同期同時対角化音源分離におけるアクティベーション行列を用いた更新回数の最適化*

☆泉太貴 (大分大), 太刀岡勇気 (デンソーアイティーラボラトリ), 上ノ原進吾, 古家賢一 (大分大)

1 はじめに

現在、様々な音源分離手法が提案されている。例えば、各音源が互いに独立であると考えることで分離を行う独立成分分析 (Independent Component Analysis:ICA)[1]、音源信号の非ガウス性を活用する独立ベクトル分析 (Independent Vector Analysis : IVA)[2]、頻出する周波数特徴ごとに分離を行う非負値行列因子分解 (Nonnegative Matrix Factorization : NMF)[3]、NMF を多チャネル拡張し、周波数情報に加え空間情報を利用することにより高性能な分離が行えるようになったマルチチャネル NMF(Multichannel NMF : MNMF)[4]、IVA と MNMF を組み合わせた独立低ランク分析 (Independent Low-Rank Matrix Analysis : ILRMA)[5] などが挙げられる。ICA、IVA、ILRMA は優決定条件 (音源数 ≤ チャネル数) における音源分離であり、NMF は劣決定条件 (音源数 > チャネル数) の音源分離となる。MNMF は優決定、劣決定のどちらの条件でも利用できる。そして、音源信号の非定常性を活用する新しいブラインド音源分離手法として、複数の時間区間の相関行列の同時対角化を行いつつ、同じ信号源に対する対角成分を時間的に同期させる同期同時対角化 (Synchronized Joint Diagonalization : SJD)[6] が提案された。SJD は IVA と似た手法でありながら、パーミュテーション問題をより正しく揃え、性能は IVA に優ると報告されている。そして、これらの音源分離は、反復更新アルゴリズムによって達成される。この中でも SJD は新しい音源分離手法であるため、更新回数がどの程度必要であるかわかっていない。

本稿では、SJD を対象に、アクティベーション行列 $\boldsymbol{V}$ の推定の際の差分を利用した更新回数の最適化手法を提案する。そして、その有効性を楽器音による音源分離実験により評価した。

2 SJD によるブラインド音源分離

2.1 概要

信号の非定常性を活用するブラインド音源分離手法として、複数の時間区間の相関行列の同時対角化 (Joint Diagonalization : JD)[7] が提案

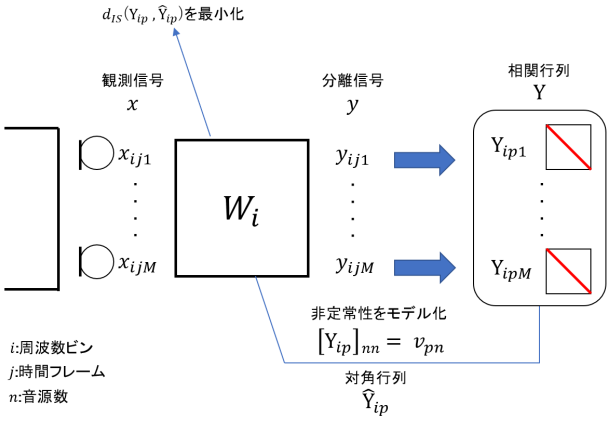

Fig. 1 SJD による音源分離

されており、SJD[6] は JD が複数の同時対角化問題を解く際に、同じ信号源に対応する対角成分を時間的に同期させるようにした音源分離手法である。図 1 に SJD による音源分離の原理を示す。

2.2 定式化

各マイクチャネル $m = 1, \cdots, M$ の観測信号に短時間フーリエ変換を行い、時間周波数表現 x_{ijm} を得る。$i = 1, \cdots, I$ は周波数ビン、$j = 1, \cdots, J$ は時間フレームを表す。観測信号が独立な音源信号 $s_{ijn} = [\boldsymbol{s}_{ij}]_n, n = 1, \cdots, N$ の線形混合であると仮定することで、$M \times N$ の混合行列 $\boldsymbol{A}_i$ が定義でき、観測信号は以下の式で表現できる。

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij} \tag{1}$$

ブラインド音源分離の目的は、観測信号のみから $N \times M$ の分離行列 $\boldsymbol{W}_i$ を周波数ビン毎 $i = 1, \cdots, I$ に求め、できるだけ源信号に近い分離信号 $y_{ijn} = [\boldsymbol{y}_{ij}]_n, n = 1. \cdots, N$ を以下の式で推定することである。

