音源のアクティベーションを事前情報とした独立ベクトル分析による音源分離*

☆小野拓磨 (東大院・情報理工), 小野順貴 (NII), 嵯峨山茂樹 (東大院・情報理工)

1 はじめに

様々な雑音が重畳する実環境において目的とする音源信号のみを取得するための枠組みとして、マイクロホンアレーを用いた音源分離技術の研究が行われている。観測信号のみから元信号を推定するブラインド音源分離の枠組みに対し、近年は音源信号の先験情報を積極的に利用することで分離性能を向上させることを目的とした枠組みも試みられてきている [1, 2, 3]。

本研究では、音源のパワーの時間変化（以下ではアクティベーション情報と記す）を利用することによるブラインド音源分離の性能向上の枠組みについて論じる。アクティベーション情報は、1) カメラや他のセンサの情報により取得する、2) 録音信号に対する分離を行う際に人間が分離信号を聴きながら与える、といった場面を想定している。以下では問題設定として、1) 音源のアクティベーション情報が既知、2) マイクロホン数 $\geq$ 音源数といった条件下での定式化とその解法、並びに実験結果について述べる。

2 問題の定式化

時間周波数表現を用いて観測信号 $\mathbf{X}_{\tau\omega}$ に対し、復元信号を

$$\mathbf{Y}_{\tau\omega} = W_\omega \mathbf{X}_{\tau\omega} \tag{1}$$

とモデル化し、分離行列 W_ω の推定を考える。

本研究では音源のアクティベーション情報を利用できることを仮定しているが、各音源のパワーの時間変化を正確に与えることは一般に困難であるため、これを確率的な分布として与える方が妥当であると考えられる。よってここでは、従来のブラインド信号処理の枠組みも参考に音源信号に適切な確率モデルを仮定し、音源のアクティベーション情報を事前分布として扱い、事後確率最大化により分離行列 W_ω を推定する枠組みを考える。

まず音源信号のモデルは、時変性を表現することが必要となる。ここでは音源信号を周波数方向にまとめたベクトル列 $\tilde{\mathbf{Y}}_{m\tau} = [Y_{m\tau 1}, \cdots, Y_{m\tau\Omega}]^T$ が複素ガウス分布

$$p_y(\tilde{\mathbf{Y}}_{m\tau}) \propto \frac{1}{\sigma^2_{m\tau}} \exp\left\{ -\frac{|\tilde{\mathbf{Y}}_{m\tau}|^2}{\sigma^2_{m\tau}} \right\} \tag{2}$$

に従うことを仮定する。これは独立ベクトル分析 [4, 5] において、音源の確率密度分布に時変ガウス分布を想定したことに等しい。

式 (2) のモデルにおいては、分散 $\sigma^2_{m\tau}$ が音源 m の時間フレーム τ におけるパワーを表しているので、音源のアクティベーション情報としては、$\sigma^2_{m\tau}$ の事前分布の形で与えるのが自然であると考えられる。ここでは計算の簡便さから、ガウス分布の分散の共役分布である逆ガンマ分布により、

$$p_{\sigma^2}(\sigma^2_{m\tau}) \propto \left(\frac{1}{\sigma^2_{m\tau}} \right)^{\frac{\alpha}{1-\alpha}} \exp\left\{ -\frac{\alpha \bar{\sigma}^2_{m\tau}}{(1-\alpha)\sigma^2_{m\tau}} \right\} \tag{3}$$

のように事前分布をモデル化する。ここで、音源のアクティベーション情報として事前分布の最頻値 $\bar{\sigma}^2_{m\tau}$ を与えた。また α は分布の広がりを表すパラメータで $0 \leq \alpha < 1$ である。

3 事後確率最大化による分離行列の推定

3.1 目的関数

対数事後確率 $J_1 = \log p(\mathbf{\Sigma}, \mathbf{W}|\mathbf{X})$ の最大化は対数尤度と対数事前確率の和

$$J_2 = \sum_{m,\tau} \log p_x(\tilde{\mathbf{X}}_{m\tau}|\mathbf{\Sigma}, \mathbf{W}) + \log p(\mathbf{\Sigma}, \mathbf{W}) \tag{4}$$

