

マルチチャンネルNMFのための MUSICに基づくパーミュテーション解決のスケジューリング Scheduling of MUSIC-based Permutation Solution for Multi-channel NMF

太刀岡勇気
デンソーアイティラボラトリ

Yuuki TACHIOKA
Denso IT Laboratory

アブストラクト ブラインド音源分離では、周波数ごとに分離した音源を音源ごとにまとめるパーミュテーション解決の必要がある。マルチチャンネル非負値行列因子分解 (MNMF) による音源分離では、基底とアクティベーションにより、パーミュテーション解決が自然に行われることが期待されているが、実際にはパーミュテーション不整合が起こっている。まずはこれを実験的に示す。そしてそれを解決するために多重信号分類 (MUSIC) により、MNMF で求められる空間相関行列から分離音の到来方向を推定し、それに基づきパーミュテーション解決を行う方法を提案する。また、パーミュテーション解決を行うスケジュールを管理する方法を併せて提案する。劣決定条件での音源分離実験により、提案法の有効性を確認した。

1 はじめに

混合音を分離する音源分離において、独立成分分析のような周波数ごとの音源分離では、音源のパーミュテーションを解決する必要がある。音源の性質に注目したモデル化 (隣接周波数間での差異に注目したモデル化 [1] やスペクトル包絡のモデル化 [2], [3]) と到来方向に基づくクラスタリング [4] が提案されており、主に前者に関する検討が盛んである。例えば独立ベクトル分析 [5]–[7] では、異なる周波数間の成分が同期して励起されることを利用している。

マルチチャンネル NMF は、高精度の音源分離を達成できる方法である。音源基底とアクティベーションによるスペクトル包絡の拘束により、前者を内包していると考えられるが、後者の到来方向については明示的な拘束を使っていない。実際、推定された空間相関行列の類似度と音源分離の性能には明確な関係がある [8]。また本稿中の実験にて、音源分離性能が低い場合にはパーミュテーション解決がうまくいっていないことを示す。ここでは、マルチチャンネル NMF がその音源分離プロセスの過程で空間相関行列を推定していることに着目し、空間相関

行列の最大固有値に対応する固有ベクトルとしてステアリングベクトルが求まることから、それに対して多重信号分類 (Multiple signal classification; MUSIC) 法を適用する。従来、MUSIC を独立成分分析に対して適用した手法 [9] が提案されているが、この手法では分離後の信号に対して到来方向推定を行いパーミュテーション解決を行っていたのに対し、提案法では空間相関行列を直接扱うことから、音源分離の繰り返し最適化中に本処理を挿入することが可能である。これにより、パーミュテーション解決と音源分離を協調しておこなうことができる。

本報では、MUSIC スペクトルに基づいてパーミュテーションを解決する方法を提案する。さらにパーミュテーション解決を行うスケジュールの管理を行う方法を併せて提案する。提案法の有効性を確認するため、音楽データの分離実験を行った。

2 マルチチャンネル NMF

NMF では観測行列 \mathbf{X} を、基底行列 \mathbf{T} とアクティベーション行列 \mathbf{V} に因子分解する。さらに、MNMF は観測行列 \mathbf{X} を 4 行列に因子分解する [10]。2 行列 \mathbf{H} と \mathbf{Z} は、それぞれ空間相関行列とクラスター指示隠れ変数である。MNMF は K 個のスペクトル基底を L 個の音源に分解する際に、空間情報を使う。

2.1 定式化

周波数ビン i ($1 \leq i \leq I$)、時間フレーム j ($1 \leq j \leq J$) の観測ベクトル \mathbf{x}_{ij} を、 $[x_1, \dots, x_m, \dots, x_M]_{ij}^\top$ とする。ここで、 \top は転置、 x_m は m ($1 \leq m \leq M$) 番目のマイクで観測された短時間フーリエ変換 (STFT) の複素スペクトルである。これから、観測行列 $\mathbf{X} \in (\mathbb{C}^{M \times M})^{I \times J}$ の要素 i, j は以下のように表される。

$$\mathbf{X}_{ij} = \mathbf{x}_{ij} \mathbf{x}_{ij}^H = \begin{bmatrix} |x_1|^2 & \cdots & x_1 x_M^* \\ \vdots & \ddots & \vdots \\ x_M x_1^* & \cdots & |x_M|^2 \end{bmatrix}_{ij} \quad (1)$$