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij} \tag{2}$$

2.3 同時対角化

相関行列の同時対角化は周波数ビン i 毎に処理を行う。J 個の時間フレームを P 個の時間区間 $\mathcal{J}_p (p = 1, \cdots, P)$ に分割し、各時間区間 p で観測信号の相関行列 $\boldsymbol{X}_{ip}$ および分離信号の相関行列 $\boldsymbol{Y}_{ip}$ を以下の式で求める。

*Optimization of Number of Updates using Activation Matrix of Sound Source Separation by Synchronized Joint Diagonalization by Taiki Izumi(Oita University), Yuuki Tachioka(Denso IT Laboratory), Shingo Uenohara, Ken'ichi Furuya(Oita University)

$$\boldsymbol{X}_{ip} = \frac{1}{P}\sum_{j\in\mathcal{J}_p} \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^H \tag{3}$$

$$\boldsymbol{Y}_{ip} = \frac{1}{P}\sum_{j\in\mathcal{J}_p} \boldsymbol{y}_{ij}\boldsymbol{y}_{ij}^H = \boldsymbol{W}_i\boldsymbol{X}_{ip}\boldsymbol{W}_i^H \tag{4}$$

H はエルミート転置である。そして、分離信号の P 個の相関行列 $\boldsymbol{Y}_{ip}$ を同時に対角化する分離行列 $\boldsymbol{W}_i$ を求める。$P = 2$ の場合は、厳密に対角化することができるが、$P \geq 3$ になると、一般的に厳密解を求めることができない。

2.4 SJD

SJD には、$\boldsymbol{Y}_{ip}$ と対角行列 $\hat{\boldsymbol{Y}}_{ip}$ との多チャンネル板倉斎藤距離を最小化する。

$$d_{IS}(\boldsymbol{Y}_{ip}, \hat{\boldsymbol{Y}}_{ip}) = \mathrm{tr}(\boldsymbol{Y}_{ip}\hat{\boldsymbol{Y}}_{ip}^{-1}) - \log\det\boldsymbol{Y}_{ip}\hat{\boldsymbol{Y}}_{ip}^{-1} - N \tag{5}$$

信号源が持つ非定常性をモデル化するために、対角行列 $\hat{\boldsymbol{Y}}_{ip}$ を次式で表す。

$$[\hat{\boldsymbol{Y}}_{ip}]_{nn'} = \begin{cases} v_{pn} & \text{if } n = n' \\ 0 & \text{if } n \neq n' \end{cases} \tag{6}$$

ここで、右辺は時間区間 p と信号 n のみに依存し周波数 i には依存しない。以後、定式化のため $\boldsymbol{V}$ をサイズ $P \times N$ の行列とし、$v_{pn} = [\boldsymbol{V}]_{pn}$ とする。SJD は以下のすべての周波数をまとめた総和が最小化するコスト関数である。

$$\mathcal{C} = \sum_{i=1}^{I}\sum_{p=1}^{P}\left[\sum_{n=1}^{N}\left(\frac{[\boldsymbol{Y}_{ip}]_{nn}}{v_{pn}} + \log v_{pn}\right) - 2\log|\det\boldsymbol{W}_i|\right] \tag{7}$$

2.5 音源分離アルゴリズム

分離行列 $\boldsymbol{W}$ とアクティベーション行列 $\boldsymbol{V}$ を交互に更新する。式 (7) を v_{pn} で微分して 0 とおくことで、アクティベーション行列 $\boldsymbol{V}$ に関する以下の更新式が導出される。

$$v_{pn} = \frac{1}{I}\sum_{i=1}^{I}[\boldsymbol{Y}_{ip}]_{nn} \tag{8}$$

次に周波数毎の分離行列 $\boldsymbol{W}_i$ は以下の手順で更新する。

$$\boldsymbol{U}_{in} = \frac{1}{P}\sum_{p=1}^{P}\frac{1}{v_{pn}}\boldsymbol{X}_{ip} \tag{9}$$

これから N 個の行列 $\boldsymbol{U}_{in}$ をハイブリッド同時対角化する行列として $\boldsymbol{W}_i$ を以下のように更新する。

$$\boldsymbol{w}_{in} = (\boldsymbol{W}_i\boldsymbol{U}_{in})^{-1}\boldsymbol{e}^n \tag{10}$$

