深層パーミュテーション解決法に基づくブラインド音源分離の性能評価*

☆蓮池郁也，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，事前情報を用いることなく，複数の音源が混合した観測信号から混合前の各音源信号を推定する技術である．BSS は周波数領域独立成分分析（frequency-domain independent component analysis: FDICA）[1] を起源として発展してきた [2]．FDICA には Fig. 1 に示すように，分離信号成分の順序が周波数間で不揃いになる問題が生じる．この問題はパーミュテーション問題と呼ばれる．

BSS の歴史では，様々なパーミュテーション問題解決法（permutation solver: PS）が提案されており，例えば，音源の到来方向（direction of arrivals: DOA）の違いに基づく PS [3] などがある．また，深層ニューラルネットワーク（deep neural network: DNN）を用いた PS（deep PS: DPS）の学習も検討されている [5–9]．特に文献 [6–9] では，多層パーセプトロンや長･短期記憶（long-short term memory: LSTM）を用いた双方向再帰ニューラルネットワーク（bidirectional recurrent neural network using LSTM: BiLSTM）に基づく DPS が検討され，周波数ビン単位のパーミュテーション問題の解決ができることが示された．しかしながら，これらの文献では FDICA が周波数毎の完全な音源分離を達成することを仮定して DPS の性能を評価していた．実際は，FDICA の周波数毎の推定音源には分離誤差が含まれる．そこで本稿では，実際に残響を含む混合信号を FDICA に適用して得られる分離信号に対して，提案 DPS を用いた際の性能について評価する．

2　FDICA とパーミュテーション問題

2.1　信号の定義

短時間 Fourier 変換（short-time Fourier transform: STFT）を適用して得られる時間周波数領域の音源信号，観測信号，及び FDICA の分離信号を次式でそれぞれ表す．

$$\boldsymbol{s}_{ij} = [s_{ij1}, s_{ij2}, \cdots, s_{ijn}, \cdots, s_{ijN}]^{\mathrm{T}} \in \mathbb{C}^N \quad (1)$$

$$\boldsymbol{x}_{ij} = [x_{ij1}, x_{ij2}, \cdots, x_{ijm}, \cdots, x_{ijM}]^{\mathrm{T}} \in \mathbb{C}^M \quad (2)$$

$$\boldsymbol{z}_{ij} = [z_{ij1}, z_{ij2}, \cdots, z_{ijn'}, \cdots, z_{ijN}]^{\mathrm{T}} \in \mathbb{C}^N \quad (3)$$

ここで，$i = 1, 2, \cdots, I$，$j = 1, 2, \cdots, J$，$n = 1, 2, \cdots, N$，$m = 1, 2, \cdots, M$，及び $n' = 1, 2, \cdots, N$ はそれぞれ周波数ビン，時間フレーム，音源信号，観測チャネル，及び分離信号のインデクスを示す．ここで，分離信号は音源の順序が必ずしも n と一致しているとは限らないため，n と n' を使い分けている．また，$^{\mathrm{T}}$ は転置を表す．さらに，分離信号の複素スペクトログラムを $\boldsymbol{Z}_{n'} \in \mathbb{C}^{I\times J}$ と定義する．本稿では，以後 $M = N$ を仮定する．

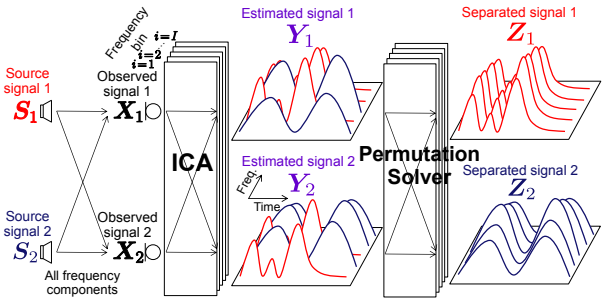

Fig. 1　Permutation problem in FDICA ($N = 2$).

