) の初期値については、NG で 3 秒間の混合信号をもとに学習させ、その結果を分割スペクトルとビームフォーミングに基づいてスケールの不定性と成分置換を是正して得られ

Fig. 1 Placement of sound sources and microphones

る最終的な分離荷重を採択した．その後，8秒間の混合信号に対して分離性能を評価した．

また，ConvFDICAの場合，アルゴリズムの最大更新回数 It_{max} と探索ステップ幅 η を，分離フィルタのタップ長 $M = 4, 8, 16$ に応じてそれぞれ $(It_{max}, \eta) = (1000, 0.007) (2000, 0.002) (4000, 0.002)$ と決めた．これは，TDICAでは224, 352, 510タップ長に相当することになる．改良型ConvFDICAの場合，最大更新回数と探索ステップ幅はフィルタ長 $M = 4, 8, 16$ に応じてそれぞれ $(It_{max}, \eta) = (1000, 0.02) (2000, 0.006) (4000, 0.002)$ ， γ はすべて0.1に固定した．これらの値は，本実験データを用いた予備実験によって，NRR (Noise Reduction Rate) が最大となる組合せである．

4.2 分離性能のNRRによる評価

NRR (Noise Reduction Rate) は分離後の目的信号に占められる干渉信号の割合 (出力SNR) が分離前のそれ (入力SNR) と比べてどれだけ改善されたかを評価するもので， $NRR \equiv \text{出力SNR[dB]} - \text{入力SNR[dB]}$ と定義される．

Fig. 2にConvFDICAと改良型ConvFDICAによるNRRと分離フィルタのタップ長 M との関係を示す．ここに，NRRは12パターンの混合信号に対して得られたNRRの平均値である．NRRではConvFDICAは改良型ConvFDICAよりも高い値となっている．これには以下に述べるように白色化歪みが影響している．

Fig. 2 NRR of ConvFDICA and modified ConvFDICA

NRRの評価については，FDICAの場合，分離フィルタは遅れ項を含まない複素の定数行列であるため収束が安定しており，繰り返し回数を固定してもほぼ正確にNRRのピーク値が得られる．しかし，ConvFDICAや改良型ConvFDICA，TDICAのように分離フィルタが遅れ項を含む場合，FDICAと違って収束が安定せず，いったんNRRの値がピークに達しても，繰り返し回数が増加するとNRRの値は低下していく．すなわち，ICAのNRRは混合信号のパターンや探索ステップ幅，繰り返し回数に依存して変化する．そのため，NRRだけでは分離性能を正しく評価することはできない．このことは文献[13]で示唆されている．したがって，本論文ではSDR (Signal Distortion Ratio) を用いて白色化歪みもあわせて評価する．

4.3 SDRによる白色化歪みの評価

NRRには白色化歪みが生じる場合に正確な評価ができない問題があるため，ここでは歪みの尺度 (SDR: Signal Distortion Ratio) を

$$SDR \equiv \frac{1}{2} \sum_{i=1}^2 10 \log_{10} \frac{\sum_t |a_{ii}(z) s_i(t)|^2}{\sum_t |a_{ii}(z) s_i(t) - g_{ii}(z) s_i(t)|^2}$$

と定義して，このSDRにより白色化歪みを評価する．ここに， $g_{ii}(z)$ は分離フィルタ $W(z)$ と混合フィルタ $A(z)$ の積 $G(z) = W(z)A(z)$ の i 行 i 列要素である．これは，音響分野ではSQ (Sound Quality) として知られており，音質の評価量として用いられている[13]．この値が高いほど分離信号がより忠実に原信号を再現したことになり，歪みなく信号を分離できたことになる．

ここでは，NRRがピークをとるときのSDRを求める．これは，分離能力が最大と期待されるとき白色化歪みを評価することに相当する．Fig. 3に分離信号のSDRを示す．Fig. 3の縦軸はSDR，横軸はタップ長 M を示す．結果から，改良型ConvFDICAはConvFDICAよりSDRを約6dB改善しており，原音をより忠実に再現していることが読み取れる．

Fig. 4に， $M = 16$ のときの8秒間の分離信号の平均パワースペクトルを示す．a) は一人の話者に対するパワー

Fig. 3 SDR of ConvFDICA and modified ConvFDICA

スペクトル, b) はもう一人の話者に対するパワースペクトルで, 実線は原信号, 破線は ConvFDICA, 鎖線は改良型 ConvFDICA による結果を表している. 図より, 改良型 ConvFDICA の場合, 低域を除けば参照信号とほぼ一致しており, 原音をより忠実に再現できている.