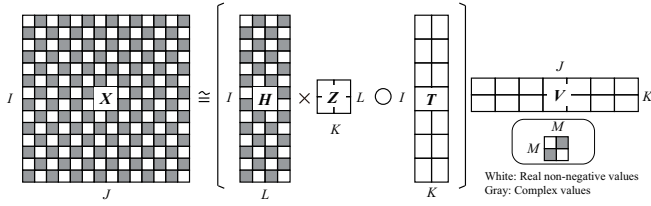


図 1: An example of factorizing an observation matrix \mathbf{X} into four matrices \mathbf{H} , \mathbf{Z} , \mathbf{T} , and \mathbf{V} by the multi-channel NMF algorithm. ($I = J = 7$ and $K = L = M = 2$)

ここで、 $*$ は複素共役、 \mathbf{H} はエルミート転置である。行列 \mathbf{X} は階層的な行列であり、その要素 X_{ij} は $M \times M$ の複素半正定値エルミート行列となる。この行列 \mathbf{X} は以下のように 4 行列 (\mathbf{H} , \mathbf{Z} , \mathbf{T} , \mathbf{V}) に分解される。

$$\mathbf{X} \cong \hat{\mathbf{X}} = [(\mathbf{H}\mathbf{Z}) \circ \mathbf{T}] \mathbf{V} \quad (2)$$

\circ はアダマール積で、図 1 は、式 (2) を図示したものである。 $\mathbf{H} \in (\mathbb{C}^{M \times M})^{I \times L}$ は、 L 個の音源の空間情報を示す空間相関行列である。行列 $\mathbf{Z} \in \mathbb{R}^{L \times K}$ は、空間情報を各基底に関連付けるクラスター指示隠れ変数行列である。基底行列 $\mathbf{T} \in \mathbb{R}^{I \times K}$ は K 個の基底から構成され、 $\mathbf{V} \in \mathbb{R}^{K \times J}$ は各基底のアクティベーションを表す。式 (2) の右辺は、以下のように表される。

$$\hat{X}_{ij} = \sum_k \left[\sum_l H_{il} z_{lk} \right] t_{ik} v_{kj} \quad (3)$$

理想条件では、 $\hat{\mathbf{X}}_{ij}$ を要素に持つ再構築された行列 $\hat{\mathbf{X}}$ は、元の観測行列 \mathbf{X} に一致するが、一般的には、これらの行列は誤差により一致しない。NMF において、 \mathbf{X} と $\hat{\mathbf{X}}$ の間の任意の距離を定義し、式 (2) の右辺の 4 行列は、この距離を最小化するように更新される。ここでは、板倉斎藤擬距離 (ISD) を使う。

3 MUSIC

MUSIC スペクトルを空間相関行列から以下のように求める。空間相関行列 $\mathbf{H}_{il} \in \mathbb{C}^{M \times M}$ を固有値分解

$$\mathbf{H}_{il} = \mathbf{V}_{il} \mathbf{D}_{il} \mathbf{V}_{il}^{-1} \quad (4)$$

する。 $\mathbf{D} \in \mathbb{R}^{M \times M}$ は (\mathbf{H} がエルミート行列のため) 実数の固有値を対角成分に持つ対角行列で、降順にソートされている。 $\mathbf{V} \in \mathbb{C}^{M \times M}$ は固有値に対応する固有ベクトルを列に並べたもので、互いに直交する。

到来方向を S 方向に離散化して、マイク間隔 d の直線アレイで平面波仮定すると、音源の各到来方向 $\theta_s (s \in \{1, \dots, S\})$ に対応するステアリングベクトル $\mathbf{a}(i, \theta_s) = [a_1, \dots, a_m, \dots, a_M]^T$ は

$$a_m = \exp \left[j \left(m - \frac{M+1}{2} \right) \frac{2\pi d}{c} \varphi(i) \sin \theta_s \right] \quad (5)$$

