7N-05

逐次的音源検出と高速マルチチャネル非負値因子分解に基づくオンライン音源分離

藤原 啓梧[1]　関口 航平[3,2]　Yicheng Du[2]　吉井 和佳[2,3]

[1]京都大学 工学部情報学科　[2]京都大学 大学院情報学研究科　[3]理化学研究所 AIP

1. はじめに

マルチチャネル音源分離は，複数の音源が存在しうる実環境でリアルタイム音声インタラクションを行うための基盤技術であり，遅延の小さいオンライン処理が求められている．近年，マイクアレイや周囲の環境に関する事前情報を用いずに，観測信号のみから音源信号を推定する汎用的なブラインド音源分離 (BSS) 手法として，高速マルチチャネル非負値行列因子分解 (FastMNMF) [1,2] が注目されている．本手法は本来，オフライン処理を行うよう設計されており，パラメータを最適化する際に，観測信号全体を繰り返し参照する必要がある．しかしながら，高い分離精度を達成しつつも，高速に動作する点で，オンライン処理に拡張するのに適している．

本研究では，マイク配置が既知のマイクアレイを用いたオンライン音源分離に取り組む．これは完全なブラインド条件ではないが，通常，マイクアレイの形状は変化することはないので，実用上の制約にはならない．オンライン分離においては，音源の出現や消失への対処が必要になるが，FastMNMF のパラメータを単純に逐次更新してしまうと，音源が新たに出現した際に全体が大幅に更新され，分離が不安定になる．

この問題を解決するため，本研究では，MUSIC 法に基づく音源定位を前段に用いて，新規音源を検出した際には，既存音源に関するパラメータを固定した上で新規音源に関するパラメータを推定したのち，全体を逐次更新する方法を提案する．本手法の肝は，アクティブでない分離音に対応するステアリングベクトルを，新規音源方向に対して幾何的に計算されるステアリングベクトルで初期化したうえで，限定的に更新する点にある．しかし，FastMNMF は，一種の分離行列，すなわち，各音源に対応するステアリングベクトルを並べて得られる混合行列の逆行列に対して更新則の導出がなされており，特定の音源に着目した限定的な更新は容易ではない．そこで，独立ベクトル分析 (IVA) のために最近提案された反復音源ステアリング (ISS) 法 [3] を用いて，混合行列中の特定のステアリングベクトルに限定して更新を行うのと等価な，分離行列全体に対する更新を行う．

2. FastMNMF

FastMNMF は，NMF に基づく音源モデルと同時対角化制約付きフルランク空間モデルを統合した BSS 手法である．いま，音源数を N，マイク数を M，周波数ビン数を F，フレーム数を T とする．また，$\{\mathbf{x}_{nft} \in \mathbb{C}^M\}_{f,t=1}^{F,T}$ を音源 n の多チャネル複素スペクトログラム（イメージ），$\{\mathbf{x}_{ft} \in \mathbb{C}^M\}_{f,t=1}^{F,T}$ を観測される混合音の多チャネル複素スペクトログラムとする．まず，音源イメージが複素ガウス分布に従うことを仮定する．

$$\mathbf{x}_{nft} \sim \mathcal{N}_{\mathcal{C}}\left(\mathbf{0}_M, \lambda_{nft}\mathbf{G}_{nf}\right) \tag{1}$$

ここで，$\lambda_{nft} > 0$ は音源 n の周波数 f，フレーム t におけるパワースペクトル密度，$\mathbf{G}_{nf} \in \mathbb{C}^{M\times M}$ は音源 n の周波数 f における空間相関行列を表す．FastMNMF では，$\{\lambda_{nft}\}_{f,t=1}^{F,T}$ は低ランク構造を持つことを仮定する．

$$\lambda_{nft} = \sum_{k=1}^{K} w_{nkf} h_{nkt} \tag{2}$$

ここで，w_{nkf} および h_{nkt} はそれぞれ，各音源 n，基底 k の周波数 f におけるパワースペクトル密度およびフレーム t におけるアクティベーションである．一方，$\{\mathbf{G}_{nf}\}_{n,f=1}^{N,F}$ は，正則行列 $\mathbf{Q}_f = [\mathbf{q}_{f1}, \ldots, \mathbf{q}_{fM}]^{\mathsf{H}}$（一種の分離行列）により同時対角化可能であると仮定する．

$$\mathbf{G}_{nf} = \mathbf{Q}_f^{-1}\mathrm{Diag}(\tilde{\mathbf{g}}_n)\,\mathbf{Q}_f^{-\mathsf{H}} \tag{3}$$

