

時間領域 ICA と周波数領域 ICA を併用した多段 ICA によるブラインド音源分離*

猿渡洋†(P) 西川剛樹† 荒木章子‡ 牧野昭二‡ (†奈良先端大, ‡NTT・CS 研)

Blind Source Separation Based on Multi-Stage ICA Combining Time-Domain ICA and Frequency-Domain ICA

Hiroshi Saruwatari†(P) Tsuyoki Nishikawa† Shoko Araki‡ Shoji Makino‡ (†NAIST, ‡NTT CS Lab.)

E-mail: sawatari@is.aist-nara.ac.jp

Abstract

We propose a new algorithm for blind source separation (BSS), in which frequency-domain independent component analysis (ICA) and time-domain ICA are combined to achieve a superior source-separation performance under reverberant conditions. The results of signal separation experiments under reverberant conditions reveal that the signal separation performance of the proposed method is superior to that of the conventional frequency-domain ICA-based BSS method.

1 はじめに

複数の音源信号が混在されて観測された場合に、観測信号のみから音源信号を同定する技術を Blind Source Separation (BSS) と呼び、独立成分分析 (Independent Component Analysis: ICA) [1,2] の観点から音源を分離する手法が多く検討されている。本技術により、高精度な Hands-free 通信や雑音に対して頑健な音声認識の実現が期待できる [3]。

この BSS においては、残響が少ない環境では十分な分離性能が得られている [3,4]。しかし、高残響下での音源分離問題に関しては、(1) 従来の時間領域 ICA では反復学習規則が複雑であるため収束性が悪い、(2) 従来の周波数領域 ICA では帯域分割数を過渡に増やすと狭帯域信号間の独立性が低下する [5]、といった問題より高精度な音源分離は困難である。本稿では、各 ICA におけるこれらの問題点を実験結果に基づいて指摘するとともに、実環境での高精度な音源分離を目標とした周波数領域 ICA と時間領域 ICA を併用した BSS 手法の紹介を行う。

2 実環境における信号混合問題の定式化

実環境における音響信号の BSS を考える場合、(1) 観測点 (マイクロホン入力) 間に時間のずれがある、(2) 室内の壁・床等における反響により時間差をもった多数の反射波が同時に到来する、といった状態を考慮する必要がある。よって、実環境における BSS においては、時間差の無い単純な瞬時混合問題ではなく「畳み込み混合問題」を解くことになる。本問題を定式化すると、以下のように書くことができる。ここで、 $\mathbf{s}(t)=[s_1(t), \dots, s_L(t)]^T$ は L 個の音源信号を表すベクトル、 $\mathbf{x}(t)=[x_1(t), \dots, x_K(t)]^T$ はマイクロホン (K 素子) での観測信号ベクトル、 $\mathbf{a}(n)=[a_{kl}(n)]$ は畳み込み混合を表すインパルス応答を要素に持つ混合行列であり $\mathbf{A}(z)$ はその z 変換である。ただし、 z は単位遅延演算子であり、ここでは $z^{-\tau} \cdot \mathbf{x}(t) = \mathbf{x}(t-\tau)$ と表記している。

$$\mathbf{x}(t) = \left(\sum_{n=0}^{N-1} \mathbf{a}(n)z^{-n} \right) \cdot \mathbf{s}(t) = \mathbf{A}(z) \cdot \mathbf{s}(t) \quad (1)$$

3 従来の ICA に基づく BSS の問題点

3.1 時間領域 ICA

式 (1) をブラインドで解く手段として、まず時間領域で分離フィルタ行列 (以下逆混合行列を呼ぶ) $\mathbf{W}(z)=\mathbf{A}^{-1}(z)$ を ICA によって求めることを考える。これを一般に「時間領域 ICA」と呼ぶ。