となる。ここで φ は周波数ビンを周波数 [Hz] に変換する関数、 c は音速である¹。

\mathbf{H}_{il} は 1 つの音源 l に対応する空間相関行列なので、最大固有値以外に対応する固有ベクトルは雑音部分空間を張る。これより、最大固有値で重みづけされた MUSIC スペクトルは

$$S_{il}(\theta_s) = \frac{\sqrt{D_{il}(1,1)}}{\mathbf{a}^H(i, \theta_s) \mathbf{V}(:, 2:M) \mathbf{V}^H(:, 2:M) \mathbf{a}(i, \theta_s)} \quad (6)$$

のように表される²。

音源のパワーが大きいところの MUSIC スペクトルのほうが小さいところよりも信頼性が高いと考えられるので、各音源に属するパワーで MUSIC スペクトルを除したパワー正規化された MUSIC スペクトル

$$S'_{i,l}(\theta_s) = \frac{S_{i,l}(\theta_s)}{\sum_j \sum_k z(l,k) t(i,k) v(k,j)} \quad (7)$$

を使うことも考えられる。

周波数全体での MUSIC スペクトルは

$$S_l(\theta_s) = \sum_{i=\varphi^{-1}(f_{min})}^{\varphi^{-1}(f_{max})} S_{il}(\theta_s) \quad (8)$$

で、MUSIC スペクトルの信頼性が高いと考えられる $f_{min} \sim f_{max}$ [Hz] までの周波数を用いる。

4 パーミュテーション評価

周波数全体での MUSIC スペクトルを音源ごとに比較することで、分離ができていないか評価に使う。3~7 の基準を使う際には MUSIC スペクトルを、その総和が 1 になるように正規化しておく。

$$\bar{S}_l(\theta_s) = \frac{S_l(\theta_s)}{\sum_s S_l(\theta_s)} \quad (9)$$

4.1 最大値 (max) 基準

音源 l に対するピーク

$$\phi_l = \arg \max_{\theta_s} S_l(\theta_s) \quad (10)$$

を算出し、ピークをとるインデックスの距離

$$d_{max}(\phi_l, \phi_{l'}) = |\phi_l - \phi_{l'}| \quad (11)$$

で分離度を評価する。これらを音源ペアごとに足し合わせることで最終的な分離度

$$D_{max} = \sum_l \sum_{l' > l} d_{max}(\phi_l, \phi_{l'}) \quad (12)$$

を推定する。

¹ただし等間隔であれば、マイク間隔 d は不明でもよい。 d を仮定して実際には d' であった場合でも、 $a'_m(\theta_s) = a_m(\theta_s)^{d'/d}$ となるだけなので、MUSIC スペクトルの概形は変わらない。

² \mathbf{V} に関するインデックス il は煩雑になるので省略した。

4.2 内積 (IP) 基準

スペクトルの類似度を内積を正規化したコサイン類似度 d_{IP} で測る.

$$d_{IP}(\mathbf{S}_l, \mathbf{S}_{l'}) = -\frac{\mathbf{S}_l^\top \mathbf{S}_{l'}}{|\mathbf{S}_l| |\mathbf{S}_{l'}|} \quad (13)$$

ここでは分離度を判定したいのでコサイン類似度の符号を変えている. 上と同様に, 音源ペアごとに和を取る.

$$D_{IP} = \sum_l \sum_{l' > l} d_{IP}(\mathbf{S}_l, \mathbf{S}_{l'}) \quad (14)$$

$\mathbf{S} = [S(\theta_1), \dots, S(\theta_S)]^\top$ である. $\mathbf{S} > 0$ なので, 内積は必ず 0 以上となる.

4.3 二乗誤差 (SE) 基準

スペクトル間の距離を 2 乗誤差

$$d_{SE}(p_1, p_2) = \sum_\theta |p_1(\theta) - p_2(\theta)|^2 \quad (15)$$

で測り,

$$D_{SE} = \sum_l \sum_{l' > l} d_{SE}(\bar{S}_l, \bar{S}_{l'})$$

で評価する.