2.2　BSS の定式化と FDICA

FDICA では，観測信号を次式で表す．

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij} \quad (4)$$

ここで，$\boldsymbol{A}_i \in \mathbb{C}^{M\times N}$ は周波数毎の時不変混合行列である．混合行列 $\boldsymbol{A}_i$ が正則であれば，周波数毎の分離行列 $\boldsymbol{W}_i = \boldsymbol{A}_i^{-1} \in \mathbb{C}^{N\times M}$ が存在し，これを用いて理想的な分離信号を次式で表せる．

$$\boldsymbol{z}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij} \quad (5)$$

従って FDICA は，観測信号 $\boldsymbol{x}_{ij}$ の各周波数ビンに対して独立に分離行列 $\boldsymbol{W}_i$ を推定する．

2.3　パーミュテーション問題

FDICA は，分離信号成分の周波数毎のスケール及び順序が不定である．従って，推定分離行列を $\hat{\boldsymbol{W}}_i \in \mathbb{C}^{N\times M}$ とすると，たとえ完全な推定が実現できたとしても，真の分離行列 $\boldsymbol{W}_i$ に対して次式の不定性が残る．

$$\hat{\boldsymbol{W}}_i = \boldsymbol{D}_i \boldsymbol{P}_i \boldsymbol{W}_i \quad (6)$$

ここで，$\boldsymbol{D}_i \in \mathbb{R}^{N\times N}$ は，$\boldsymbol{w}_{in}$ のスケールを変化させる可能性のある対角行列である．また，$\boldsymbol{P}_i \in \{0,1\}^{N\times N}$ は分離行列 $\boldsymbol{W}_i$ の行ベクトル $\boldsymbol{w}_{in}$ の順序を入れ変えうるパーミュテーション行列（置換行列）である．例えば，$N = 2$ であれば $\boldsymbol{P}_i$ は

$$\boldsymbol{P}_i = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (7)$$

の 2 通りの内のいずれかを取る．そのため，FDICA で得られる信号を $\boldsymbol{y}_{ij}$ とすると，次式のように推定信号成分の順序やスケールが周波数間で不揃いである．

$$\boldsymbol{y}_{ij} = \hat{\boldsymbol{W}}_i \boldsymbol{x}_{ij} \quad (8)$$

$$= [y_{ij1}, y_{ij2}, \cdots, y_{ijn'_i}, \cdots, y_{ijN}]^{\mathrm{T}} \in \mathbb{C}^N \quad (9)$$

ここで，$n'_i = 1, 2, \cdots, N$ は周波数ビン i 毎に音源の順序が異なっている状態を表すための新たな音源イ

*Evaluation of blind source separation performance based on deep permutation solver. By Fumiya HASUIKE and Daichi KITAMURA (NIT Kagawa).

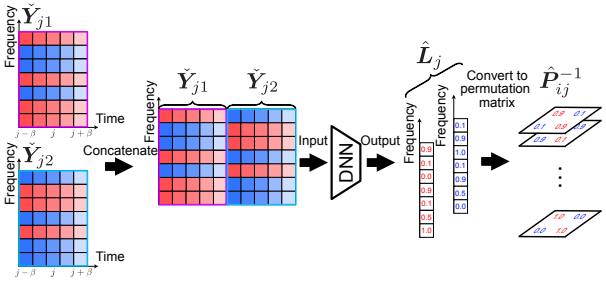

Fig. 2　Estimation of permutation matrix ($N = 2$).

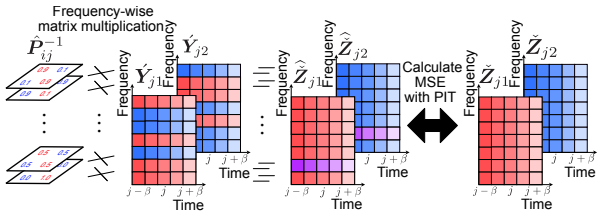

Fig. 3　Loss function using MSE with PIT ($N = 2$).

