局所時間周波数構造に基づく深層パーミュテーション解決法の実験的評価*

☆山地修平,北村大地 (香川高専)

1 はじめに

ブラインド音源分離(blind source separation: BSS)とは、複数の音源が混合した観測信号から、混 合前の音源を推定する技術である.特に独立成分分 析 (independent component analysis: ICA) [1] の 登場以降,音源信号数>マイクロホン数である優決 定条件下の音源分離問題に広く適用されている.

一般的には、周波数領域での瞬時混合を仮定し ICA を周波数毎に適用する周波数領域 ICA (frequencydomain ICA: FDICA) [2] に基づく手法がよく用い られる.しかし、ICA は分離信号の順番が不定であ り、FDICA は周波数毎に分離を行うため、周波数毎 の分離信号の順番がバラバラになってしまう. これを 正しい順番に並び替える問題は一般にパーミュテー ション問題 [3] と呼ばれている.近年では FDICA に 対して音源の時間周波数成分の共起関係を新たに仮定 して,パーミュテーション問題を起こさず分離信号を 推定する BSS が登場している. 例えば, 独立ベクト ル分析 (independent vector analysis: IVA) [4] や独 立低ランク行列分析 (independent low-rank matrix analysis: ILRMA) [5, 6] が有名である. さらに, 深 層学習(deep neural networks: DNN)を用いて音源 の時間周波数構造の仮定を学習し、FDICA に適用す る独立深層学習行列分析 [7] も提案されている. しか しながら、上記いずれの手法を用いても完璧にパー ミュテーション問題を解くことは難しく、とくに複数 音声の混合信号では IVA や ILRMA 等でも分離に失 敗してしまうことが多い.

文献 [8] では、局所的な時間周波数構造から分離信 号の正しいパーミュテーションを予測する DNN を 新たに提案し, FDICA が周波数毎に理想的な分離を 行った状況では高精度な BSS が達成できることを示 している.本稿では、上記の DNN に基づくパーミュ テーション解決法を拡張し, FDICA の分離結果に対 してより頑健な予測を達成できるアルゴリズムを新 たに提案する.また,提案手法を実際に FDICA のポ スト処理として用いて、従来の BSS と音源分離性能 を比較する.提案するパーミュテーション解決法は, モノラル音源の推定値のみを必要とするため、劣決定 条件の場合にも一般性を損なうことなく適用できる.

FDICA とパーミュテーション問題 2

2.1 定式化

音源数と観測チャネル数(マイクロホン数)をそれ ぞれ N 及び M とし、各時間周波数における音声信 号,混合信号,及び分離信号をそれぞれ

$$\boldsymbol{s}_{ij} = (s_{ij,1}, \cdots, s_{ij,n}, \cdots, s_{ij,N})^{\mathrm{T}} \in \mathbb{C}^{N}$$
(1)

$$\boldsymbol{x}_{ij} = (x_{ij,1}, \cdots, x_{ij,m}, \cdots, x_{ij,M})^{\mathrm{T}} \in \mathbb{C}^{M} \quad (2)$$
$$\boldsymbol{z}_{ij} = (z_{ij,1}, \cdots, z_{ij,n}, \cdots, z_{ij,N})^{\mathrm{T}} \in \mathbb{C}^{N} \quad (3)$$

と表す. ここで, $i = 1, 2, \cdots, I$, $j = 1, 2, \cdots, J$, $n = 1, 2, \cdots, N$,及び $m = 1, 2, \cdots, M$ はそれぞれ

周波数,時間,音源,チャネルのインデクスを示し,一 は転置を表す.また,各信号の複素スペクトログラム 行列を $S_n \in \mathbb{C}^{I \times J}$, $X_m \in \mathbb{C}^{I \times J}$,及び $Z_n \in \mathbb{C}^{I \times J}$ で表す.周波数毎の時不変な混合行列 $oldsymbol{A}_i \in \mathbb{C}^{M imes N}$ を定義し、混合信号が次式で表現できると仮定する.

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij} \tag{4}$$

以後,決定的な系 (M = N) を仮定する. A_i が正則 であれば,分離行列 $oldsymbol{W}_i = oldsymbol{A}_i^{-1} \in \mathbb{C}^{N imes M}$ を用いて, 分離信号を次式で表せる.

$$\boldsymbol{z}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij} \tag{5}$$

従って, 音源分離は分離行列 W: の推定問題となる.