4.4 重なり面積 (Overlapping area; OA) 基準

正規化されたスペクトル間の重なり面積

$$d_{CA}(p_1, p_2) = \sum_\theta -\min(p_1(\theta), p_2(\theta)) \quad (16)$$

で測り,

$$D_{CA} = \sum_l \sum_{l' > l} d_{CA}(\bar{S}_l, \bar{S}_{l'})$$

で評価する.

4.5 Kullback-Leibler divergence (KLD) 基準

スペクトル間の距離を KLD

$$d_{KL}(p_1, p_2) = \sum_\theta p_1(\theta) \log \left(\frac{p_1(\theta)}{p_2(\theta)} \right) \quad (17)$$

で測り,

$$D_{KL} = \frac{1}{2} \sum_l \sum_{l' \neq l} d_{KL}(\bar{S}_l, \bar{S}_{l'})$$

で評価する.

4.6 Itakura-Saito divergence (ISD) 基準

スペクトル同士の距離を測るのによいとされている ISD

$$d_{ISD}(p_1, p_2) = \sum_\theta \left[\frac{p_1(\theta)}{p_2(\theta)} - \log \left(\frac{p_1(\theta)}{p_2(\theta)} \right) - 1 \right] \quad (18)$$

で測り,

$$D_{ISD} = \frac{1}{2} \sum_l \sum_{l' \neq l} d_{ISD}(\bar{S}_l, \bar{S}_{l'})$$

で評価する.

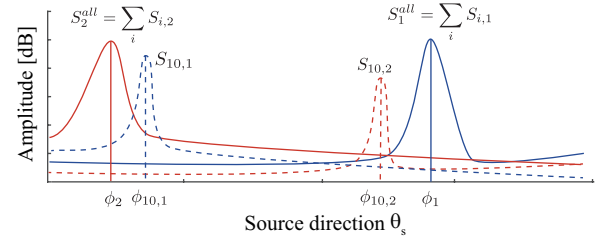


図 2: Finding permutation at each frequency bin.

4.7 Density power divergence (DPD) 基準

効率性とロバスト性を調整できる位相回復の分野で使われている DPD[11]

$$d_{DPD}(p_1, p_2) = \sum_\theta \left[\frac{1}{\gamma} (p_1(\theta)^\gamma - p_2(\theta)^\gamma) - \frac{1}{1+\gamma} (p_1(\theta)^{1+\gamma} - p_2(\theta)^{1+\gamma}) \right] \quad (19)$$

で測り,

$$D_{DPD} = \frac{1}{2} \sum_l \sum_{l' \neq l} d_{DPD}(\bar{S}_l, \bar{S}_{l'})$$

で評価する. ただし $\gamma = 0.2$ とした.

5 パーミュテーション解決

周波数全体での MUSIC スペクトルと各周波数ビンでの MUSIC スペクトルを以下の基準で比較することで, パーミュテーション解決を行う. 以下 $\mathcal{P} = \{p_1, p_2, \dots, p_L\}$ を音源 $1, \dots, L$ のパーミュテーションとする. 例えば $L = 2$ の場合, $\mathcal{P} = \{p_1, p_2\} = \{(1, 2), (2, 1)\}$ である.

図 2 のように, 全体のスペクトル $\sum_i S_{i,l}$ と個別の周波数ビン (ここでは $f = 10$) のスペクトル $S_{10,l}$ を比較する. $S_{10,1}$ のピーク位置は, $\sum_i S_{i,1}$ のピーク位置よりも, $\sum_i S_{i,2}$ のピーク位置に近く, ここにおいてパーミュテーションが起きていると判断できる.