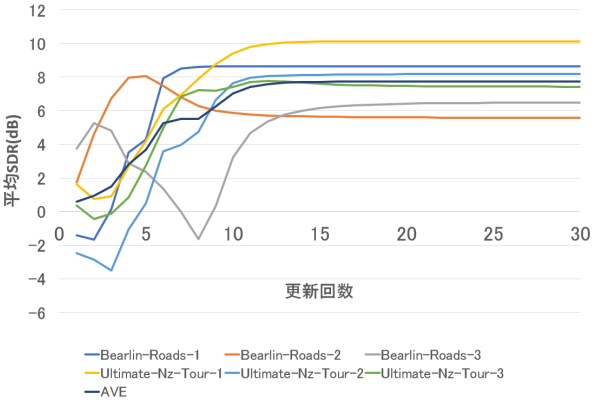

Fig. 2　各更新回数での分離性能 ($P = 404$)

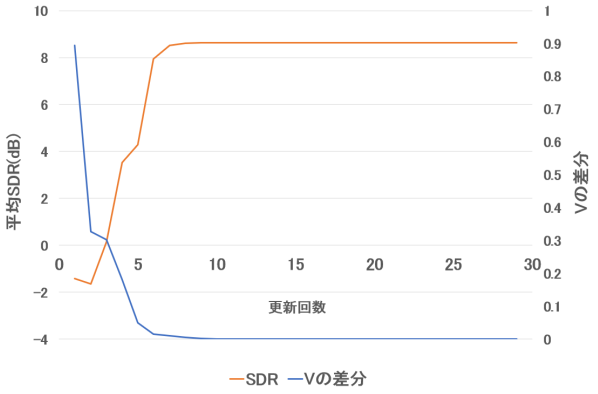

Fig. 3　$\boldsymbol{V}$ の差分と SDR の関係

ここで、$\boldsymbol{e}^n$ は n 行目だけが 1 になる N 次元のベクトルである。そして、以下の式により正規化を行う。

$$\boldsymbol{w}_{in} = \frac{\boldsymbol{w}_{in}}{\sqrt{\boldsymbol{w}_{in}^H\boldsymbol{U}_{in}\boldsymbol{w}_{in}}} \tag{11}$$

3 提案手法

3.1 着眼点

図 2 に更新回数と平均 SDR の結果を示す。また、図 3 には楽曲での、1 から 100 回までの各更新回数での SDR と $\boldsymbol{V}$ の差分を示す。この結果、10 数回の更新で $\boldsymbol{V}$ の差分に変化がなくなり、SDR も同様に変化しなくなることを確認した。また、アクティベーション行列 $\boldsymbol{V}$ の差分と SDR には負の相関があるということが確認できた。これによりアクティベーション行列 $\boldsymbol{V}$ が変化しなくなったタイミングで更新を終了することで無駄な処理時間を削減することができ、音源ごとに最適な更新回数で処理することが可能になると考えた。

3.2 更新回数と SDR の関係

ここでは、SJD の更新回数と分離性能の関係を調査する。実験には表 1 の音楽データを使用

している。実験結果は図 2 に示すとおりであり、数十回の更新で分離が行えていることが確認できる。文献 [6] においても 30 回程度の更新で分離を達成できることが示されている。しかしながら、一般的に、更新回数は処理前に設定するため、音源ごとに適した更新回数がどの程度であるかわからない。

3.3 アクティベーション行列 $\boldsymbol{V}$ を用いた更新回数の最適化

更新回数の最適化には、アクティベーション行列 $\boldsymbol{V}$ の更新前と更新後の推定値の差分を利用する。繰り返し q 回目の $\boldsymbol{V}$ を $\boldsymbol{V}^q$ として、以下のように判定する。

$$\left|\boldsymbol{V}^{q-1} - \boldsymbol{V}^{q}\right| < \epsilon \left|\boldsymbol{V}^{1} - \boldsymbol{V}^{2}\right|$$

$$\left|\frac{\boldsymbol{V}^{q-1}}{max(\boldsymbol{V}^{q-1})} - \frac{\boldsymbol{V}^{q}}{max(\boldsymbol{V}^{q})}\right| < \epsilon \left|\frac{\boldsymbol{V}^{1}}{max(\boldsymbol{V}^{1})} - \frac{\boldsymbol{V}^{2}}{max(\boldsymbol{V}^{2})}\right|$$

$$\left|\frac{\boldsymbol{V}^{q-1}}{norm(\boldsymbol{V}^{q-1})} - \frac{\boldsymbol{V}^{q}}{norm(\boldsymbol{V}^{q})}\right| < \epsilon \left|\frac{\boldsymbol{V}^{1}}{norm(\boldsymbol{V}^{1})} - \frac{\boldsymbol{V}^{2}}{norm(\boldsymbol{V}^{2})}\right|$$