FDICA は音源間の統計的独立性のみに基づいて分 離行列を推定するため,分離音源のスケール及び順 番に関しては任意である.従って, FDICA の推定分 離行列を Ŵ_iとすると,次式のような不定性が残る.

$$\hat{\boldsymbol{W}}_i = \boldsymbol{D}_i \boldsymbol{P}_i \boldsymbol{W}_i \tag{6}$$

ここで, P_i は任意のパーミュテーション行列, D_i は 任意の対角行列である.また,式(8)の音源毎の複素 スペクトログラム行列を $Y_n \in \mathbb{C}^{I imes j}$ と定義する.す なわち, FDICA で推定される分離信号

$$\boldsymbol{y}_{ij} = \hat{\boldsymbol{W}}_i \boldsymbol{x}_{ij} \tag{7}$$

$$= (y_{ij,1}, \cdots, y_{ij,n}, \cdots, y_{ij,N})^{\mathrm{T}} \in \mathbb{C}^{N}$$
 (8)

は,推定音源の順番やスケールが周波数毎にばらば らになっている状態である.このうち, D_i によって 生じるスケールの任意性は、プロジェクションバック 法 [9] で復元可能である.一方で, **P**_i によって生じ る分離信号の順番の任意性(パーミュテーション)を 復元することは容易ではない.

パーミュテーション問題を解決して得られる分離信 号は次式となる.

$$\boldsymbol{z}_{ij} = \boldsymbol{P}_i^{-1} \boldsymbol{D}_i^{-1} \boldsymbol{y}_{ij} \tag{9}$$

本稿では、 P_i^{-1} を推定することが課題となる.

提案手法 3

3.1動機

文献 [10] では, BSS の短時間フーリエ変換 (shorttime Fourier transform: STFT) おける最適な窓長 を実験的に検討している. 文献 [10] 中の Fig. 6 (b) よ り, IVA 及び ILRMA では, 残響状態 T₆₀ = 470 ms の条件では音声を分離できないことがわかる.一方 で,FDICA に対して,音源信号 $s_{i,j}$ を用いる理想 的なパーミュテーション解決法 (ideal permutation solver: IPS) を適用した結果では 10 dB 以上の信号 対歪み比(signal-to-distortion ratio:SDR)[11]の改

(2)

^{*}Experimental evaluation of deep permutation solver based on local time-frequency structure. By Shuhei YAMAJI and Daichi KITAMURA (NIT Kagawa).

善を達成している.この事実は、高残響下での音声 混合信号であっても、 \hat{W}_i は FDICA で正確に推定で き, **P**⁻¹の推定のみ失敗していることを示している. そこで, 文献 [8] において, 著者らは DNN に基づく 教師付きパーミュテーション解決法を提案し, FDICA が周波数毎に理想的な分離を行った状況では高精度 なBSS が達成できることを示した.本稿では、音源 数 N = 2 及びチャネル数 M = 2 を仮定したうえで 上記のパーミュテーション解決法を拡張し, FDICA の分離結果に対してより頑健な予測を達成できるア ルゴリズムを新たに提案する.この手法の概要は以 下の通りである.

- 分離信号 Y₁ 及び Y₂ から同じ 2 つの周波数の時 系列パワーをそれぞれ抽出し DNN に入力する
- DNN は入力された2つの時系列パワーが同一音 源か否かを予測し0または1として出力する
- Y₁ 及び Y₂ の全時間及び全周波数に対して DNN が適用される
- 最終的な推定値 **P**⁻¹ は,予測値の時間方向及び 周波数方向の多数決結果から決定される

3.2 DNN の入出力

提案する DNN モデルの入力ベクトルを Fig. 1 に 示す. 観測された混合信号 X_n に FDICA を適用する と、パーミュテーション問題が生じた分離信号 Y_n が 得られる. $|Y_n|^2$ から, 2つの周波数 $(i,i+\omega)$ の短 時間時系列パワー(長さ τ)を以下のように集める.