5.1 最大値 (max) 基準

全体でのピーク ϕ_l と各周波数ビンでのピーク

$$\phi_{il} = \arg \max_{\theta_s} S_{il}(\theta_s) \quad (20)$$

を算出し, ピーク同士がなるべく近くなるようにパーミュテーション解決を行う.

$$\min_m \sum_l d_{max}(\phi_l, \phi_{ip_m(l)}) = \min_m \sum_l |\phi_l - \phi_{ip_m(l)}| \quad (21)$$

5.2 内積 (IP) 基準

内積も同様だが, こちらは最大となるようにする.

$$\max_m \sum_l d_{IP}(\mathbf{S}_l, \mathbf{S}_{ip_m(l)}) = \max_m \sum_l \frac{\mathbf{S}_l^\top \mathbf{S}_{ip_m(l)}}{|\mathbf{S}_l| |\mathbf{S}_{ip_m(l)}|} \quad (22)$$

である.

表 1: Setup for music source separation.

Sampling frequency	16 kHz
Frame size and shift	1024 and 256
# base s	30
# iterations	500

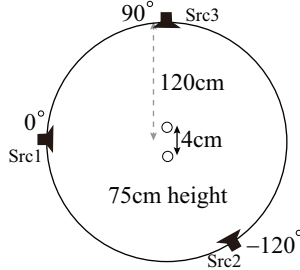


図 3: Source locations.

5.3 その他の基準

その他の基準に関しても, $\min_m \sum_l d. (\bar{S}_l, \bar{S}_{ip_m(l)})$ とすることで同様にパーミュテーション解決できる.

6 パーミュテーション解決のスケジュール管理

音源分離アルゴリズムの繰り返しの当初では, 音源位置に関する事前知識がないので, パーミュテーション解決を行うことは不可能である. 何回か音源分離アルゴリズムを動作させたのちに (すなわち音源分離アルゴリズムの繰り返し途中), パーミュテーション解決を始める. ここでは 2 種類のパーミュテーション解決のスケジュール管理法を提案する.

6.1 固定繰り返し数

1 つめは, 固定繰り返し数でパーミュテーション解決を始めるものである. 予備実験の結果により, 数回ある程度の間隔をあけてパーミュテーション解決を行った方が, 一度だけ行う場合よりも結果がよいことが分かっている.

6.2 固定閾値

2 つめのタイプは, 指標 D がある固定閾値を超えたらパーミュテーション解決を行う固定閾値型である. パーミュテーション解決を行うためには, はっきりとした MUSIC スペクトルが必要である. 始めは, すべての音源に対して, MUSIC スペクトルが重なっており, すべての指標 d は最小値を取る. 図 5 で実験的に示すように, 繰り返しに従い, すべての指標がほぼ単調に増加する.

7 実験

7.1 実験条件

3 音源を 2 マイクで観測して分離する劣決定条件音源分離実験を行った³. 図 3 に音源の配置を示す. 表 2 に示

³文献 [12] と同条件である.

表 2: Source music

ID	Author/Song	Snip	Part
1	Bearlin Roads	85-99 (14 sec)	piano ambient vocals
2	Another Dreamer The Ones We Love	69-94 (25 sec)	drums vocals guitar
3	Fort Minor Remember The Name	54-78 (24 sec)	drums vocals violin+synth
4	— Ultimate Nz Tour	43-61 (18 sec)	drums guitar synth

表 3: Schedule of permutation alignment for the fixed count type.

sID	Schedule	sID	Schedule
1	40,45,50,55,60	3	100,150
2	50,100,150	4	50,100,150,200

す 3 音源 ($L = 3$) の音楽データに, それぞれ正面 (0 度), -90 度, 60 度方向からのインパルス応答 (インパルス応答長 300) を畳み込み, 4cm 間隔の 2 マイク ($M = 2$) で分離した結果を信号対歪比 (SDR)[dB] で評価した.

10 の異なる乱数シードを設定し, \mathbf{T} , \mathbf{V} , \mathbf{Z} には乱数を \mathbf{H} は単位行列として実験した. 従来研究に倣い, 最初の 20 回は \mathbf{H} を固定で乗法更新則を適用した.

表 3 に示す更新回数においてパーミュテーション解決を行った. 音源分離のパラメータは, 表 1 の通り設定した. $f_{min} = 500$, $f_{max} = \frac{c}{2d} = 4250[\text{Hz}]$ とした.