ンデクスである．$\boldsymbol{D}_i$ で生じる周波数間のスケールの不整合は，プロジェクションバック法 [10] で解析的に復元できる．しかし，$\boldsymbol{P}_i$ で生じる周波数間の音源順序の不整合を全周波数ビンにわたって復元（整列）すること（$\boldsymbol{P}_i^{-1}$ の推定）は容易ではなく，パーミュテーション問題と呼ばれる．Fig. 1 では，FDICA で得られる推定信号 $\boldsymbol{y}_{ij}$ の n' 番目のスペクトログラムを $\boldsymbol{Y}_{n'} \in \mathbb{C}^{I\times J}$ と定義している．

理想的なパーミュテーション問題の解決は

$$\boldsymbol{z}_{ij} = \boldsymbol{P}_i^{-1}\boldsymbol{D}_i^{-1}\boldsymbol{y}_{ij} \tag{10}$$

と表せる．但し厳密には，周波数間の音源順序の整列後も，全周波数をまとめた音源信号全体の順序の不定性は残るため，分離信号は次式となる．

$$\boldsymbol{z}_{ij} = \boldsymbol{P}_{\mathrm{all}}\boldsymbol{P}_i^{-1}\boldsymbol{D}_i^{-1}\boldsymbol{y}_{ij} \tag{11}$$

ここで，$\boldsymbol{P}_{\mathrm{all}} \in \{0,1\}^{N\times N}$ は周波数に非依存なパーミュテーション行列である．本稿では，この音源信号全体の順序の復元は対象としない．

2.4　深層パーミュテーション解決法 [9]

FDICA からはパーミュテーション問題が生じた状態の推定信号の複素スペクトログラム $(\boldsymbol{Y}_{n'})_{n'=1}^{N}$ が得られる．DPS ではまず，これらの信号を次式で正規化パワースペクトログラムに変換する．

$$\overline{\boldsymbol{Y}}_{n'} = \frac{|\boldsymbol{Y}_{n'}|^{.2}}{\sum_{n'}^{N}|\boldsymbol{Y}_{n'}|^{.2}} \in [0,1]^{I\times J} \tag{12}$$

ここで，$|\cdot|^{.2}$ 及び括線はそれぞれ行列の要素毎の絶対値の 2 乗及び要素毎の割り算を示す．次に，$(\overline{\boldsymbol{Y}}_{n'})_{n'=1}^{N}$ から，次式のように時間フレーム j を中心とする局所時間パワースペクトログラムを抽出する．

$$\check{\boldsymbol{Y}}_{jn'} = [\overline{\boldsymbol{y}}_{(j-\beta)n'} \ \cdots \ \overline{\boldsymbol{y}}_{(j+\beta)n'}] \in [0,1]^{I\times(2\beta+1)} \tag{13}$$

ここで，$\overline{\boldsymbol{y}}_{jn'} \in [0,1]^I$ は $\overline{\boldsymbol{Y}}_{n'}$ の j 列目の列ベクトルを表す．また，β（0 以上の整数）は時間フレーム j の近傍時間フレームをどの程度 DNN に入力するかを決めるハイパーパラメータである．DPS では $(\check{\boldsymbol{Y}}_{jn'})_{n'=1}^{N}$ を時間方向に結合した行列 $[\check{\boldsymbol{Y}}_{j1} \ \cdots \ \check{\boldsymbol{Y}}_{jN}] \in [0,1]^{I\times N(2\beta+1)}$ を BiLSTM に入力する（Fig. 2 参照）．各音源の周波数方向の関係性を明確に学習するため，周波数方向に対して BiLSTM を適用する．なお，DNN モデルの詳細は文献 [9] を参照されたい．

DPS は，予測結果として行列 $\hat{\boldsymbol{L}}_j \in [0,1]^{I\times N!}$ を出力する．$\hat{\boldsymbol{L}}_j$ はパーミュテーション行列の予測確率値 $\hat{l}_{iqj} \geq 0$ から構成される行列であり，$q = 1, 2, \cdots, N!$ は N 個の音源に対する $N!$ 通りの順列のインデクスを表す．DNN の出力の際に Softmax 関数を適用することにより，$\hat{\boldsymbol{L}}_j$ の要素は $\hat{l}_{iqj} \geq 0$ かつ $\sum_q \hat{l}_{iqj} = 1$ が保証されている．この時，$N = 2$ を例とすると予測パーミュテーション行列は $\hat{l}_{i1j}$ と $\hat{l}_{i2j}$ を用いて次式のように表せる．

$$\hat{\boldsymbol{P}}_{ij}^{-1} = \hat{l}_{i1j}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} + \hat{l}_{i2j}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} \in [0,1]^{N\times N} \tag{14}$$