差分が 1 回目と 2 回目の差分の ϵ 以下になった段階で処理を終了する。差分の取り方は上記のように絶対値で差分を計算した場合、最大値で正規化した場合、ノルムで正規化した場合を試したが、結果にほとんど差がなかったため、今回の実験では一番上の式を用いて判定を行う。

4 評価実験

4.1 実験条件

実験に用いる混合信号は音楽データとし、2 音源 2 マイクの環境で作成した。これを表 1 に示す。混合信号は各音源に対して、図 4 の環境で測定した RWCP 実環境音声・音響データベースのインパルス応答 (E2A)[9] を畳み込んで作成した。混合信号の長さは、文献 [6] に合わせ 6.4 秒にしている。図 4 においてマイクロホンは右から順に 1-14 まで番号がついており、今回の実験で利用したマイクロホン番号は 6 と 8 である。SJD のパラメータは表 2 に示す。時間区間数 P は最大の $P = 404$、$P = 202$ としている。プログラムは MATLAB 上で実装しており、Intel Corei7-4770 3.4Ghz で処理を行った。計算時間の評価には、音源の長さに対する処理時間の割合を表すリアルタイムファクタ (Real Time Factor : RTF) を用い、分離性能の評価には、SDR(Signal-to-Distortion Ratio)[10] を用いた。更新終了のための ϵ は 0.001 とした。

4.2 実験結果

表 3，4 には、それぞれ時間区間数 P が 404，202 における各楽曲での提案手法により最適化した更新回数、それに伴う分離性能および RTF を示してあり、表中の平均 SDR は各更新回数における ID1〜6 の SDR の平均を示している。これら 2 つの表から、提案手法によって最適化された更新回数の性能は 30 回程度の性能と同等であることが確認できる。そして、10 回では性能が不十分な場合があることが分かった。提案のアクティベーション行列 $\boldsymbol{V}$ の差分を用いることで更新回数が最適化されているのがわかる。加えて、RTF が 30 回更新した場合に比べて小さくなっており、リアルタイム性が向上していることが確

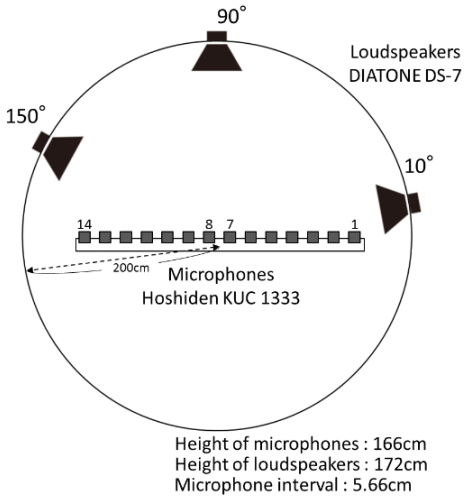

Fig. 4　音源とマイクロホンの配置

Table 1　使用した音楽データ

ID	Author/Song	Part
1	Bearlin-Roads-1	Piano(90°) Vocal(10°)
2	Bearlin-Roads-2	Vocal(10°) Ambient(150°)
3	Bearlin-Roads-3	Piano(90°) Ambient(150°)
4	Ultimate-Nz-Tour-1	Guitar(90°) Synth(10°)
5	Ultimate-Nz-Tour-2	Synth(10°) Drum(150°)
6	Ultimate-Nz-Tour-3	Guitar(90°) Drum(150°)

Table 2　SJD のパラメータ

残響時間	300ms
サンプリング周波数	16kHz
フレームサイズ	1024
シフトサイズ	256
音源数	2
マイク数	2
時間区間数 P	404，202

Table 3　$P = 404$ の場合の実験結果まとめ
(太字は提案手法による更新回数)