7.2 MUSIC スペクトルの考察

10 の分離結果のうち, SDR 最良 (14.63[dB]) の場合 (best) と最悪 (1.63[dB]) の場合 (worst) で, 最終的に得られた \mathbf{H} を対象として, 提案法により求めた MUSIC スペクトルを, 図 4 に示す. 明らかに worst の場合は, best の場合に比べてピークが重なっている. これから, MNMF においても, パーミュテーション解決がうまくいかない場合があることがうかがえる. 例えば 4.1 の基準に従えば, 左はピーク位置が音源 1,2,3 について $\{0.4, -1, -1.25\}$ なので, それぞれの絶対値の総和は $|0.4 - (-1)| + |-1 - (-1.25)| + |0.4 - (-1.25)| = 3.3$ である. これに対して, 右はそれぞれ $\{0.25, 0.25, -1\}$ なので, $|0.25 - 0.25| + |0.25 - (-1)| + |0.25 - (-1)| = 2.5$ で左の方が指標が大きい.

7.3 分離性能の予測

表 4 に SDR と 4 節での指標の相関係数を示す. 曲 1 を除くと相関はあまり高くなく, SDR の直接的な予測はパーミュテーションの程度だけからは難しいことが分かった.

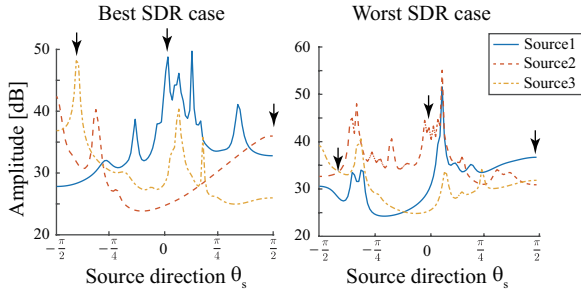


図 4: MUSIC spectrum of \mathbf{H} corresponding to the best and worst SDR among results starting from 10 random initializations.

表 4: Correlation coefficients between SDR[dB] and criteria.

ID	d_{IS}	max	IP	DPD	KL	IS	SE	OA
1	-0.56	0.89	0.89	0.80	0.77	-0.29	0.64	0.91
2	-0.15	-0.14	-0.41	-0.58	-0.57	-0.35	-0.79	-0.47
3	-0.83	0.36	0.56	0.46	0.50	-0.12	0.01	0.51
4	-0.62	0.66	0.55	0.49	0.45	-0.59	0.26	0.51

7.4 指標 D の変化

図 5 には、2 楽曲に対する繰り返し数と規格化された D を示す。2 楽曲に対して収束速度が異なるものの、100 繰り返しで指標はほぼ収束した。6 節で述べたように、すべての指標はほぼ単調に増加する。ただし、はじめの 20 繰り返しは \mathbf{H} を更新しないため、20 繰り返しまで、すべての指標は変化がない。

7.5 分離実験の結果 (固定繰り返し数)

分離実験の結果を表 5 に示す。表中「proposed (Eq. (6))」は通常の MUSIC スペクトルを使った場合、「proposed (Eq. (7))」はパワーで正規化した MUSIC スペクトルを使った場合である。このように楽曲により差異があるが、平均的にみるとすべての場合で性能の改善がみられる。最良の基準は楽曲によって異なるが、max か IP が最も安定していることがわかる。

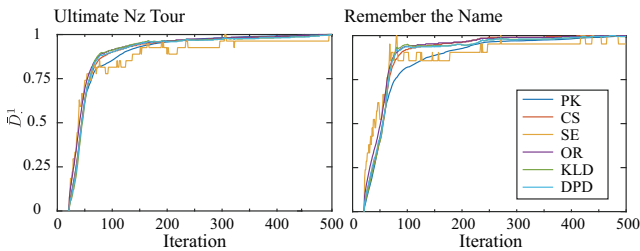


図 5: Normalized total discrepancy \bar{D} and the number of iterations where \bar{D} is $(D - \min(D))/(\max(D) - \min(D))$.

表 5: Average SDR[dB] of each song with fixed iteration numbers.

