

周波数領域 ICA における初期値の短時間データからの学習

Initial Value Estimation for Frequency Domain ICA from Limited Data

荒木章子¹ 伊藤信貴² 澤田宏¹ 小野順貴² 牧野昭二¹ 嵯峨山茂樹²
Shoko Araki Nobutaka Ito Hiroshi Sawada Nobutaka Ono Shoji Makino Shigeki Sagayama

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
NTT Communication Science Laboratories, NTT Corporation
東京大学大学院 情報理工学系研究科²

Graduate School of Information Science and Technology, The University of Tokyo

1 まえがき

独立成分分析 (ICA) によるブラインド音源分離手法は、近年、実環境音源分離において高い性能をあげている。ICA の代表的なアルゴリズムに InfoMax[1] がある。これは高い分離性能が得られるが、収束が遅いため分離行列の適切な初期値設定が必要である。そこで例えば [2] では、少ない反復で収束する FastICA[3] で分離行列を得、それを InfoMax の初期値としている。しかし FastICA は、観測データ長が短い場合に良い初期値を与えない。一方、短いデータからの初期値計算は、音源位置などが変化する環境で ICA を適用する時に重要である。そこで本稿では、信号のスパース性を用いた、短時間データからの InfoMax の初期値計算法を提案する。

2 提案手法

本稿では、時間領域で観測した信号 $x_j(t)$ ($j = 1, \dots, M$) に短時間フーリエ変換 (STFT) を適用し、時間周波数領域にて信号を取り扱う。時間周波数領域における観測信号 $x_j(f, \tau)$ は近似的に

$$\mathbf{x}(f, \tau) \approx \sum_{k=1}^N \mathbf{h}_k(f) s_k(f, \tau) \quad (1)$$

と書ける。ここで $\mathbf{x} = [x_1, \dots, x_M]^T$, $\mathbf{h}_k = [h_{1k}, \dots, h_{Mk}]^T$ は信号源 k から各マイク $j = 1, \dots, M$ への周波数応答、 $s_k(f, \tau)$ ($k = 1, \dots, N$) は原信号の STFT 結果である。本稿では、音源数 $N \leq$ マイク数 M の場合を取り扱う。音源分離の目的は、 $\mathbf{x}(f, \tau)$ から分離信号 $\mathbf{y} = [y_1, \dots, y_N]^T$ を得ることである。

従来法では、各周波数の観測信号 $\mathbf{x}(f, \tau)$ に対し、まず FastICA を適用し分離行列 $\mathbf{W}_F^{(0)}(f)$ を求め、それを初期値とした InfoMax にて、最終的な分離行列 $\mathbf{W}(f)$ を得て、 $\mathbf{y}(f, \tau) = \mathbf{W}(f)\mathbf{x}(f, \tau)$ にて分離を行っていた。しかし FastICA は、観測信号が短い時に良い初期値を与えない。

そこで提案法では、信号のスパース性を利用して分離行列の初期値 $\mathbf{W}_P^{(0)}$ を推定する。まず、各時間周波数 (f, τ) においてある 1 つの音源 s_k のみが支配的であること、すなわち式 (1) が

$$\mathbf{x}(f, \tau) \approx \mathbf{h}_k(f) s_k(f, \tau) \quad (2)$$

と近似されることを仮定する。そして、マイク間のレベル差や位相差を用いて観測信号 $\mathbf{x}(f, \tau)$ をクラスタリングし、各音源 s_k が支配的な (f, τ) の集合 Ω_k を推定する [4]。次に、 $\mathbf{h}_k(f)$ を推定する。ここでは (2) における音源 s_k への依存性を無くすため、(2) を

$$\mathbf{x}(f, \tau) / x_J(f, \tau) \approx \mathbf{h}_k(f) / h_{Jk}(f) = \mathbf{h}'_k(f) \quad (3)$$

と変形する。ここで J は任意のマイク番号である。(3) において、 \mathbf{h}'_k には h_{Jk} のスケールリングの任意性があるが、ICA では問題とならない。(3) より、 $\mathbf{x}(f, \tau) \approx \mathbf{h}'_k(f) x_J(f, \tau)$ であるので、各音源 k に関する周波数応答 $\mathbf{h}'_k(f)$ は、最小二乗推定により

$$\mathbf{h}'_k(f) = \frac{\sum_{(f, \tau) \in \Omega_k} \mathbf{x}(f, \tau) x_J^*(f, \tau)}{\sum_{(f, \tau) \in \Omega_k} |x_J(f, \tau)|^2} \quad (4)$$

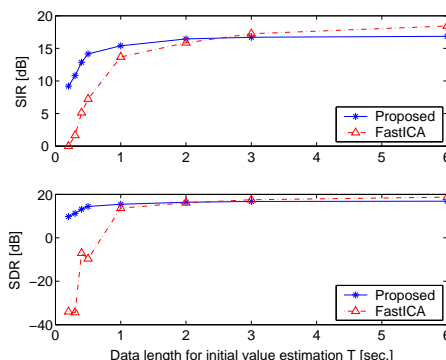


図 1 実験結果 (5 組の音声組合せの平均値)

にて推定できる。最後に、 $\mathbf{W}_P^{(0)}(f) = \mathbf{H}'(f)^{(+)}$ (+ は一般化逆行列) により InfoMax の初期値を与える。ここで $\mathbf{H}'(f) = [\mathbf{h}'_1, \dots, \mathbf{h}'_N]$ である。

3 実験と結果

音源数 $N = 2$ マイク数 $M = 2$ とし、音源に音声信号を用いて実験を行った。分離対象の混合信号は、6 秒間の英語音声に、残響時間 200ms の部屋で計測したインパルス応答を畳み込んで作成した。サンプリング周波数は 8kHz、フレーム長は 2048、シフトは 256 とした。初期値の学習には分離対象混合信号の冒頭 T 秒を用い、この T をパラメタとした。

提案法、FastICA それぞれによる初期値 $\mathbf{W}_P^{(0)}(f)$, $\mathbf{W}_F^{(0)}(f)$ を用いた分離性能 (信号対妨害音比 (SIR) 改善量と信号対歪比 (SDR)) を、図 1 に示す。初期値の学習に用いるデータが短い場合 (2 秒以下)、提案法が有効であった。尚、初期値推定と分離に要する時間は、 $T = 6$ (秒) の場合、FastICA による方法では 7.4 秒であり、提案法による方法では 5.6 秒であった (Intel Pentium4, 2.2GHz, Matlab6.5)。以上より、スパース性による InfoMax の初期値設定の有効性が示された。

参考文献

- [1] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [2] R. Mukai, H. Sawada, S. Araki, S. Makino, "Blind source separation of many signals in the frequency domain," *ICASSP 2006*, vol. V, pp. 969–972, 2006.
- [3] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *International Journal of Neural Systems*, vol. 10, no. 1, pp. 1–8, 2000.
- [4] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors," *Signal Processing*, vol. 87, pp. 1833–1847, 2007.