の最大化と等価である。ここで、$\mathbf{X} = \{\mathbf{X}_{m\tau}\}, \mathbf{\Sigma} = \{\sigma^2_{m\tau}\}, \mathbf{W} = \{W_\omega\}$ である。$\mathbf{\Sigma}$ は $\mathbf{W}$ と独立と仮定できるので、J_2 の最大化は

$$\begin{aligned} J_3 &= \sum_{m,\tau} \left\{ \log p_x(\tilde{\mathbf{X}}_{m\tau}|\mathbf{\Sigma}, \mathbf{W}) + \log p_{\sigma^2}(\sigma^2_{m\tau}) \right\} & (5) \\ &= \sum_{m,\tau} \left\{ \log p_y(\tilde{\mathbf{Y}}_{m\tau}|\mathbf{\Sigma}, \mathbf{W}) + \log p_{\sigma^2}(\sigma^2_{m\tau}) \right\} \\ &\quad + T \sum_\omega \log\det|W_\omega| & (6) \end{aligned}$$

の最大化と等価である。ここで T はフレーム数である。式 (2), (3) を代入すれば

$$\begin{aligned} J_3 &= -\sum_{m,\tau} \left\{ \frac{1}{1-\alpha} \log \sigma^2_{m\tau} + \frac{(1-\alpha)|\tilde{\mathbf{Y}}_{m\tau}|^2 + \alpha\bar{\sigma}^2_{m\tau}}{(1-\alpha)\sigma^2_{m\tau}} \right\} \\ &\quad + \sum_\omega \log\det|W_\omega| \end{aligned} \tag{7}$$

と表せる。

3.2 分散の更新

分散についての更新は $\frac{\partial J_3}{\partial \sigma^2_{m\tau}} = 0$ を解いて

$$\sigma^2_{m\tau} \leftarrow (1-\alpha)|\tilde{\mathbf{Y}}_{m\tau}|^2 + \alpha\bar{\sigma}^2_{m\tau} \tag{8}$$

を得る。これは、復元信号のパワーと与える分散の重み付き和となっている。

---

*Independent vector analysis with prior distribution exploiting a source activation for source separation. by ONO Takuma (The University of Tokyo), ONO Nobutaka (National Institute of Informatics), SAGAYAMA Shigeki (The University of Tokyo)

Table 1　シミュレーション実験の条件

sources	TIMIT database
impulse responses	RWCP Sound Scene Database[8]
reverberation time	300 ms (E2A), 470 ms (JR2)
sampling rate	16 kHz
data length	5 s
frame length	4096 points (256 ms)
frame shift	1024 points (64 ms)
Evaluative criterion	SDR (Signal-to-Distortion)

3.3　分離行列の更新

W_ω について対数尤度関数は

$$J_3 = -\sum_\omega \left(\sum_{m,\tau} \frac{|\mathbf{w}_{m\omega}^H \mathbf{X}_{\tau\omega}|^2}{\sigma_{m\tau}^2} - T \log\det|W_\omega| \right) + C \tag{9}$$

と書ける。ここで、$\mathbf{w}_{m\omega}$ は W_ω の m 行目であり、C は W_ω に依らない定数である。J_3 の最大化は解析的には解けないが、W_ω を一行ごとに更新するアルゴリズム [6, 7] により

$$V_{m\omega} \quad = \quad \frac{1}{T}\sum_\tau \left(\frac{\mathbf{X}_{\tau\omega}\mathbf{X}_{\tau\omega}^H}{\sigma_{m\tau}^2} \right) \tag{10}$$

$$\mathbf{w}_{m\omega} \quad \leftarrow \quad (W_\omega V_{m\omega})^{-1}\mathbf{e}_m \tag{11}$$

$$\mathbf{w}_{m\omega} \quad \leftarrow \quad \mathbf{w}_{m\omega} / \sqrt{\mathbf{w}_{m\omega}^H V_{m\omega} \mathbf{w}_{m\omega}} \tag{12}$$

と効率的に最適化できる。ここで $\mathbf{e}_m$ は m 列目にのみ 1 で他が 0 の値を持つ単位ベクトル $\mathbf{e}_m = [0, \cdots, 1, \cdots, 0]^T$ である。この手法は、自然勾配法による最適法と比べパラメータチューニングの必要がなく、収束が速いことが知られている [6]。