Criterion	sID	Music ID				Avg.
		1	2	3	4	
baseline	-	2.18	5.55	3.07	7.52	4.58
Based on Eq. (6)						
max	1	2.64	7.24	2.59	9.08	5.39
max	2	2.35	6.77	3.08	7.00	4.80
max	3	2.18	4.88	2.88	6.92	4.22
max	4	2.40	5.67	3.08	5.94	4.27
IP	1	1.60	6.63	2.42	8.17	4.70
IP	2	2.51	7.57	2.74	8.75	5.39
IP	4	2.78	7.63	2.75	8.48	5.41
DPD	1	2.09	6.89	2.95	8.79	5.18
DPD	2	1.28	7.25	2.92	8.48	4.98
DPD	4	1.38	7.21	2.89	7.78	4.81
KLD	1	1.81	5.76	2.61	8.72	4.72
KLD	2	1.15	6.69	2.99	8.31	4.78
KLD	4	1.18	6.63	2.94	8.40	4.79
ISD	1	1.82	5.85	2.69	8.62	4.75
ISD	2	1.11	6.62	2.82	8.26	4.70
ISD	4	1.30	6.47	2.85	8.46	4.77
SE	1	2.03	6.06	2.67	8.52	4.82
SE	2	1.24	6.58	2.81	8.35	4.75
SE	4	1.22	6.60	2.95	8.08	4.71
OA	1	1.34	6.72	2.65	8.46	4.79
OA	2	2.01	7.29	2.81	8.75	5.22
OA	4	1.85	7.23	2.82	8.86	5.19
Based on Eq. (7)						
max	1	2.85	6.31	2.53	8.13	4.95
max	2	3.51	7.38	3.57	7.88	5.58
IP	1	1.93	8.11	2.45	8.81	5.32
IP	2	2.18	5.89	2.98	9.12	5.04
DPD	1	1.47	5.44	2.47	9.37	4.69
DPD	2	1.47	6.06	2.92	9.48	4.98
KLD	1	2.00	5.57	2.35	9.42	4.84
KLD	2	1.26	6.05	2.80	9.45	4.89
ISD	1	1.36	5.41	2.35	8.59	4.43
ISD	2	1.41	5.87	2.74	9.78	4.95
SE	1	1.03	6.63	2.79	9.14	4.90
SE	2	1.84	6.36	2.81	9.21	5.05
OA	1	2.43	7.21	2.32	8.51	5.12
OA	2	1.57	5.78	3.30	9.09	4.93

7.6 分離実験の結果 (固定閾値)

固定閾値の場合の平均 SDR を表 6 に示す。固定繰り返し数の場合に比べて、固定閾値の方が性能が良い。これは、楽曲ごとに異なる収束速度に合わせて、パーミュテーション解決を適切なタイミング行うことができるためと考えられる。

8 まとめ

音源分離の際のパーミュテーション解決を目的として、空間相関行列の MUSIC スペクトルを用いたパーミュテーション解決法を提案した。音楽データの分離実験により、

表 6: Average SDR[dB] of each song with fixed threshold.