推定パーミュテーション行列 $\hat{\boldsymbol{P}}_{ij}^{-1}$ を求めた後の処理を Fig. 3 に示す．ここで，Fig. 3 中の $(|\check{\boldsymbol{Y}}_{jn'}|)_{n'=1}^{N}$ は $(\boldsymbol{Y}_{n'})_{n'=1}^{N}$ の局所時間振幅スペクトログラムである．DNN を用いて求めた予測分離信号 $(|\hat{\check{\boldsymbol{Z}}}_{n'}|)_{n'=1}^{N}$ と予測分離信号に対する正解ラベル $(|\check{\boldsymbol{Z}}_{n'}|)_{n'=1}^{N}$（分離信号 $(|\boldsymbol{Z}_{n'}|)_{n'=1}^{N}$ の局所時間振幅スペクトログラム）を用意し，$(|\hat{\check{\boldsymbol{Z}}}_{n'}|)_{n'=1}^{N}$ と $(|\check{\boldsymbol{Z}}_{n'}|)_{n'=1}^{N}$ の間で損失関数として平均二乗誤差（mean squared error: MSE）を用いる．ここで，DPS は $\boldsymbol{P}_{\mathrm{all}}^{-1}$ の推定を目的としないため，順序不変学習（permutation invariant training: PIT）[11] を導入した損失関数 $\mathcal{L}$ を用いる．

$$\mathcal{L} = \min(C_1, C_2, \cdots, C_q, \cdots, C_{N!}) \tag{15}$$

$$C_q = \sum_{n'}^{N} || \ |\hat{\check{\boldsymbol{Z}}}_{jn'}| - |\check{\boldsymbol{Z}}_{j\mathcal{P}(q,n')}| \ ||_2^2 \tag{16}$$

ここで，$\min(\cdot)$ は入力の最小値を返す関数であり，$\mathcal{P}(q,n')$ は $N!$ 個の全てのありうる順列の内，q 番目の順列における n' 番目の値を返す処理を表す．

パーミュテーション問題は時不変な分離行列 $\hat{\boldsymbol{W}}_i$ で生じることから，正しい音源順序は時間フレーム方向には常に一定である．そのため，テストデータへの適用時は，様々な時間 j の局所時間パワースペクトログラム $(\check{\boldsymbol{Y}}_{jn'})_{n'=1}^{N}$ を DPS に入力し，出力 $(\hat{\boldsymbol{P}}_{ij}^{-1})_{j=1}^{J}$ を次式のように多数決処理することで，更なる精度向上が期待できる．

$$\hat{\boldsymbol{P}}_i^{-1} = \mathrm{round}\left(\frac{1}{J}\sum_{j}^{J}\hat{\boldsymbol{P}}_{ij}^{-1}\right) \in \{0,1\}^{N\times N} \tag{17}$$

ここで，round(·) は入力行列の要素毎の四捨五入を表す．最終的な推定分離信号は次式で得られる．

$$\hat{\boldsymbol{z}}_{ij} = \hat{\boldsymbol{P}}_i^{-1}\boldsymbol{y}_{ij} \tag{18}$$

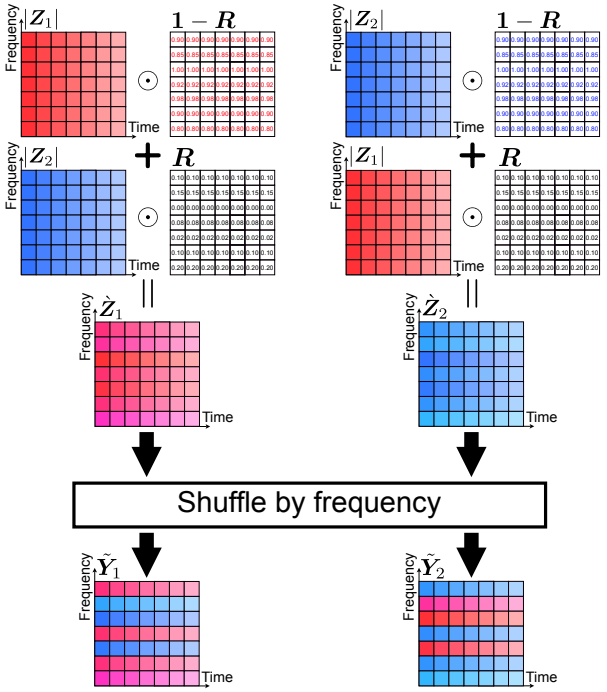

Fig. 4 Simulation of estimation error in FDICA ($N = 2$).