$$\boldsymbol{d}_{i,\omega,\gamma} = (\tilde{\boldsymbol{r}}_{i,\gamma}^{\mathrm{T}}, \tilde{\boldsymbol{g}}_{i,\omega,\gamma}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}_{>0}^{4\tau \times 1}$$
(10)

$$\tilde{\boldsymbol{r}}_{i,\gamma} = (\boldsymbol{r}_{i,\gamma,1}^{\mathrm{T}}, \boldsymbol{r}_{i,\gamma,2}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2\tau \times 1}_{>0}$$
(11)

$$m{r}_{i,\gamma,n} = (|y_{i,(\gamma-1)\eta+1,n}|^2, |y_{i,(\gamma-1)\eta+2,n}|^2,$$

$$(y_{i,(\gamma-1)\eta+\tau,n}|^2)^{\mathrm{T}} \in \mathbb{R}^{\tau \times 1}_{>0}$$
 (12)

$$\tilde{\boldsymbol{g}}_{i,\omega,\gamma} = (\boldsymbol{g}_{i,\omega,\gamma,1}^{\mathrm{T}}, \boldsymbol{g}_{i,\omega,\gamma,2}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2\tau \times 1}_{\geq 0} \qquad (13)$$

$$\boldsymbol{g}_{i,\omega,\gamma,n} = (|y_{i+\omega,(\gamma-1)\eta+1,n}|^2, |y_{i+\omega,(\gamma-1)\eta+2,n}|^2, \\ \cdots, |y_{i+\omega,(\gamma-1)\eta+\tau,n}|^2)^{\mathrm{T}} \in \mathbb{R}_{>0}^{\tau \times 1} \quad (14)$$

ここで, 行列の |-|-² は, 要素ごとの二乗絶対値を返す. $\sharp \hbar, \ \omega = -\Omega, -\Omega + 1, \cdots, -1, 0, 1, \cdots, \Omega \ k, \ r_{i,\gamma,n}$ と $g_{i,\omega,\gamma,n}$ の周波数の差であり、 η は、短時間セグメ ントの時間軸に沿ったストライド幅, $\gamma = 1, 2, \cdots, \Gamma$ は,短時間セグメントのインデクスである.なお, Γ は、短時間のアクティベーションの長さ τ とストラ イド幅 η によって決まる.ベクトル $r_{i,\gamma,n}$ は、参照 周波数 i の短時間時系列パワーに対応し、ベクトル $g_{i,\omega,\gamma,n}$ は、Fig. 1 に示すように、隣接又は局所周波 数 $i + \omega$ の短時間時系列パワーに対応する. DNNの 入力ベクトルは, (10) を正規化したものである.

$$\tilde{d}_{i,\omega,\gamma} = \frac{d_{i,\omega,\gamma}}{\|d_{i,\omega,\gamma}\|_2} \in \mathbb{R}^{4\tau \times 1}_{\geq 0}$$
(15)

ここで, ||·||₂ は, *L*₂ ノルムを表す. 提案 DNN モデルは, 0 または 1 を出力する 2 値 分類器である.推定結果が「0」の場合は、 $r_{i,\gamma,1}$ と $oldsymbol{g}_{i,\omega,\gamma,1}$ が同一音源であることを意味し,同様に $oldsymbol{r}_{i,\gamma,2}$ と $g_{i,\omega,\gamma,2}$ も同一音源である,一方,推定結果が[1]の場合は $r_{i,\gamma,1}$ と $g_{i,\omega,\gamma,1}$ (同様に $r_{i,\gamma,2}$ と $g_{i,\omega,\gamma,2}$)が 異なる音源成分であることを意味している. これら の推定処理を Fig. 2 に示す. 実際には, DNN の予測 結果は2値ではなく次式のような値である.

$$q_{i,\omega,\gamma} = \text{DNN}\left(\tilde{d}_{i,\omega,\gamma}\right) \in [0,1]$$
 (16)



Fig. 2 DNN predictions in subband frequency bins.

3.3 DNN の構造

提案する DNN の構造は、入力層、隠れ層 6 層、及 び出力層の計8層からなる全結合構成となっており, 1~5番目の隠れ層には ReLU 関数,最終隠れ層には sigmoid 関数を適用している.16番目の隠れ層の次 元数はそれぞれ 128, 128, 128, 64, 64, 1 である.予測 結果 $q_{i,\omega,\gamma}$ と正解ラベルとの誤差関数には,平均二乗 誤差(mean squared error: MSE)を使用している.