Fig. 4 Power spectrum of separated signals

以上のことから, 提案した改良型 ConvFDICA は Servière の ConvFDICA に比べて, 原音をより忠実に再現できていることがわかる. また, Fig. 4 の結果から, 提案法では成分置換も改善されていることがわかる.

5. あとがき

本論文では, 高残響下で有効に機能するブラインド音源分離を開発するため, 従来法の利点と欠点を検討し, 新しいICA (改良型 ConvFDICA) を提案した. その結果, 従来法の問題点を以下のように改善した.

インパルス応答を各周波数ビンに分割するので, 残響が長くても各周波数ビンでフィルタ長は短くてすむ. これにより, 高残響下で TDICA や MDP に基づく TDICA のフィルタ長を長くすると収束しない欠点が解消された.

インパルス応答を各周波数ビンに分割するので, 短い窓幅でもすべての残響を包含できる. したがって, FDICA の窓幅を大きくすると十分な統計量が得られない欠点が解消された.

各周波数ビンに MDP を適用することにより, 白色化による歪みの影響を回避できる. これにより, ConvFDICA の白色化による歪みの影響を軽減するとともに, スケールの不定性の問題も解消した.

検証実験の結果, 提案法では白色化歪みによる劣化を抑えて良好に分離できており, 上述の改善点を裏付ける結果になっていた. このことから, 提案法は, 残響時間 300ms の高残響環境下において, 分離信号の音質の劣化

を最小限に抑え音声を良好に分離できることがわかった.

今回はアルゴリズムの高速化のため, タップ長をブロックサイズとしてブロックサイズ分のデータを溜め込んで分離フィルタを更新するブロックアルゴリズムで計算した. これにより, データが一つ入力されるごとに更新を行う on-line アルゴリズムよりも高速に分離を行うことができると考えられる. 今後は実用に向けて, 提案アルゴリズムのさらなる高速化の研究開発を行う予定である.

参考文献

- [1] J. Bell and T. J. Sejnowski: An information maximisation approach to blind separation and blind deconvolution; *Neural Computation*, Vol. 7, pp. 1129–1159 (1995)
- [2] A. Cichocki and S. Amari: *Adaptive Blind Signal and Image Processing, Learning Algorithm and Applications*, John Wiley & Sons, Ltd. (2002)
- [3] A. Hyvärinen, J. Karhunen and E. Oja: *Independent Component Analysis*, John Wiley & Sons, Ltd. (2001)
- [4] S. Ikeda and N. Murata: An method of blind separation based on temporal structure of signals; *Proc. 5th Int. Conf. on Neural Information Processing (ICONIP'98)*, Vol. 2, pp. 737–742 (1998)
- [5] S. Araki, S. Makino, R. Aichner, T. Nishikawa and H. Saruwatari: Subband based blind source separation with appropriate processing for each frequency band; *Proc. ICA2003*, pp. 499–504 (2003)
- [6] C. Servière: Separation of speech signals with segmentation of the impulse responses under reverberant conditions; *Proc. ICA2003*, pp. 511–516 (2003)
- [7] C. Servière and D. T. Pham: A novel method for permutation correction in frequency-domain in blind separation of speech mixtures; *Proc. ICA2004*, pp. 807–815 (2004)
- [8] S. Amari, S. C. Douglas, A. Cichocki and H. H. Yang: Novel online adaptive learning algorithms for blind deconvolution using the natural gradient approach; *Proc. SYSID*, Vols. 8–11, pp. 1057–1062 (1997)
- [9] K. Matsuoka and S. Nakashima: Minimal distortion principle for blind source separation; *Proc. ICA2001*, pp. 722–727 (2001)
- [10] 石橋, 井上, 五反田, 熊丸: 実環境下での伝達特性を利用した周波数領域 ICA の成分置換問題の解決; システム制御情報学会論文誌, Vol. 19, No. 12, pp. 471–478 (2006)
- [11] (社)日本音響学会: 新聞記事読み上げ音声コーパス; *JNAS*, Vols. 1–16 (1997)
- [12] Real World Computing Partnership: RWCP sound scene database in real acoustic environments (2001)
- [13] T. Takatani, T. Nishikawa, H. Saruwatari and K. Shikano: High-fidelity blind separation of acoustic signals using SIMO-model-based independent component analysis; *IEICE Trans. Fundamentals*, Vol. E87-A, No. 8, pp. 2063–2072 (2004)