3 提案 DPS の BSS への応用

3.1 提案 DPS を BSS に応用する際の問題と解決法

これまで，音源の種別に依存しづらく汎化性能の高い PS の構築を目的として DPS を検討してきた [5–9]. 特に文献 [9] では，10 秒程度の音楽信号 2 ファイルのみでワンショット学習した DPS が音声信号のパーミュテーション問題を解決できることを示しており，省データで汎化性能を獲得できる可能性が示唆された. しかし，FDICA の推定信号を用いた評価実験は未実施である．実際の音源分離においては，FDICA の推定分離行列 $\hat{\boldsymbol{W}}_i$ が推定誤差を含むため，周波数毎に異なる分離誤差を含む信号に対してパーミュテーション問題を解決する必要がある．現状の DPS を FDICA の分離誤差を含む信号に対して適用した場合，分離誤差の影響によって PS としての性能が低下することが予想される．この問題を解決するために，本稿では FDICA で生じる周波数毎の分離誤差を模倣した信号を DPS の学習データに用いる．また，学習データに鏡像法を用いた様々な残響を付与することで，残響に対する汎化性能の獲得も狙う.

3.2 提案 DPS における学習データ

学習データ作成時には，まず FDICA の推定分離誤差を模倣した振幅スペクトログラム $(|\grave{\boldsymbol{Z}}_{n'}|)_{n'=1}^{N}$ が必要となる．そこで，推定誤差量の相対的な割合を表す時間周波数行列

$$\boldsymbol{R} = \begin{bmatrix} r_1 & r_1 & \cdots & r_1 \\ r_2 & r_2 & \cdots & r_2 \\ \vdots & \vdots & \ddots & \vdots \\ r_I & r_I & \cdots & r_I \end{bmatrix} \in [0, \alpha]^{I \times J} \tag{19}$$

を作成する．$r_1, r_2, \cdots, r_I$ の各要素は区間 $[0, \alpha]$ の一様分布から生成される乱数が割り当てられ，周波数ビン i 毎にランダムな値を持つ．FDICA の推定信号に含まれる分離誤差を模倣した振幅スペクトログラムは，完全分離信号 $(\boldsymbol{Z}_{n'})_{n'=1}^{N}$ を用いて

$$|\grave{\boldsymbol{Z}}_{n'}| = \boldsymbol{R} \odot \left(\sum_{\tilde{n} \neq n'}^{N} |\boldsymbol{Z}_{\tilde{n}}| \right) + (\boldsymbol{1} - \boldsymbol{R}) \odot |\boldsymbol{Z}_{n'}| \tag{20}$$

と表せる．ここで，$\boldsymbol{1}$ はサイズ $I \times J$ の行列であり，全ての要素が 1 である．Fig. 4 に示すように，式 (20) の処理により推定分離誤差を模倣した後に，各周波数において成分を不揃いにすることで，FDICA の分離誤差とパーミュテーション問題を含む振幅スペクトログラム $(|\tilde{\boldsymbol{Y}}_{n'}|)_{n'=1}^{N}$ を作成する.

3.3 提案 DPS の損失関数

提案 DPS では，分離誤差の修正（即ち，FDICA の推定分離行列 $\hat{\boldsymbol{W}}_i$ の精度向上）は行わず，パーミュテーション問題の解決のみを目的とする．そのため，損失値を計上する際に，完全分離信号を用いるのではなく，FDICA の分離誤差のみ（パーミュテーション問題は解決されている）を含む信号 $(|\grave{\boldsymbol{Z}}_{n'}|)_{n'=1}^{N}$ から，対応する局所時間振幅スペクトログラムを抽出したものと DNN を用いて求めた予測分離信号 $(|\widehat{\boldsymbol{Z}}_{n'}|)_{n'=1}^{N}$ との間で損失関数として MSE を用いる．また，これまでの DPS と同様に損失計算時には PIT を導入する．DNN 構造，DNN の出力及びテストデータに対する多数決処理はこれまでの DPS と同一である.

