

非定常信号の同時無相関化に基づく時間領域 ICA の比較 *

©西川 剛樹† 高谷 智哉† 猿渡 洋† 鹿野 清宏† 荒木 章子† 牧野 昭二†

(† 奈良先端大科学技術大学院大学 情報科学研究科, ‡ NTT コミュニケーション科学基礎研究所)

1 はじめに

複数の音源信号が混在して観測される場合、観測信号のみから音源信号を同定する技術をブラインド音源分離 (BSS) と呼び、独立成分分析 (ICA) [1] に基づく手法が多く検討されている [2, 3, 4, 5]. 本技術により、高精度なハンズフリー音声認識システムの実現が期待できる. ICA に基づく BSS は分離フィルタの学習領域に関して、周波数領域 ICA (FDICA) と時間領域 ICA (TDICA) に分類される. FDICA は、ある程度の分離を行うことのできるフィルタを安定に推定することが可能である. しかし、帯域分割処理を行うため、十分な性能に達しない段階で性能が飽和してしまうという欠点があった [2]. この問題を解決するために、我々は FDICA と TDICA を統合した多段 ICA (MSICA) を提案した [3]. MSICA では、FDICA の残留クロストーク成分を TDICA を用いて分離することが可能であった. よって、FDICA の分離性能をさらに向上させる TDICA 部は非常に重要である. 本稿では、TDICA に注目し、我々が提案した非定常信号の同時無相関化に基づく TDICA (NS-TDICA) [3] と、Kullback-Leibler 情報量最小化に基づく TDICA (KLD-TDICA) [4] の比較を行う.

2 音源分離手法

TDICA によるブラインド音源分離の処理過程を図 1 に示す. N 個の音源信号が線形に混合されて M 個のマイクロホンで観測されたとする. 時間 t における観測信号を $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]^T$ と表し、分離フィルタ $\mathbf{w}(k)$ を用いて音源分離を行う. 分離フィルタ $\mathbf{w}(k)$ は、 N 個の出力信号 $\mathbf{y}(t) = [y_1(t), \dots, y_N(t)]^T$ が互いに独立となるように学習される. この処理は、以下のように示される.

$$\mathbf{y}(t) = \sum_{k=1}^{K-1} \mathbf{w}(k) \mathbf{x}(t-k) = \mathbf{W}(z) \cdot \mathbf{x}(t) \quad (1)$$

ここで $\mathbf{W}(z)$ は次式で与えられるような分離フィルタ $\mathbf{w}(k)$ の z 変換である.

$$\mathbf{W}(z) = \sum_{k=0}^{K-1} \mathbf{w}(k) z^{-k} = \left[\sum_{k=0}^{K-1} w_{ij}(k) z^{-k} \right]_{ij} \quad (2)$$

ここで z は単位遅延演算子であり、 $z^{-k} \cdot \mathbf{x}(t) = \mathbf{x}(t-k)$ と表記する. また $[\cdot]_{ij}$ は行列の i 行 j 列要素を表す.

2.1 NS-TDICA [3]

NS-TDICA は非定常信号の同時無相関化と Time-Delayed Decorrelation (TDD) を統合した手法であり、シンプルな同時無相関化 ICA や TDD よりも高精度な音源分離を実現できる [3]. この手法では、式 (1) における分離フィルタを以下の式に基づいて最適化する.

$$\mathbf{w}_{i+1}^{(NS)}(k) = \mathbf{w}_i^{(NS)}(k) + \frac{\alpha}{B} \sum_{b=1}^B \{ (\text{diag} \mathbf{R}_y^{(b)}(0))^{-1} \text{diag} \mathbf{R}_y^{(b)}(k) \} \mathbf{W}_i^{(NS)}(z) \quad (3)$$

*Comparison of time-domain ICA methods based on minimization of KL divergence and simultaneous decorrelation of nonstationary signal, by T. Nishikawa, T. Takatani, H. Saruwatari, K. Shikano (Nara Institute of Science and Technology), S. Araki, and S. Makino (NTT Communication Science Laboratories, NTT Corporation).

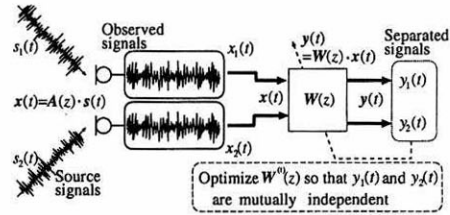


図 1: TDICA によるブラインド音源分離の処理過程

$$- (\text{diag} \mathbf{R}_y^{(b)}(0))^{-1} \mathbf{R}_y^{(b)}(k) \} \mathbf{W}_i^{(NS)}(z) \quad (3)$$

ここで B は局所分析区間数、 α はステップサイズパラメータである. また、 $\mathbf{R}_y^{(b)}(n) = \langle \mathbf{y}(t) \mathbf{y}(t-n)^T \rangle_t^{(b)}$ は分離信号の局所分析区間 b における相関行列である. ただし、 $\langle \cdot \rangle_t^{(b)}$ は b 番目の分析区間に関する時間平均である.

