

観測ベクトルのクラスタリングによるブラインド音源分離

Blind Signal Separation by Observation Vector Clustering

荒木章子
Shoko Araki

澤田宏
Hiroshi Sawada

向井良
Ryo Mukai

牧野昭二
Shoji Makino

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
NTT Communication Science Laboratories, NTT Corporation

1 まえがき

ブラインド音源分離で広く研究がなされている独立成分分析による方法 (e.g., [1]) はセンサ数 $M \geq$ 信号数 N の場合のみを扱い、 $M < N$ の場合も扱える信号のスパース性 (後述) による方法の多く (e.g., [2]) は、2 センサの位相差またはゲイン差を用いるため 3 個以上のセンサ情報を効果的に利用できない。そこで本稿では、センサや信号源の数によらず、3 個以上のセンサ情報を全て利用しながら、更に任意のセンサ配置にてブラインド音源分離を可能とする方法について実験結果を交えて示す。

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

$$x_j(f, \tau) \approx \sum_{k=1}^N h_{jk}(f) s_k(f, \tau), \quad (1)$$

と書ける。ここで $s_k(f, \tau)$ ($k = 1, \dots, N$) は原信号の STFT 結果、 $h_{jk}(f)$ は信号源 k からセンサ j への周波数応答である。またこれらをまとめた $\mathbf{x} = [x_1, \dots, x_M]^T$ を本稿では観測信号ベクトルと呼ぶ。目的は $\mathbf{x}(f, \tau)$ から分離信号 $\tilde{s}_k(f, \tau)$ ($k = 1, \dots, N$) を得ることである。

提案法では、各時間周波数 (f, τ) において原信号のうちの 1 つ s_k のみが支配的であること、すなわち式 (1) が $x_j(f, \tau) \approx h_{jk}(f) s_k(f, \tau)$ と近似されることを仮定する。これは音声信号などで確認され (広義で) スパース性と呼ばれる。スパース信号の分離には、各時間周波数 (f, τ) でどの s_k が支配的かを判定し、ある s_k が支配的な時間周波数 (f, τ) の観測のみを集めれば良い。

これを本手法では、観測信号ベクトルのクラスタリングで実現する (詳細は [3])。まず、クラスタリングをしやすくするため、観測信号ベクトルを

$$\bar{x}_j(f, \tau) \leftarrow |x_j(f, \tau)| \exp \left[j \frac{\arg[x_j(f, \tau)/x_J(f, \tau)]}{4fc^{-1}d_{\max}} \right] \quad (2)$$

$$\bar{\mathbf{x}}(f, \tau) \leftarrow \bar{\mathbf{x}}(f, \tau) / \|\bar{\mathbf{x}}(f, \tau)\| \quad (3)$$

と正規化する (c は伝播速度、 d_{\max} はセンサ J (基準) と $\forall j \in \{1, \dots, M\}$ の距離の最大値)。次に正規化された $\bar{\mathbf{x}}(f, \tau)$ を全ての (f, τ) について例えば k-means 法などでクラスタリングすると、信号源とセンサの位置関係に基づいてクラスタ C_1, \dots, C_N が形成される。それぞれのクラスタ C_k が個々の信号 s_k に相当するので、

$$\tilde{s}_k(f, \tau) = \begin{cases} x_J(f, \tau) & \bar{\mathbf{x}}(f, \tau) \in C_k \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

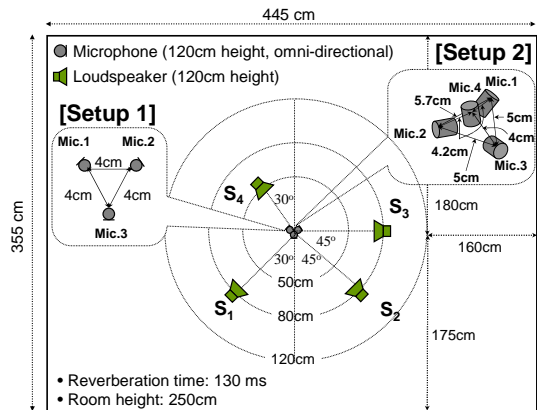


図 1 実験条件

表 1 実験結果 (SIR 改善量 [dB] と SDR [dB])

[Setup 1]	\tilde{s}_1	\tilde{s}_2	\tilde{s}_3	\tilde{s}_4
入力 SIR	-7.4	-6.2	-6.0	-1.0
SIR	17.9	11.8	15.9	12.3
SDR	5.6	4.8	5.8	8.8
[Setup 2]	\tilde{s}_1	\tilde{s}_2	\tilde{s}_3	\tilde{s}_4
入力 SIR	-8.1	-5.3	-6.6	-0.6
SIR	19.3	17.3	10.1	17.4
SDR	4.9	6.6	4.2	12.2

により分離信号 $\tilde{s}_k(f, \tau)$ を抽出することができる。

3 実験と結果

音声信号を用い実験を行った。図 1 に実験環境を示す。図のように、センサは非直線配置である。混合信号は、5 秒間の英語音声に、図 1 の環境にて計測したインパルス応答を畳み込んで得た。サンプリング周波数は 8kHz、フレーム長 L は 512、シフトは $L/8$ とした。

分離結果を、信号対妨害音比 (SIR) 改善量と信号対歪比 (SDR) で評価した (表 1)。センサ数 $M <$ 信号数 N の時やセンサ配置が不規則な時でも、高い分離性能を得ることができた (表 1 および [4])。このような条件における分離を実現したのは本提案が初めてである。

参考文献

- [1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.
- [2] Ö. Yılmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. on SP*, vol. 52, no. 7, pp. 1830–1847, 2004.
- [3] S. Araki, H. Sawada, R. Mukai and S. Makino, "A novel blind source separation method with observation vector clustering," in *Proc. IWAENC2005*, Sept. 2005.
- [4] http://www.kecl.ntt.co.jp/icl/signal/araki/xcluster_fine.html