分離行列 W_ω は結局、式 (8), (10), (11), (12) を各周波数ごとに反復的に更新する。

4　シミュレーション実験

4.1　実験目的および条件

提案手法の有用性を示す目的で、シミュレーションによる実験を行った。Table 1 に実験条件を示す。予め録音された残響時間の異なる二種類の室内インパルス応答 [8] をクリーン音声とそれぞれ畳み込み加算することで観測信号とした。音源数 2 とし、音源の到来方向と話者を変化させ実験した。またマイクロホン数 4 とし、28.3 mm の等間隔線形アレーを用いた。部屋 E2A(残響時間 T_{60}=300 ms) では、$\{-80°, -40°, -20°, 20°, 60°\}$ の 5 方向から 2 方向を全ての組み合わせについて 10 通りを試し、部屋 JR2(T_{60}=470 ms) では、$\{-30°, -10°, 10°, 20°\}$ の 4 方向から 2 方向を全ての組み合わせについて 6 通りを試した。また音声として男声 2、女声 3 の異なる 5 話者から 2 話者を全ての組み合わせについて 10 通りとした。総合的な分離性能を測る指標として SDR(Signal-to-distortion Ratio)[9] を算出し、E2A, JR2 のそれぞれの部屋について 100 または 60 パターンの平均を定量的な分離性能とした。

提案法の事前分布のパラメータである式 (3) の α は 0.2, 0.5, 0.8 の 3 パターンについて試した。今回

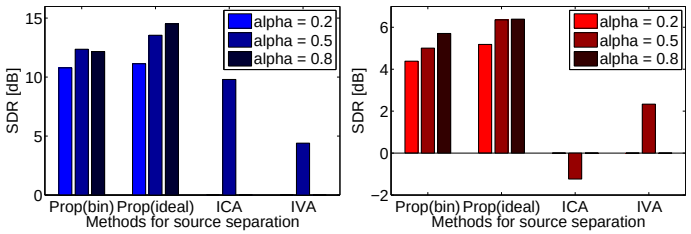

Fig. 1　SDR による部屋 E2A(T_{60}=300 ms, 左図) と部屋 JR2(T_{60}=470 ms, 右図) における分離性能比較 (提案法 (Prop) は左から α=0.2, 0.5, 0.8)

は、提案法で用いる $\tilde{\sigma}_{m\tau}^2$ を正解データから算出した。加算する前のインパルス応答を畳み込んだそれぞれの元信号のパワーの時間変化を 1) 閾値で二極化したアクティベーションパターン (bin) 2) そのままを用いた理想的なアクティベーションパターン (ideal) の二種類を実験に用いた。閾値で二極化した条件 (bin) についてはユーザが十分に与えられうる、またはカメラを用いて取得できるパターンであるとし与える事前情報として妥当であると考えた。

比較として独立成分分析による手法 (ICA)[10] と復元信号の生成モデルにラプラス分布を仮定した独立ベクトル分析による手法 (IVA)[5] の SDR 平均値を算出した。全ての手法に共通して前処理に主成分分析による次元圧縮と白色化を行った。

4.2　実験結果

実験結果を Fig. 1 に示す。一般に独立ベクトル分析はデータ長が短い場合に十分な分離性能を得られないが、提案法は事前確率を導入したことで ICA と比べても優位な結果が得られ、高残響下でも性能向上を実現した。

5　まとめと今後の展望

本研究では、アクティベーション情報を復元信号の分散の事前分布とし、事後確率最大化独立ベクトル分析による音源分離を行った。正解データから音源の大まかなアクティベーションを与えたシミュレーション実験により、提案手法に 3dB 程度の性能向上が見られ有用性を確認できた。今後の展望として、動画中の唇の動きを用いて音声のアクティベーションを与え、より応用に近い実験を行う予定である。

参考文献

[1] M. Parvaix *et al.*, *Proc. ICASSP*, pp.245–248, 2010.
[2] R. Hennequin *et al.*, *Proc. ICASSP*, pp.45–49, 2011.
[3] A. Ozerov *et al.*, *Proc. ICASSP*, pp.257–260, 2011.
[4] A. Hiroe, *Proc. ICA*, pp.601–608, 2006.
[5] T. Kim *et al.*, *Proc. ICA*, pp.165–172, 2006.
[6] N. Ono *et al.*, *Proc. LVA/ICA*, pp.165–172, 2010.
[7] A. Yeredor, *Proc. CAMSAP*, pp. 312–315, 2009.
[8] S. Nakamura *et al.*, *Proc. LREC*, pp. 965–968, 2000.
[9] E. Vincent *et al.*, *Trans. ASLP*, pp.1462–1469, 2006.
[10] H. Sawada *et al.*, *Proc. ICASSP*, pp. 381–384, 2003.