ここで，$\tilde{\mathbf{g}}_n = [\tilde{g}_{n1}, \ldots, \tilde{g}_{nM}]$ は M 次元の非負値ベクトルである．$\mathbf{x}_{ft} = \sum_{n=1}^{N} \mathbf{x}_{nft}$ は，次式で与えられる．

$$\mathbf{x}_{ft} \sim \mathcal{N}_{\mathcal{C}}\left(\mathbf{0}_M, \mathbf{Q}_f^{-1}\left(\sum_{n=1}^{N} \lambda_{nft}\mathrm{Diag}(\tilde{\mathbf{g}}_n)\right)\mathbf{Q}_f^{-\mathsf{H}}\right) \tag{4}$$

FastMNMF では，式 (4) で与えられる尤度関数を最大化するよう $\mathbf{W} \triangleq \{w_{nfk}\}_{n,k,f=1}^{N,K,F}$, $\mathbf{H} \triangleq \{h_{nkt}\}_{n,k,t=1}^{N,K,T}$, $\tilde{\mathbf{G}} \triangleq \{\tilde{\mathbf{g}}_n\}_{n=1}^{N}$, $\mathbf{Q} \triangleq \{\mathbf{Q}_f\}_{f=1}^{F}$ を求める．具体的には，乗法更新則および反復射影 (IP) 法に基づく収束保証付きの最適化が可能である（更新式は省略）．

3. 提案法

本章では，マイク配置から幾何的に計算されるステアリングベクトルを用いて，検出された新規音源に安定して対応する方法を提案する．

3.1 FastMNMF のオンライン拡張

FastMNMF のオンライン処理では，ミニバッチ（T フレームで構成）を 1 フレームずつずらしながら，逐次的にパラメータを更新することが基本となる．各ミニバッチで $\mathbf{W}, \tilde{\mathbf{G}}, \mathbf{Q}$ は前のミニバッチのものを引き継ぎ，$\mathbf{H}$ はフレームシフトを行い ($h_{nkt} \leftarrow h_{n,k,t+1}$ $(0 \le t < T)$)，新たなフレームに対応する部分 (h_{nkT}) はランダムに初期化したうえで，パラメータ全体の更新を行う．

FastMNMF では，式 (4) より，$\mathbf{Q}_f\mathbf{x}_{ft}$ の各要素は独立となることから，$\mathbf{Q}_f$ および $\mathbf{Q}_f^{-1}$ はそれぞれ ICA における分離・混合行列と同様の働きを持ち，$\mathbf{Q}_f^{-1}$ の列ベ

Online Source Separation Based on Sequential Source Detection and FastMNMF: *Keigo Fujiwara (Kyoto Univ.), Kouhei Sekiguchi (RIKEN/Kyoto Univ.), Yicheng Du (Kyoto Univ.), Kazuyoshi Yoshii (Kyoto Univ./RIKEN)*

クトルは一種のステアリングベクトルとみなせる．設定した N が実際の音源数より多ければ，少なくとも 1 つはアクティブでない音源が存在することに注意する．

3.2 音源定位に基づく新規音源への対応

提案法では，まず，分離処理の前段として，10 フレームごとに MUSIC 法による音源検出・定位を行う．新規音源を検出すると，当該方向に対応する幾何的なステアリングベクトルを，$\mathbf{Q}_f^{-1}$ のある列（i 列目とする）に挿入し，推定済みの既存音源の情報を利用しつつ，集中的に更新を行う．ここで，挿入先 i の選び方は性能に影響するので，慎重な検討が必要である．本研究では，検出された音源方向と挿入先の列のインデックスの組をそのつど記録しておき，未検出の期間が最も長い音源方向に対応する列のインデックス i を選ぶことにした．

FastMNMF [2] では，もともと IP 法を用いて $\mathbf{Q}_f$ を行ベクトル単位ですべて更新していたのに対し，提案法では，ISS 法 [3] を用いて $\mathbf{Q}_f^{-1}$ の特定の列ベクトル $\tilde{\mathbf{q}}_{fi}$ のみを更新する．具体的には，幾何的なステアリングベクトルの挿入したのち，$\mathbf{W}, \mathbf{H}, \tilde{\mathbf{G}}$ のうちで i に対応する部分を初期化し，$\tilde{\mathbf{q}}_{fi}$ と交互に更新する．ただし，$\tilde{\mathbf{g}}_i$ は $\tilde{\mathbf{q}}_{fi}$ に対する重みが大きくなるよう $[\epsilon, \ldots, 1, \ldots, \epsilon]$ で初期化する $(\epsilon = 0.01)$．ISS 法に基づく更新式は，AuxIVA と同様に次式で与えられる．

$$\mathbf{Q}_f \leftarrow \mathbf{Q}_f - \mathbf{u}_{fi}\mathbf{q}_{fi}^{\mathsf{H}} \tag{5}$$