4 実験

4.1 実験条件

提案 DPS の BSS の性能評価をするために，PS を用いない FDICA（PS: none），音源の到来方向 (direction of arrivals: DOA) 情報による PS [3] を用いた FDICA（PS: DOA），提案 DPS を用いた FDICA（PS: DPS），IVA [4]，及び真の音源信号を用いた理想的なパーミュテーション解決法（ideal PS: IPS）を用いた FDICA（PS: IPS）を比較した．FDICA (PS: IPS) は，FDICA に基づく BSS の上限性能を示している.

提案 DPS の学習データには，Table 1 に示すドラムとギターの音楽信号を用いた．音楽信号に対して pyroomacoustics [12] を使用して 100 部屋分のシミュレーションを行った．部屋のサイズは横幅 5 から 12 m，奥行き 5 から 10 m，及び高さ 3 から 5 m の範囲の一様分布から生成される乱数に設定した．壁面の反射回数と反射係数を調節し，T_{60} は 220 ms 程度となるように設定した．2 個のマイクロホンを用いて録音し，マイクロホンの間隔は横軸方向に 5 cm とした．音源とマイクロホンの配置は，高さのみ 1.5 m に固定し，横幅及び奥行きは部屋の範囲内の一様分布から生成される乱数に設定した．但し，2 つの音源とマイクロホンがなす角が必ず 30° 以上になるように設定した．DNN 学習時には，FDICA の分離誤差及びパーミュテーション問題を模倣するために，学習データ中の各

Table 1　Dry sources obtained from SiSEC2011 [13]

Source	Data name	Length
Drums	`dev1_wdrums_src_3.wav`	11.0 s
Guitar	`dev1_wdrums_src_2.wav`	11.0 s

サンプルに対して，異なる $\boldsymbol{R}$ を用いた Fig. 4 の処理を行った．テストデータには，JVS コーパス [14] に含まれる男女の 100 セット分の音声信号 (nonpara30) を用いた．テストデータに対しても pyroomacoustics を用いて，学習データと同一の部屋の生成条件を用いてインパルス応答を生成し，観測信号を作成した．サンプリング周波数は 16 kHz とした．STFT における分析窓関数長（短時間信号長）は 4096 点（256 ms），シフト長は 2048 点（128 ms）と設定し，窓関数にはハン窓を用いた．

提案 DPS における，式 (19) は 0.2 及び式 (13) の β は 13 とした．最適化手法は Adam，ミニバッチサイズは 8，エポック数は 500 とした．評価指標には，信号対歪み比（source-to-distortion ratio: SDR）[15] の改善量を用いた．

4.2　実験結果

Fig. 5 にテストデータ 100 セットにおける各手法のバイオリン図を示す．バイオリン図中の色のついた点は，ひとつのテストデータにおける 2 種類の男女の音声信号の平均 SDR 改善量を示している．中央の白い点は中央値，グレーの縦棒は四分位範囲，曲線はカーネル密度推定分布を表す．Fig. 5 より，提案 DPS のテストデータにおける SDR 改善量の中央値は 7.5 [dB] 程度であり，実際に FDICA を適用した信号に対してパーミュテーション問題をある程度解決できている．IVA の SDR 改善量の中央値は 8.2 [dB] 程度であった．提案 DPS の SDR 改善量の最小値は-2.2 [dB] 程度であり，各手法と比べると比較的 SDR 改善量におけるばらつきが少ない．FDICA（PS: IPS）と比較すると，提案 DPS の性能が劣っていることより，パーミュテーション問題を完璧には解決できておらず，また，提案 DPS の性能の中央値は IVA のそれにやや劣る結果となったが，それでも Table 1 に示すワンショットの音楽信号で学習したモデルが，音声信号のパーミュテーション問題を解決できる結果を示唆していることが分かる．