ID	更新回数	SDR	RTF	平均 SDR
1	30 回	8.64 dB	0.73	
	10 回	8.64 dB	0.092	
	12 回	8.64 dB	0.10	30 回
2	30 回	5.57 dB	0.73	7.73 dB
	10 回	5.86 dB	0.092	
	13 回	5.69 dB	0.11	
3	30 回	6.45 dB	0.73	
	10 回	3.22 dB	0.092	
	18 回	6.36 dB	0.14	10 回
4	30 回	10.12 dB	0.73	7.02 dB
	10 回	9.41 dB	0.092	
	16 回	10.11 dB	0.13	
5	30 回	8.19 dB	0.73	
	10 回	7.56 dB	0.092	
	15 回	8.11 dB	0.12	提案手法
6	30 回	7.42 dB	0.73	7.75 dB
	10 回	7.42 dB	0.092	
	16 回	7.56 dB	0.13	

Table 4　$P = 202$ の場合の実験結果まとめ
(太字は提案手法による更新回数)

ID	更新回数	SDR	RTF	平均 SDR
1	30 回	8.65 dB	0.17	
	10 回	8.65 dB	0.063	
	12 回	8.65 dB	0.075	30 回
2	30 回	5.71 dB	0.17	7.67 dB
	10 回	6.02 dB	0.063	
	13 回	5.83 dB	0.08	
3	30 回	6.60 dB	0.17	
	10 回	5.40 dB	0.063	
	16 回	6.45 dB	0.091	10 回
4	30 回	10.17 dB	0.17	7.36 dB
	10 回	9.48 dB	0.063	
	17 回	10.16 dB	0.097	
5	30 回	7.92 dB	0.17	
	10 回	7.45 dB	0.063	
	16 回	7.95 dB	0.094	提案手法
6	30 回	6.97 dB	0.17	7.69 dB
	10 回	7.18 dB	0.063	
	17 回	7.10 dB	0.1	

認できる。よって、提案手法により無駄な更新を回避し、結果として、計算時間の削減が行えている。今回は、楽器音のみでの実験であったが、非定常である音声の場合は異なる結果が得られる可能性が考えられる。

5　まとめと今後の展望

本稿では、新しい音源分離手法である SJD に着目し、各更新回数におけるアクティベーション行列 $\boldsymbol{V}$ の差分と SDR には負の相関があることを確認したことから、アクティベーション行列 $\boldsymbol{V}$ の推定による更新回数の最適化手法を提案した。提案手法により、無駄な更新を免れることができたため、楽曲ごとに最適な更新で処理を行え、処理時間の削減を達成した。今後の展望としては、多様な音源および環境で実験を行い、主に SJD の性能を向上させる手法を検討する。

参考文献

[1] T.-W. Lee, “Independent Component Analysis-Theory and Applications,” Norwell, MA: Kluwer, 1998.

[2] I. Lee *et al.*,“Fast fixedpoint independent vector analysis algorithms for convolutive blind source separation,” Signal Processing 87(8), 2007

[3] D.D. Lee *et al.*, “Learning the Parts of Objects with Nonnegative Matrix Factorization,” Nature, vol.401, pp.788-791, 1999.

[4] H. Sawada *et al.*, “Multichannel Extensions of Non-Negative Matrix Factorization with Complex-Valued Data,” IEEE Trans. ASLP, vol.21, no.5, pp.971-982, 2013.

[5] D. Kitamura *et al.*, “Dtermined blind source separation unifying independent vector analysis and nennegative matrix factorization,” IEEE/ACM Transaction on Audio, Speech, and Language Processing, 24(9):1626-1641, 2016

[6] 澤田 宏, “同期同時対角化によるブラインド信号分離,” 第 32 回信号処理シンポジウム講演論文集, pp.332-337, 2017

[7] A.Ziehe *et al.*,“A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation,” Journal of Machine Learning Research, pp777-800, 2004

[8] S. Araki *et al.*, “The 2011 Signal Separation Evaluation Campaign (SiSEC2011): - Audio Source Separation,” Latent Variable Analysis and Signal Separation(Springer, Bearlin, 2012), pp. 414-422.

[9] RWCP: “実環境音声・音響データベース (RWCP-SSD)” 音声資源コンソーシアム, http://research.nii.ac.jp/src/ RWCP-SSD.html, 閲覧日:2018/08/21.

[10] E. Vincent *et al.*, “First Stereo Audio Source Separation Evaluation Campaigh: Data Algprithm and Results,” Independent Component Analysis and Signal Separation(Springer, Bearlin, 2007), pp.552-559.