3.4 サブバンド周波数帯域における DNN 推定

DNN の予測例を Fig. 2 に示す. ここで, f_1, f_2, \cdots, f_5 はサブバンド内の周波数であり, 簡単のために γ は省略している. この図では, 参照周波数を $i = f_3$ とし,その近傍周波数を $i + \omega = f_1, f_2, \cdots, f_5$ と定義している. Y_1 の参照周 波数 f_3 の成分は赤色の音源成分であり、 Y_1 の f_2 、 f₄,及び f₅ は青色の音源成分である.この内,2本 の短時間時系列パワー $(\tilde{r}_i, \tilde{g}_{i,\omega})$ の全組み合わせが DNN に入力される.入力された短時間時系列パワー $(i \ge i + \omega)$ のパーミュテーションが正しい場合, DNN は理想的には「0」を出力する. 逆に, パーミュ テーションが正しくない場合は, DNN は理想的には 「1」を出力する. その結果 Fig. 2 の右側に示すよう に、参照周波数 *i* に基づくパーミュテーション問題 の発生個所の推定が可能となる.以後このベクトル は、サブバンドベクトルと呼ぶ.実際には、DNNの 出力は [0,1] の範囲内の値となるので、サブバンドベ クトルの生成時は以下の閾値処理を行う.

$$\tilde{q}_{i,\omega,\gamma} = \operatorname{round}(q_{i,\omega,\gamma}) \in \{0,1\}$$
 (17)

$$\tilde{\boldsymbol{q}}_{i,\gamma} = (\tilde{q}_{i,-\Omega,\gamma}, \tilde{q}_{i,-\Omega+1,\gamma}, \cdots, \tilde{q}_{i,-1,\gamma}, \tilde{q}_{i,0,\gamma}, \cdots, \tilde{q}_{i,-1,\gamma}, \tilde{q}_{i,0,\gamma})$$

$$q_{i,1,\gamma},\cdots,q_{i,\Omega,\gamma}) \in \{0,1\}^{2n+1}$$
(18)

ここで, round(·) は, 丸め演算子である.

音声信号は本来, 無音区間が多く存在することか ら,長さauの短時間時系列パワー $\tilde{r}_{i,\gamma}$ や $\tilde{g}_{i,\omega,\gamma}$ はほ ぼ零ベクトルになる可能性があり,その場合 DNN の 予測は不安定になる.この問題に対処するために,提 案手法では, Fig. 3 に示すように, 長さτの入力ベ クトルをストライド幅 n でシフトさせて, 全時間フ



Fig. 3 DNN predictions for all short-time subbands and their majority decision.



Fig. 4 Estimation of subband permutation vectors in all frequency bins.

レームに DNN 予測を走査する.そして,DNN の予 測結果を時間軸に関して多数決することで,より信 頼性の高いサブバンドベクトル \tilde{v}_i を得る.この処理 は,次のように示される.

$$\boldsymbol{v}_i = \frac{1}{\Gamma} \sum_{\gamma} \tilde{\boldsymbol{q}}_{i,\gamma} \in \{0,1\}^{2\Omega+1}$$
(19)

$$\tilde{\boldsymbol{v}}_i = \operatorname{round}(\boldsymbol{v}_i) \in \{0, 1\}^{2\Omega + 1}$$
(20)

実際に,音源の順番は時間に依存しないため,多数決 によって予測誤差の悪影響を大幅に軽減できる.

3.5 フルバンドベクトルの推定

提案する DNN パーミュテーション解決法は (a) サ ブバンドのストライドによる全周波数のサブバンド ベクトルの推定 (3.5.1 項及び Fig. 4) 及び (b) 類似 度比較と多数決に基づくフルバンドベクトルの構築 (3.5.2 項及び Fig. 5) で構成される.

3.5.1 全周波数におけるサブバンドベクトルの推定 サブバンドベクトル \tilde{v}_i は、基準周波数 i をシフ トすることにより全周波数で推定する. この処理を Fig. 4 に示す. 全部で I 個のサブバンドベクトル $\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_I$ を推定している. ここで、各サブバン ドベクトル $\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_I$ 内の2値(「0」及び「1」) は、同じ意味を持たない. これはサブバンド内の周波 数成分が、参照周波数 i の成分と同一音源か否かを示 しているに過ぎず、参照周波数 i の変化(サブバンド のシフト)とともに対応音源が変化するためである. 例えば、Fig. 4 の \tilde{v}_{f_3} の0と1は、それぞれ赤色と 青色の音源成分を示している. 一方で、 \tilde{v}_{f_4} の0と1

はそれぞれ青色と赤色の音源成分を示している.これ らのサブバンドベクトルの整列は次項で処理される.