この時間領域 ICA の一つとして、音源が非定常信号であるとき 2 次の相互相関値が零になったときのみ最小値をとる次式のような非負の評価関数を最小化することによって音源分離する手法が提案されている [6]。

$$Q = \frac{1}{2} \sum_{b=1}^B \left\{ \sum_{l=1}^L \log \langle y_l(t-D) \rangle_b \right\}$$

$$- \log \det \langle \mathbf{y}(t-D) \mathbf{y}(t-D)^T \rangle_b \quad (2)$$

ここで $\mathbf{y}(t)=[y_1(t), \dots, y_L(t)]^T$ は分離信号ベクトル、 $\langle \cdot \rangle_b$ は局所分析区間 b ($b=1, \dots, B$) における時間平均であり、 B, D はそれぞれ分析区間数、フィルタに因果性をもたせるための遅延である。また $\mathbf{W}^{(t)}(z)$ は分離フィルタ係数 $\mathbf{w}^{(t)}(n)$ の z 変換である。式 (2) はすべての局所分析区間 b において $y_i(t-D)$ と $y_j(t-D)$ が無相関になったときのみ零となる。式 (2) の評価関数を最小化するための更新式は次式のように与えられる。

$$\mathbf{w}_{i+1}^{(t)}(n) = \alpha z^{-n} \sum_{b=1}^B \left\{ \mathbf{I} - \left(\text{diag} \langle \mathbf{y}(t-D) \mathbf{y}(t-D)^T \rangle_b \right)^{-1} \times \langle \mathbf{y}(t-D) \mathbf{y}(t-D)^T \rangle_b \right\} \mathbf{W}_i^{(t)}(z) + \mathbf{w}_i^{(t)}(n) \quad (3)$$

時間領域 ICA における特徴を以下にまとめる。(1) 帯域分割を行わず音源信号そのものの統計的性質を利用しているので、収束点近傍では音源分離を高精度に行うことが出来る。(2) しかし、分離行列を学習する規則が複雑かつ不安定であり、実際の音環境における分離に適用するのは困難である。例えば、典型的な室内残響は約 300 ms (これは 8 kHz サンプリング時に 2400 taps の FIR フィルタに相当) であり、この場合の分離行列 $\mathbf{w}(n)$ も数千 taps のフィルタとなる。しかし、従来の研究では、このような長いフィルタの学習はほぼ不可能であり、実用的には 100 taps 程度のフィルタの学習のみが可能であるとされている。

3.2 周波数領域 ICA

式 (1) を解く別の手法として、観測信号に短時間 DFT 分析を適用して複数の狭帯域「時間・周波数信号」に分解し、各帯域にて別々に ICA を適用する手法が提案されている。これを「周波数領域 ICA」と呼ぶ。

観測信号 $\mathbf{x}(t)$ を時間・周波数信号に変換したものを $\mathbf{X}(f, t)$ と書く場合、分離出力 $\mathbf{Y}^{(f)}(f, t)$ は以下のように書くことができる。

$$\mathbf{Y}^{(f)}(f, t) = \mathbf{W}^{(f)}(f) \mathbf{X}(f, t) \quad (4)$$

ここで $\mathbf{W}^{(f)}(f)$ は複素係数を要素に持つ逆混合行列であり、時間遅れの項を含まない。よって、この手法においては各狭帯域にて瞬時混合問題となるため、比較的簡単かつ安定に学習を行うことが可能となる。

逆混合行列 $\mathbf{W}^{(f)}(f)$ の計算法としては、Kullback-Leibler divergence の最小化に基づいた教師なし適応アルゴリズムが提案されている [7]。このとき、最適な $\mathbf{W}^{(f)}(f)$ は以下の更新式で得られる。

$$\mathbf{W}_{i+1}^{(f)}(f) = \eta \left[\text{diag} \left\{ \langle \Phi(\mathbf{Y}^{(f)}(f, t)) \mathbf{Y}^{(f)}(f, t)^H \rangle \right\} - \langle \Phi(\mathbf{Y}^{(f)}(f, t)) \mathbf{Y}^{(f)}(f, t)^H \rangle \right] \mathbf{W}_i^{(f)}(f) + \mathbf{W}_i^{(f)}(f) \quad (5)$$

ただし $\langle \cdot \rangle$ は、入力信号をフレーム分析した際の全フレームに関する時間平均であり、 $\Phi(\cdot)$ はシグモイド関数等を要素に持つ適当な非線形ベクトル関数である。

周波数領域 ICA における特徴を以下にまとめる。(1) 各狭帯域にて別々に瞬時混合問題を解く ICA を行うことより、安定に解が求まる反面、各帯域間にて音源入れ替わり・利得不定等の問題が生じる。これらは、周波数分解処理に付随して生じた問題であり、いくつかの有効な解決策が既に提案・導入されている [7,8]。(2) 長い残響に対応するためには、帯域分割数を増やす必要がある。しかし、帯域分割数を過度に増やすと狭帯域信号間の独立性が低下するという現象により、残響を含む音の分離に関しては性能限界が存在する [5]。