Criterion	Thres.	Music ID				Avg.
		1	2	3	4	
baseline	-	2.18	5.55	3.07	7.52	4.58
Based on Eq. (6)						
max	40,5/5	3.99	7.92	2.75	8.45	5.78
max	40,4/25	3.80	6.70	2.83	7.57	5.23
max	40,3/50	3.67	6.36	3.06	7.21	5.07
max	50,5/5	4.13	7.40	3.28	8.54	5.84
max	50,4/25	2.92	6.61	3.65	6.70	4.97
max	50,3/50	3.94	5.45	3.40	6.77	4.89
IP	-2.5	3.32	7.22	2.52	8.70	5.44
IP	-2.25	2.80	7.13	3.14	9.45	5.63
IP	-2.0	2.60	6.03	3.06	8.22	4.98
IP	-1.75	2.92	6.08	3.06	8.22	5.07
DPD	0.6	2.55	7.79	2.88	9.03	5.56
DPD	0.8	2.54	6.15	3.22	8.54	5.11
KLD	1.6	2.10	6.77	3.12	9.49	5.37
KLD	2.0	3.02	5.82	2.77	9.60	5.30
SE	0.1	2.64	7.31	2.76	8.52	5.31
SE	0.125	3.07	7.45	2.93	8.44	5.47
OA	-2.5	2.12	7.31	2.38	8.66	5.12
OA	-2.25	2.92	6.43	2.51	8.97	5.21
Based on Eq. (7)						
max	40	4.11	5.42	3.13	8.18	5.21
max	50	3.80	6.36	3.19	8.57	5.48
IP	-2.5	2.23	7.77	2.74	8.89	5.41
IP	-2.25	3.45	5.44	3.04	8.09	5.00
DPD	0.6	2.00	6.67	2.51	8.34	4.88
DPD	0.8	2.75	5.84	2.89	9.27	5.19
KLD	1.6	2.65	5.53	2.81	9.81	5.20
KLD	2.0	2.46	5.40	2.85	9.48	5.05
SE	0.1	1.96	7.06	2.91	8.76	5.17
SE	0.125	2.83	7.67	3.11	8.81	5.61
OA	-2.5	1.70	7.16	2.56	7.79	4.80
OA	-2.25	2.30	7.26	3.17	8.21	5.23

分離結果が悪い場合にはパーミュテーション解決が十分でないことを明らかにし、これを用いてパーミュテーション解決を音源分離の繰り返し最適化中に行うことで音源分離性能が向上することが示された。また更新回数のスケジュール管理を行う方法を2種類提案し、その有効性を確認した。

参考文献

[1] R. Mazur and A. Mertins, “An approach for solving the permutation problem of convolutive blind source separation based on statistical signal models,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol.17, no.1, pp.117–126, Jan. 2009.

[2] S. Saito, K. Oishi, and T. Furukawa, “Convolutive blind source separation using an iterative least-squares algorithm for non-orthogonal approximate joint diagonalization,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol.23, no.12, pp.2434–2448, Dec. 2015.

[3] A. Sarmiento, I. Duran-Diaz, A. Cichocki, and S. Cruces, “A contrast function based on generalised divergences for solving the permutation problem in convolved speech mixtures,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol.23, no.11, pp.1713–1726, Nov. 2015.

[4] H. Sawada, S. Araki, and S. Makino, “Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol.19, pp.516–527, 2011.

[5] A. Hiroe, “Solution of permutation problem in frequency domain ICA using multivariate probability density functions,” *Proceedings of International Workshop on Independent Component Analysis and Source Separation (ICA)*, pp.601–608, 2006.

[6] T. Kim, T. Eltoft, and T. Lee, “Independent vector analysis: An extension of ICA to multivariate components,” *Proceedings of International Workshop on Independent Component Analysis and Source Separation (ICA)*, pp.165–172, 2006.

[7] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp.189–192, Oct. 2011.

[8] 浦本昂伸, 太刀岡勇気, 成田知宏, 三浦伊織, 上ノ原進吾, 古家賢一, “マルチチャンネル非負値行列因子分解を用いたブラインド音源分離のためのチャンネル数増加に伴う逐次的初期化法,” *電子情報通信学会論文誌 D*, vol.J101-D, no.3, pp.569–577, March 2018.

[9] N. Mitianoudis and M. Davies, “Permutation alignment for frequency domain ICA using subspace beamforming methods,” *Proceedings of International Workshop on Independent Component Analysis and Source Separation (ICA)*, pp.127–132, 2004.

[10] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, “Multichannel extensions of non-negative matrix factorization with complex-valued data,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol.21, no.5, pp.971–982, May 2013.

[11] 塩谷浩之, 郷原一寿, “位相回復—計算アルゴリズム—,” *計測と制御*, vol.50, no.5, pp.332–337, May 2011.

[12] 三浦伊織, 太刀岡勇気, 成田知宏, 石井純, 吉山文教, 上ノ原進吾, 古家賢一, “マルチチャンネル NMF を用いた音源分離における初期値依存性の挙動解析と音声認識での評価,” *電子情報通信学会論文誌 D*, vol.J100-D, no.3, pp.376–384, March 2017.