3.5.2 フルバンドベクトルの構成

推定されたサブバンドベクトル $\tilde{v}_1, \tilde{v}_2, \cdots, \tilde{v}_I$ から, 次式で定義されたフルバンドベクトルuを構成する.

$$\boldsymbol{u} = (u_1, u_2, \cdots, u_I)^{\mathrm{T}} \in \{0, 1\}^I$$
 (21)



Fig. 5 Reconstruction of fullband permutation label: (a) initialization, (b) second, and (c) last steps.

uの構成処理を Fig. 5 に示す.前項で述べた通り,サ ブバンドベクトル $\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_I$ の2値は同じ意味を 持たない.そのため、「0」と「1」の値がそれぞれ赤色 の音源と青色の音源を示すように,全てのサブバン ドベクトルを統一する必要がある.

Fig. 5 (a) は、フルバンドベクトル u の構成にお ける最初のステップを示している. 図に示すように、 最も低い周波数のサブバンドベクトル \tilde{v}_{i_s} が、フル バンドベクトル u の対応する周波数に挿入される. ここで、 i_s は、最も低い参照周波数のインデクスを 表す. Fig. 5 (a) では、 $i_s = f_3$ 及び $\Omega = 2$ であり、 u_1, u_2, \cdots, u_5 は \tilde{v}_{f_3} により決定される.

Fig. 5 (b) は Fig. 5 (a) の次のステップを示してい る. このステップでは,最も低い周波数のサブバンド に隣接するサブバンドベクトルを算出する.推定され たサブバンドベクトル $\tilde{v}_{i_{s}+1}$ 及びその論理反転ベクト ル $\tilde{v}_{i_{s}+1}$ (Fig. 5 (b) 中の $\tilde{v}_{f_{4}}$ と $\overline{\tilde{v}_{f_{4}}}$)が用意される. 次に, uの一部を

$$\check{\boldsymbol{u}}_i = (u_{i-\Omega}, u_{i-\Omega+1}, \cdots, u_{i+\Omega-1})^{\mathrm{T}} \in \{0, 1\}^{2\Omega}$$
 (22)

の 様 に 定 義 し , $MSE(\tilde{v}_{i_s+1}, \tilde{u}_{i_s+1})$ 及 び $MSE(\tilde{v}_{i_s+1}, \tilde{u}_{i_s+1})$ を比較する. ここで $MSE(\cdot, \cdot)$ は, 2 つの入力ベクトル間の MSE 値である. その結果より, MSE が小さくなるベクトルを \tilde{v}_{i_s+1} 又は $\overline{\tilde{v}_{i_s+1}}$ から選択してメモリに格納する. フルバンドベクトル u は, Fig. 5 (b) に示すように, メモリに格納されたベクトルを基に周波数毎に多数決処理を行って更新される.



Fig. 6 Solving permutation problem.

以降,前述のステップを繰り返し,完全なフルバ ンドベクトル u が, Fig. 5 (c) のように構成される. 提案手法では、 u の構成過程の反復的多数決により、 DNN 推定誤差の悪影響を軽減している.

3.6 フルバンドベクトルに基づく成分の交換

フルバンドベクトル u は, P_i^{-1} の推定値そのもの である.従って,Fig. 6 の様に,[•] **u** に基づいて周波数 毎の分離信号成分を入れ替えることで、パーミュテー ション問題が解決できる.

実験 4

4.1実験条件

高残響下の音声信号の混合を用いて提案パーミュ テーション解決法の性能を評価した.実験では IPS を用いた FDICA [10], ILRMA [6], 及び提案手法を 用いた FDICA の 3 手法を比較した.IPS は,*s_{ii}* を 利用しており,FDICA に基づく BSS の上限性能を示 している.本実験では JVS コーパス [12] の音声信号 (nonpara30) を使用し、これらの音声信号に RWCP データベース [13] の JR2 インパルス応答(詳細は文 献 [10])を畳み込んで,1ファイル当たり 20 s,残響 長 470 ms の音声信号を男性 46 名 95 ファイル及び 女性 48 名 95 ファイル分作成した. STFT は, 窓長 512 ms 及びシフト長 128 ms に設定した.