*キーワード：信号処理、最適化、実験、応用

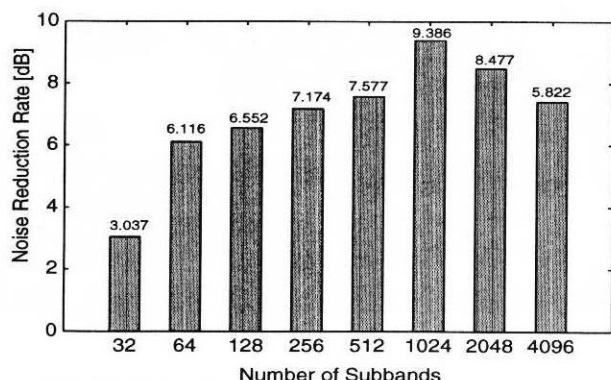


図 1: 周波数領域 ICA における帯域分割数と分離性能の関係 (RT=300 ms)

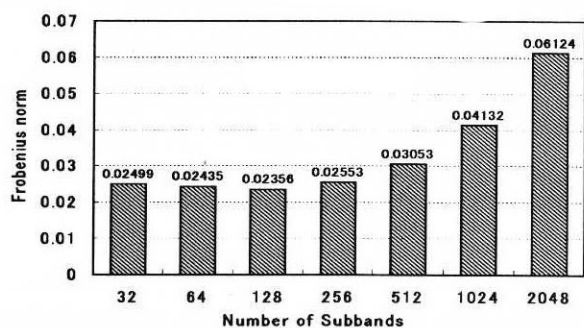


図 2: 帯域分割数と音源間独立性の関係

(2) は周波数領域 ICA における本質的かつ解決困難な問題である。以下にて本問題の一例を示す。図 1 は、帯域分割数を増やした場合の音源分離性能の例である。本実験では、素子間隔 4 cm の 2 素子アレーを使用し、残響時間 (RT) は 300 ms、方位 -30° 及び 40° から 3 秒の独立な音声が発射されるものとした。ここでは、帯域分割数 (DFT 分析フレーム長) を 32 点から 4096 点 (4 ms から 384 ms に相当) の範囲で変化させ分離性能の比較を行った。分離性能の客観評価尺度として、Noise Reduction Rate (NRR; 出力 SNR [dB] - 入力 SNR [dB]) を用いている。

図 1 より、帯域分割数を増やすと分離性能は向上していくが、過度に増やすと分離性能は劣化してしまうことが分かる。本現象の原因としては、(1) 過度に分割数を増やすと狭帯域信号間の独立性が低下する、(2) 時間サンプル数が減るため独立性の統計的推定が困難になる、といったことが考えられる。これを確認するため、今回使用した 2 音源そのものを分離信号 $\mathbf{Y}(f, t)$ と見なし、それに関して式 (5) の $[\cdot]$ 項の行列ノルムを計算したものを図 2 に示す。本図より、帯域分割数を増やすと独立性が低下・推定が困難になり、式 (5) の最小化が困難になることが分かる。

4 時間領域 ICA と周波数領域 ICA を併用した多段 ICA による BSS

本章では、周波数領域 ICA と時間領域 ICA を併用した多段 ICA に基づく BSS 手法について述べる。図 3 に提案法のフローチャートを示す。

具体的な分離処理過程としては、まず最初に観測信号を周波数領域 ICA を用いて分離する。ここでは安定な逆混合行列の推定が可能であるという周波数領域 ICA の利点を利用する。次に、周波数領域 ICA によって得られた分離信号を時間領域 ICA の入力信号とみなし、時間領域 ICA による音源分離を行う。ここでは信号間の独立性に対して広帯域信号そのものの独立性を評価できる時間領域 ICA の利点を利用する。そして時間領域 ICA によって得られた分離信号を最終的な分離信号とする。以上の処理により、安定かつ高精度な音源分離が実現でき、それぞれの ICA を単独で用いる場合よりも分離性能の向上が期待される。