2.2 KLD-TDICA [4]

KLD-TDICA は分離信号の Kullback-Leibler 情報量を最小にする手法である. この手法では、式 (1) における分離フィルタを以下の式に基づいて最適化する.

$$\mathbf{w}_{i+1}^{(KLD)}(k) = \mathbf{w}_i^{(KLD)}(k) + \beta \{ \text{diag} \{ \phi(\mathbf{y}(t)) \mathbf{y}(t-k)^T \}_t - \langle \phi(\mathbf{y}(t)) \mathbf{y}(t-k)^T \rangle_t \} \mathbf{W}_i^{(KLD)}(z) \quad (4)$$

ここで β はステップサイズパラメータであり、非線形ベクトル関数 $\phi(\mathbf{y}(t)) = [\phi(y_1(t)), \dots, \phi(y_N(t))]$ の各要素は、以下のように定義される.

$$\phi(y_n(t)) \equiv \tanh(y_n(t)) \quad (5)$$

3 音源分離実験

3.1 実験条件

実験環境で計測されたインパルス応答を畳み込んだ残響付加音声に対して検討を行う. 残響時間は 300 ms 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、方位 -30° 及び 40° から独立な音声が発射されるものとした. 音源とアレー間の距離は 1.15 m であり、混合時の SNR は 0 dB である. 音源信号としては、ASJ 研究用連続音声コーパスより選択した男女各 2 名による 2 通りの文を使用した. これらを用いて、12 通りの話者・音源位置の組合せを得ることができる. 音声信号の標準化周波数は 8 kHz である.

4 実験結果

4.1 TDICA に関する比較

本実験では、分離性能の客観評価尺度として、Noise Reduction Rate (NRR; $\equiv$ 出力 SNR [dB] - 入力 SNR [dB]) を用いる. NS-TDICA, KLD-TDICA の各 TDICA の分析条件としては、フィルタ長を 8~4096 タップ、反復回数を 500 回、ステップサイズパラメータを $5.0 \times 10^{-1} \sim 1.0 \times 10^{-7}$ とした. また、NS-TDICA における分析区間数 B については、1 から 10 (3 s の音声を 1 等分から 10 等分) 分割して実験を行った. ここで、分析区間数 (NS-TDICA のみ)、反復回数、ステップサイズパラメータに関しては、12 通りの平均の NRR が最も高くなる回数、値を選択した. 分離フィ

ルタ $w(k)$ の初期値としては、(1) ストレートパスのみを通すフィルタ (対角要素が $\delta(k-K/2)$ 、ただし、 $\delta(0)=1$, $\delta(m)=0$, ($m \neq 0$)、非対角要素が 0)、(2) $\pm 60^\circ$ に死角を持つような死角ビームフォーマ (NBF) の 2 通りを用いて実験した。

図 2 に異なるフィルタ長毎の NS-TDICA と KLD-TDICA の分離性能を示す。図 2 より、初期値に関しては、NS-TDICA、KLD-TDICA とともに NBF を用いた方が分離性能の向上がみられた。この結果より、高精度な音源分離を実現するために、良い初期値を用いることは有効であるといえる。NS-TDICA と KLD-TDICA の比較としては、フィルタ長が短い場合、KLD-TDICA の方が分離性能が高い。しかし、フィルタ長が長い場合は NS-TDICA の方が分離性能が高い。よって、残響環境下のように長いフィルタが必要な場合は NS-TDICA の方が有効であるといえる。

4.2 MSICA の TDICA 部に関する比較

MSICA[3] の後段の TDICA 部に NS-TDICA、および KLD-TDICA を適用し、それぞれの TDICA の比較を行う。MSICA の FDICA 部には Saruwatari によって提案された高速収束 FDICA[5] を用いる。この FDICA (1024 タップ) を適用することで、9.4 dB の分離性能が得られた。各 TDICA の分析条件は 4.1 節と同様である。