学習データとして,上記の畳み込み音声信号から2 話者選び,観測信号を作成した.これらに FDICA を 適用し分離した後, IPS でパーミュテーション問題を 解決した **Z**₁ 及び **Z**₂ を作成し,もう一度周波数成分を ランダムにシャッフルすることで,パーミュテーション 問題が残る信号 Y1 及び Y2 と正解ラベルを作成した. 上記の処理で計40万ファイルの学習データを作成し, 学習用及び検証用に2等分した.DNN 学習時の条件 は文献 [8] と同様である. 短時間区間長は r = 40 とし, 短時間区間のストライド幅は $\eta = 4$ とした($\Gamma = 37$ となった). テストデータは, SiSEC2011 [14] の音 声信号 (男性4名及び女性4名) を使用した.これら の音声信号を JR2 インパルス応答で畳み込み処理を 行い,28種類の観測信号を生成した.評価指標には, SDR [11] の改善量を用いた.

4.2 実験結果

各手法の平均 SDR 改善量を Fig. 7 に示す. 文献 [10] で報告されているように, ILRMA の分離性能は 4 dB 以下であるのに対し, IPS を用いた FDICA で は 10 dB 以上の改善を達成している.提案手法を用 いた FDICA は, ILRMA を上回る 8 dB 以上の改善 を達成しており、上限性能に比較的近いものである. これは、DNN に基づくパーミュテーション解決法が、 正しくを予測を行っている結果である.



5 まとめ

本論文では,DNN に基づく頑健なパーミュテーショ ン解決法を提案し、その性能を評価した、提案手法 は、DNN の予測値の示す音源成分を全周波数で統一 するアルゴリズムであり,さらに DNN の予測誤差の 影響を軽減するため時間軸と周波数軸に沿った多数 決を行う.実験結果より,提案手法の有効性が確認さ れた. 今後の課題として, N > 3 の音源数に対応で きるアルゴリズム構築が挙げられる.

謝辞 本研究の一部は JSPS 科研費 19K20306 及び 19H01116の助成を受けたものである.

参考文献

- [1] P. Comon, "Independent component analysis, a new con-
- cept?," Signal Process., vol. 36, no. 3, pp. 287–314, 1994.
 [2] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," Neurocomputing, vol. 22, pp. 21– 34, 1998
- [3] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans.* SAP, vol. 12, no. 5, pp. 530–538, Sep. 2004.
- T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency depen-dencies," *IEEE Trans. ASLP*, vol. 15, no. 1, pp. 70–79, 2007.
- D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation uni-fying independent vector analysis and nonnegative matrix factorization," IEEE/ACM Trans. ASLP, vol. 24, no. 9, pp. 1626–1641, 2016.
- D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation with independent low-rank matrix analysis," in *Audio Source* [6] D. Kitamura, Separation, S. Makino, Ed., pp. 125–155. Springer, Cham, $20\bar{1}8.$
- 2018.
 [7] N. Makishima, S. Mogami, N. Takamune, D. Kitamura, H. Sumino, S. Takamichi, H. Saruwatari, and N. Ono, "Independent deeply learned matrix analysis for determined audio source separation," *IEEE/ACM Trans. ASLP*, vol. 27, no. 10, pp. 1601–1615, 2019.
 [8] 山地修平,北村大地, "局所時間周波数構造に基づく深層パーミュテーション解決法," 日本音響学会 2020 年春季研究発表会講演論文集, pp. 317–320, 2020

- [10] D. Kitamura, N. Ono, and H. Saruwatari, "Experimental analysis of optimal window length for independent low-rank matrix analysis," *Proc. EUSIPCO*, pp. 1210–1214, 2017.
 [11] E. Vincent, R. Gribonval, and C. F, "Performance mea-
- surement in blind audio source separation," IEEE Trans.
- ASLP, vol. 14, no. 4, pp. 1462–1469, 2006.
 [12] S. Takamichi, K. Mitsui, Y. Saito, T. Koriyama, N. Tanji, and H. Saruwatari, "JVS corpus: free Japanese multi-speaker voice corpus," arXiv preprint, 1908.06248, 2019.
- [13] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, "Acoustical sound database in real environments for sound scene understanding and hands-free speech recogni-
- Sound sche understanding and hands hee speech recognition," *Proc. LREC*, pp. 965–968, 2000.
 S. Araki, F. Nesta, E. Vincent, Z. Koldovsky, G. Nolte, A. Ziehe and A. Benichoux, "The 2011 signal separation evaluation campaign (SiSEC2011): -Audio source separation," *Proc. LVA/ICA*, pp. 414-422, 2012.