5　まとめ

本稿では，提案 DPS に基づく BSS の性能評価を行った．実験結果より，実際に FDICA を適用した信号に対してもパーミュテーション問題を解決できる汎化性能の高い DPS が，省サンプルの音響信号で構築できることを示した．今後の課題として，IVA と提案 DPS を組み合わせた PS の実装を行うこと，残響長を変更した際の提案 DPS に基づく BSS の性能調査等が挙げられる．

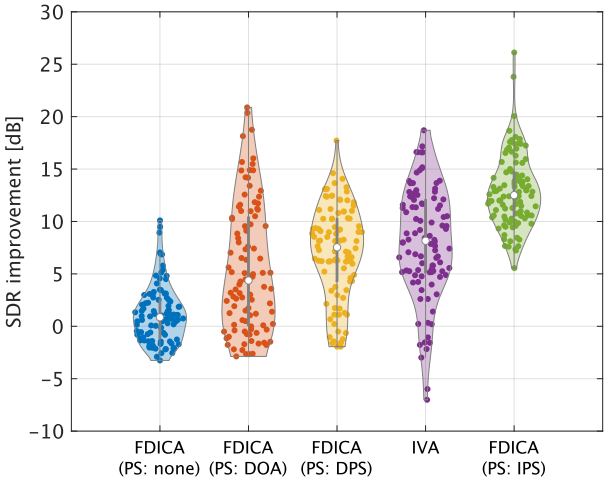

Fig. 5　Violin plot of SDR improvements.

謝辞　本研究の一部は JSPS 科研費 23K24908 の助成を受けたものである．

参考文献

[1] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain,” *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[2] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, and H. Saruwatari, “A review of blind source separation methods: Two converging routes to ILRMA originating from ICA and NMF,” *APSIPA TSIP*, vol. 8, no. e12, pp. 1–14, 2019.

[3] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, “Blind source separation based on a fast-convergence algorithm combining ICA and beamforming,” *IEEE TASLP*, vol. 14, no. 2, pp. 666–678, 2006.

[4] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” *Proc. WASPAA*, pp. 189–192, 2011.

[5] S. Yamaji and D. Kitamura, “DNN-based permutation solver for frequency-domain independent component analysis in two-source mixture case,” *Proc. APSIPA ASC*, pp. 781–787, 2020.

[6] 蓮池郁也, 渡辺瑠伊, 北村大地, “深層ニューラルネットワークに基づくパーミュテーション解決法の基礎的検討,” **信学技報**, EA2022-13, vol. 122, no. 20, pp. 62–67, 2022.

[7] 蓮池郁也, 北村大地, 渡辺瑠伊, “深層パーミュテーション解決法の汎化性能に関する実験的評価,” **日本音響学会** *2022* **年秋季研究発表会講演論文集**, pp. 351–354, 2022.

[8] F. Hasuike, D. Kitamura, and R. Watanabe, “DNN-based frequency-domain permutation solver for multichannel audio source separation,” *Proc. APSIPA ASC*, pp. 872–877, 2022.

[9] 蓮池郁也, 北村大地, 渡辺瑠伊, 川口翔也, “周波数双方再帰に基づく深層パーミュテーション解決法,” **電子情報通信学会 第** *37* **回信号処理シンポジウム**, A13–2, pp. 308–313, 2022.

[10] K. Matsuoka and S. Nakashima, “Minimal distortion principle for blind source separation,” *Proc. ICA*, pp. 722–727, 2001.

[11] D. Yu, M. Kolbak, Z.-H. Tan, and J. Jensen, “Permutation invariant training of deep models for speaker-independent multi-talker speech separation,” *Proc. ICASSP*, pp. 241–245, 2017.

[12] R. Scheibler, E. Bezzam, and I. Dokmanić, “Pyroomacoustics: a Python package for audio room simulation and array processing algorithms,” *Proc. ICASSP*, pp. 351—355, 2018.

[13] S. Araki, F. Nesta, E. Vincent, Z. Koldovský, G. Nolte, A. Ziehe, and A. Benichoux, “The 2011 signal separation evaluation campaign (SiSEC2011):-audio source separation,” *Proc. LVA/ICA*, pp. 414–422, 2012.

[14] 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.

[15] E. Vincent, R. Gribonval, and C. Févotte, “Performance measurement in blind audio source separation,” *IEEE TASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.