図 4 に、提案法と周波数領域 ICA の分離性能に関する比較を示す。横軸は実験に用いた話者組合せであり、縦軸は客観分離尺度 (NRR) である。ここでは、時間領域 ICA として 3.1 節にて述べた手法を、周波数領域 ICA として我々が以前に提案した BF に基づく高速収束周波数領域 ICA (BF-Based FDICA)[4] を使用した。

本図より、12 通りのすべての話者組合せにおいて提案法による分離性能の向上がみられ、提案法の有効性が確認される。具体的に

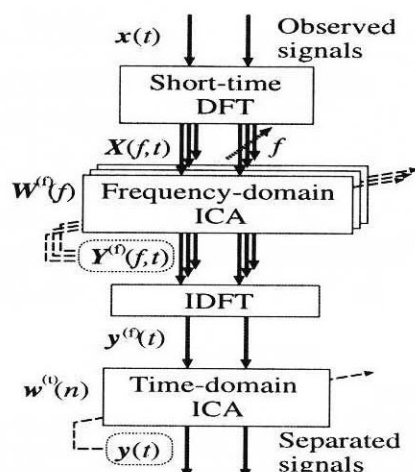


図 3: 周波数領域 ICA と時間数領域 ICA を併用した多段 ICA による音源分離手法

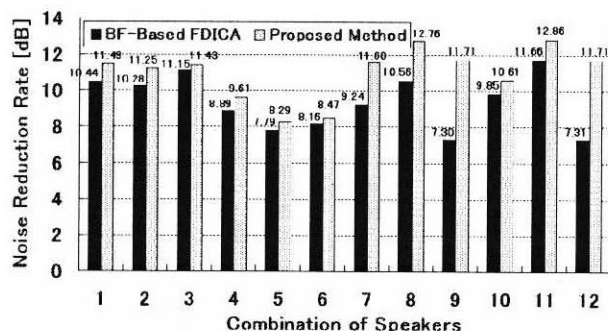


図 4: 提案法 (Proposed Method) と BF に基づく周波数領域 ICA (BF-Based FDICA) の分離性能に関する比較 (RT=300 ms)

は、12 通りの平均として従来法の分離性能が約 9.4 dB であったのに対し、提案法は約 1.6 dB の向上がみられ、最終的な分離性能は約 11 dB であった。

今回示した BF に基づく周波数領域 ICA の NRR 値は、分離性能の最大値 (飽和点) を与える反復回数のものであった。つまり、上で述べた従来法の NRR 値は、周波数領域 ICA の性能限界であったとすることができる。今回の実験では、それを上回る分離性能が得られたことより、提案法は本質的に従来の周波数領域 ICA の問題点を解決したと言える。

5 まとめ

本稿では、実環境下における BSS に関して、時間領域 ICA と周波数領域 ICA の問題点を指摘した。また、残響環境下における高精度な音源分離を実現するために、周波数領域 ICA と時間領域 ICA とを併用した多段 ICA による BSS 手法を提案した。通常の室内残響環境下での分離実験より、提案手法による音源分離性能は従来の周波数領域 ICA による音源分離性能を上回ることが確認された。

謝辞

時間領域 ICA に関して御議論頂いた島根大学の河本満博士に感謝致します。周波数領域 ICA に関して御議論頂いた NTT-CS 研究所の向井良氏に感謝いたします。本研究の一部は科学技術振興事業団による戦略的基礎研究推進事業 (CREST) の援助を受けて行われた。

参考文献

- [1] A. Bell et al.: *Neural Computation*, vol.7, pp.1129-1159, 1995.
- [2] P. Common, *Signal Processing*, vol.36, pp.287-314, 1994.
- [3] H. Saruwatari et al.: *Proc. ICSLP2000*, vol.3, pp.94-97, Oct. 2000.
- [4] 川村 他: 信学技法, vol.EA2001-2, pp.9-16, April 2001.
- [5] S. Araki et al.: *Proc. ICASSP2001*, May 2001.
- [6] M. Kawamoto et al.: *Neurocomputing* 22, pp.157-171, 1998.
- [7] N. Murata et al.: *Proc. of Int'l national Sympo. on Nonlinear Theory and its Application*, vol.3, pp.923-926, Sept. 1998.
- [8] S. Kurita et al.: *Proc. ICASSP2000*, vol.5, pp.3140-3143, June 2000.