付 録

(16) 式の導出

$$\begin{aligned}
E_t[\|\mathbf{y}(t) - \mathbf{x}(t)\|^2] &= \sum_{t=-\infty}^{\infty} \|\mathbf{y}(t) - \mathbf{x}(t)\|^2 \\
&\text{ここで } t = Nk + l \text{ とおくと} \\
&= \sum_{k=-\infty}^{\infty} \sum_{l=0}^{N-1} \|\mathbf{y}(Nk+l) - \mathbf{x}(Nk+l)\|^2 \\
&\text{パーセバルの定理より} \\
&= \sum_{k=-\infty}^{\infty} \frac{1}{N} \sum_{n=0}^{N-1} \|\mathbf{y}(\omega_n, k) - \mathbf{x}(\omega_n, k)\|^2 \\
&= \frac{1}{N} \sum_{n=0}^{N-1} E_k[\|\mathbf{y}(\omega_n, k) - \mathbf{x}(\omega_n, k)\|^2]
\end{aligned}$$

(18) 式 $\Delta \mathbf{W}_3(\omega_n, m)$ の導出

MDP に従えば $E[\|\mathbf{y}(\omega_n, k) - \mathbf{x}(\omega_n, k)\|^2]$ を最小化することにより、歪みが最小となる分離信号が得られる。したがって、 $E[\|\mathbf{y}(\omega_n, k) - \mathbf{x}(\omega_n, k)\|^2]$ の自然勾配を $\Delta \mathbf{W}_3(\omega_n, m)$ と定義して、計算を進めると、

$$\begin{aligned}
\Delta \mathbf{W}_3(\omega_n, m) &= -\frac{\partial E[\|\mathbf{y}(\omega_n, k) - \mathbf{x}(\omega_n, k)\|^2]}{\partial \mathbf{W}(\omega_n, m)} \\
&\quad \mathbf{W}^H(\omega_n, z) \mathbf{W}(\omega_n, z) \\
&= -E[2\{\mathbf{y}(\omega_n, k) - \mathbf{x}(\omega_n, k)\} \\
&\quad \mathbf{x}^H(\omega_n, k-m)] \mathbf{W}^H(\omega_n, z) \mathbf{W}(\omega_n, z) \\
&= -E[2\{\mathbf{y}(\omega_n, k) - \mathbf{x}(\omega_n, k)\} \\
&\quad \mathbf{y}^H(\omega_n, k-m)] \mathbf{W}(\omega_n, z)
\end{aligned}$$

となる。ここで、 $\mathbf{W}(\omega_n, z) = \sum_{q=0}^{M-1} \mathbf{W}(\omega_n, q) z^{-q}$ より

$$\begin{aligned}
\mathbf{y}^H(\omega_n, k-m) \mathbf{W}(\omega_n, z) &= \\
&\sum_{q=0}^{M-1} \mathbf{y}^H(\omega_n, k-m+q) \mathbf{W}(\omega_n, q)
\end{aligned}$$

となるので、 $\Delta \mathbf{W}_3(\omega_n, m)$ は

$$\begin{aligned}
\Delta \mathbf{W}_3(\omega_n, m) &= -2E[\{\mathbf{y}(\omega_n, k) - \mathbf{x}(\omega_n, k)\} \\
&\quad \sum_{q=0}^{M-1} \mathbf{y}^H(\omega_n, k-m+q)] \mathbf{W}(\omega_n, q)
\end{aligned}$$

となって、on-line の展開形は以下のように導出できる。

$$\begin{aligned}
\Delta \mathbf{W}_3(\omega_n, m) &= -2\{\mathbf{y}(\omega_n, k) - \mathbf{x}(\omega_n, k)\} \\
&\quad \sum_{q=0}^{M-1} \mathbf{y}^H(\omega_n, k-m+q) \mathbf{W}(\omega_n, q)
\end{aligned}$$

著 者 略 歴

舌 屋 武 志

2005 年近畿大学大学院博士後期課程修了、同年同大学研究生となり現在に至る。神経回路網、音声認識、独立成分分析とその応用に関する研究に従事。

いし ばし たか あき
石 橋 孝 昭 (正会員)

2007 年九州工業大学大学院博士後期課程修了。同年熊本電波工業高等専門学校情報通信工学科助教となり現在に至る。信号処理、パタン認識、独立成分分析とその応用に関する研究に従事。電子情報通信学会、計測自動制御学会などの会員。

しら つち ひろし
白 土 浩

1999 年九州工業大学大学院博士後期課程修了。同年琉球大学工学部助手、2005 年近畿大学産業理工学部講師、2009 年同大学准教授となり現在に至る。信号処理、神経回路網、独立成分分析とその応用に関する研究に従事。電子情報通信学会会員。

こ たん た ひろむ
五 反 田 博 (正会員)

1981 年九州大学大学院博士課程修了。同年近畿大学九州工学部講師、助教授、1998 年同大学教授となり現在に至る。1986～1987 年カリフォルニア大学パークレイ校客員研究員。システム同定、信号処理、神経回路網、独立成分分析とその応用に関する研究に従事。IEEE、電子情報通信学会、神経回路学会などの会員。