図 3 に MSICA の TDICA 部に関する異なるフィルタ長毎の NS-TDICA と KLD-TDICA の分離性能を示す。図 3 より、KLD-TDICA の 2048 タップにおいて、分離性能が最も高く、12.1 dB の性能が得られた。しかしながら、すべてのフィルタ長に関して NS-TDICA、KLD-TDICA の分離性能はほぼ同等である。これらの結果より、MSICA の後段の TDICA 部に NS-TDICA と KLD-TDICA のどちらを適用するかという点では優位な差がなかった。

4.3 NS-TDICA の分析区間数に関する検討

NS-TDICA において、分離性能を向上させるためには、信号の非定常性がより顕著に現れるように分析区間数を増やすべきである。そこで、分析区間数と分離性能の関係を調べる。実験は MSICA で行った。

図 4 に NS-TDICA の異なる分析区間数毎の分離性能、および KLD-TDICA の分離性能を示す。図 4 より、分析区間数を増やすことにより、分離性能が向上するという傾向がみられる。しかしながら、分析区間数が少ない方が性能が良いタップが存在する。これは、信号毎に非定常性が異なるためである。一般的に信号の非定常性を評価することは困難であるため、観測信号毎に最適な分析区間数を推定することも困難である。NS-TDICA と KL-TDICA を比較すると、KL-TDICA の分離性能は、NS-TDICA の分析区間数を変化させたときの中間的な性能、もしくは上回る性能であることが分かる。KL-TDICA は最適な分析区間の推定が不要であり、かつ NS-TDICA の最適な分割数を選択した場合とはほぼ同等の性能が得られるという点で有効であるといえる。

5 まとめ

本稿では、KLD-TDICA と NS-TDICA の比較を行った。通常の室内残響環境下での音声の分離実験より、各手法の分離性能に関しては、ほぼ同等の性能であった。また、KLD-TDICA は NS-TDICA の分析区間数を様々変化したときの中間的な分離性能、もしくはそれを上回る性能であった。よって、最適な分析区間数を推定する必要がないという点で、KLD-TDICA は有効であるといえる。今回の実験を通して、TDICA の分離信号が白色化し歪む傾向がみられた。よって、今後の課題としては、分離信号の歪み補性を行う必要がある。

謝辞 本研究に関して御議論頂いた島根大学の河本満博士、NTT-CS 研究所の Robert Aichner 氏に感謝いたします。

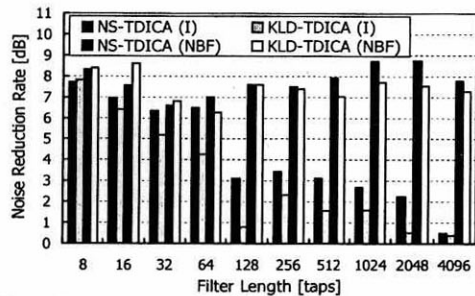


図 2: 異なるフィルタ長毎の NS-TDICA と KLD-TDICA の分離性能。“I” は初期値がストレートパスのみを通すフィルタ、“NBF” は初期値が NBF であることを表す。

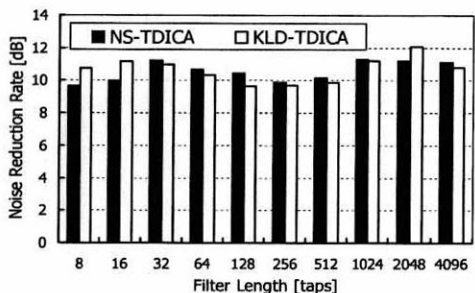


図 3: MSICA の TDICA 部に関する異なるフィルタ長毎の NS-TDICA と KLD-TDICA の分離性能

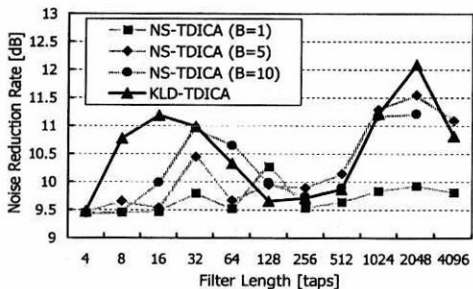


図 4: NS-TDICA における分析区間数毎の分離性能

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)「高度メディア社会の生活情報技術」、および NTT 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] S. Araki et al.: *Proc. International Conference Independent Component Analysis Signal Separation*, pp.132–137, Dec. 2001.
- [3] T. Nishikawa et al.: *Proc. International Conference on Fundamentals of Electronics, Communication and Computer Sciences*, R-1, pp.7–12, March 2002.
- [4] S. Choi et al.: *Proc. International Conference Independent Component Analysis Signal Separation*, pp.371–376, Jan. 1999.
- [5] H. Saruwatari et al.: *Proc. European Conference on Speech Communication and Technology*, vol.4, pp.2603–2606, sept. 2001.