式 (4) で定まる対数尤度関数に代入し，$\mathbf{u}_{fi}$ について偏微分が $\mathbf{0}$ となる $\mathbf{u}_{fi}^*$ を求めると次式を得る．

$$\mathbf{u}_{fi}^* = \begin{cases} \dfrac{\mathbf{q}_{fm}^{\mathsf{H}}\mathbf{V}_{fm}\mathbf{q}_{fi}}{\mathbf{q}_{fi}^{\mathsf{H}}\mathbf{V}_{fm}\mathbf{q}_{fi}} & (m \neq i) \\ 1 - (\mathbf{q}_{fm}^{\mathsf{H}}\mathbf{V}_{fm}\mathbf{q}_{fm})^{-\frac{1}{2}} & (m = i) \end{cases} \tag{6}$$

この更新は $\tilde{\mathbf{q}}_{fi}$ のみを更新するのと等価である．

4. 評価実験

本章では，比較実験について報告する．

4.1 実験設定

pyroomacoustics ライブラリ [4] を用いて，チャネル数 $M = 8$ の半径 1cm の円環マイクアレイと 3 個の音源とを図 1 の通り配置し，wsj0 [5] から選んだ 3 話者 8 発話をランダムに重畳することで，30 秒程度の混合音を 5 パターン作成した．残響時間は 137 ms，サンプリングレートは 16 kHz，短時間フーリエ変換の窓幅は 1024 サンプル，シフト長は 256 サンプルとした．マイクアレイの周囲 360° を 5° 刻みで 72 個のステアリングベクトルを用意した．MUSIC 法では，音声を想定した 250 Hz〜3200 Hz を用いて音源を定位し，±10° に既存音源が存在しない場合に新規音源が検出されたと判定した．提案法を，オフライン FastMNMF [2]（反復回数 320 回），単純な逐次更新に基づくオンライン FastMNMF（ベースライン）と比較した．各音源が存在する区間での Signal-to-Distortion Ratio(SDR) [6] を評価した．バッチサイズは 200 フレームとし，基底数 $K = 8$，音源数 $N = 8$，パラメータ更新は部分的に 10 回，全体で 1 回とした．

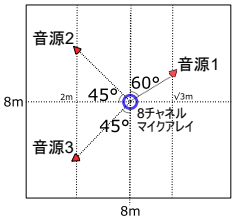

図 1: マイクアレイ・音源の配置

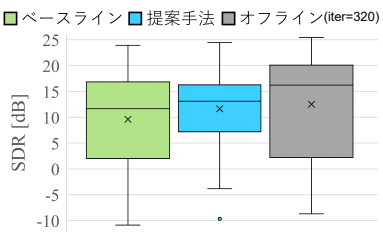

図 2: 各手法の SDR

4.2 実験結果

図 2 に示す通り，ベースライン法は SDR の分散が大きく，新規音源に適切に対応できない場合があったのに対して，提案法はステアリングベクトルを用いた $\mathbf{Q}_f^{-1}$ の初期化により，比較的安定して分離を行うことができた．平均的にも，提案法はベースライン法より優れていた．

本手法では，MUSIC 法で新たな音源が検出された際に，当該方向のステアリングベクトルを $\tilde{\mathbf{q}}_{fi}$ に挿入し，$\tilde{\mathbf{g}}_i$ を $\tilde{\mathbf{q}}_{fi}$ に対する重みが大きくなるよう初期化するため，i 番目の分離音が新規音源に対応することが期待される．しかし実際には，分離は正しくできているものの，分離音と新規音源の対応関係が期待と異なることがあった．パラメータの更新順序は任意性が大きく，さらなる改善が必要であると考えられる．

5. まとめ

本稿では，音源定位・検出と FastMNMF を用いたオンライン音源分離を提案した．本手法は，検出した新規音源に対応するステアリングベクトルを幾何的に初期化し，ISS 法を用いて集中的に更新したあとで，全体の最適化を行う．実験では，提案するパラメータ更新により，安定的したオンライン推定を実現できることを確認した．今後は，長時間録音に対する頑健性，リアルタイム処理に向けた計算量の削減に取り組む予定である．

謝辞 本研究の一部は，JSPS 科研費 Nos. 19H04137, 20K21813, 20H00602, 20H01159 の支援を受けた．

参考文献

[1] N. Ito and T. Nakatani. FastMNMF: Joint diagonalization based accelerated algorithms for multichannel nonnegative matrix factorization. In *ICASSP*, pages 371–375, 2019.

[2] K. Sekiguchi et al. Fast multichannel nonnegative matrix factorization with directivity-aware jointly-diagonalizable spatial covariance matrices for blind source separation. *IEEE/ACM TASLP*, 28:2610–2625, 2020.

[3] R. Scheibler and N. Ono. Fast and stable blind source separation with rank-1 updates. In *ICASSP*, pages 236–240, 2020.

[4] R. Scheibler et al. Pyroomacoustics: A python package for audio room simulation and array processing algorithms. In *ICASSP*, pages 351–355, 2018.

[5] J. Garofolo et al. CSR-I (WSJ0) complete. 2007.

[6] E. Vincent et al. Performance measurement in blind audio source separation. *IEEE TASLP*, 14(4):1462–1469, 2006.