1-R-25

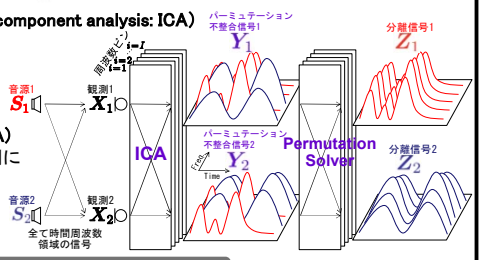
深層パーミュテーション解決法に基づくブラインド音源分離の性能評価

☆蓮池郁也，北村大地（香川高専）

1. 研究背景

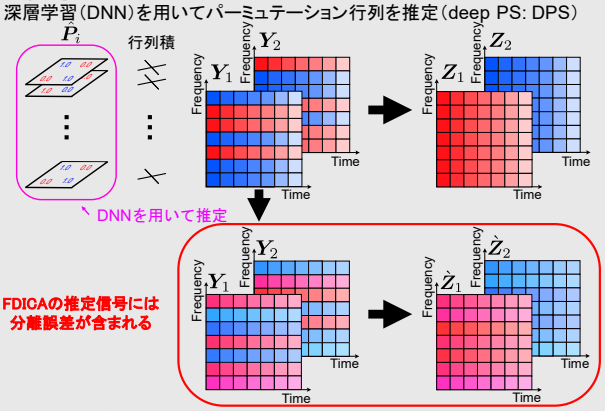
- **ブラインド音源分離（blind source separation: BSS）**
  - 観測信号をもとに特定の信号を推定する技術

- **独立成分分析（independent component analysis: ICA）**
  - 混合行列が未知の条件で分離行列を推定するが，分離信号の順序が不定
- **周波数領域ICA（frequency-domain ICA: FDICA）**
  - 各周波数ビンの複素時系列に対して独立なICAを適用
  - パーミュテーション問題が発生



本研究の動機及び目的

- **汎化性能の高いパーミュテーション解決法（permutation solver: PS）を実現** [Hasuike+, 2022]
  - 深層学習（DNN）を用いてパーミュテーション行列を推定（deep PS: DPS）



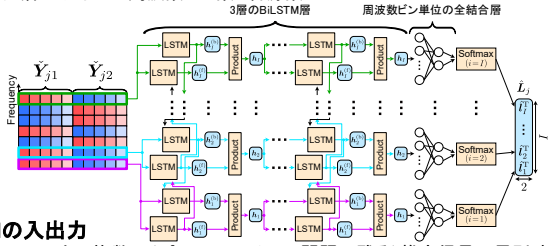
- **FDICAを適用後の推定信号に含まれる分離誤差に対する性能調査**
  - 分離誤差をシミュレーションした学習データを用いる

2. 深層パーミュテーション解決法（DPS）

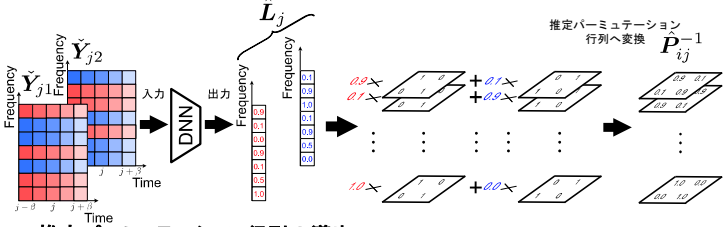
- **パーミュテーション不整合信号に対する正規化処理** [Sawada+, 2007]
  - 同一音源の成分の相関を強調かつDNNの入力の値を区間 [0, 1]に制限
  - DNNの学習が安定する

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

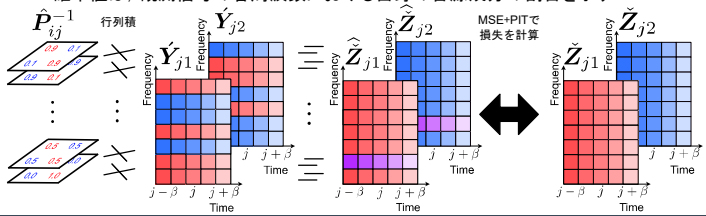
- **DNNの構造** [蓮池ら, 2022]
  - 入力層，BiLSTM層3層，出力層の計5層で構成
  - 各音源の周波数方向の関係性を明確に学習するために，BiLSTMを適用
  - 出力層にSoftmax関数を用い各周波数成分の値が足して1になるよう制約
  - 3層のBiLSTM層の後は，次元を圧縮するために全結合層を適用
  - 出力層のサイズは周波数ビン数×音源数



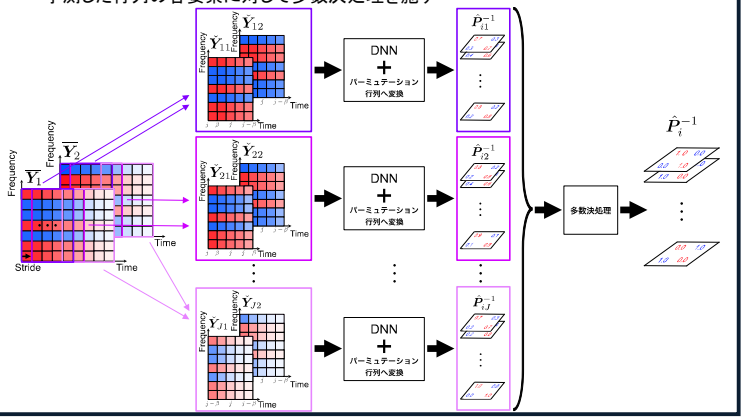
- **DNNの入出力**
  - DNNの入力に複数の（パーミュテーション問題の残る）推定信号の局所時間をとってきてベクトル化し，結合したものを使用
  - DNNの出力は各周波数における確率値であり，パーミュテーション行列の係数



- **推定パーミュテーション行列の導出**
  - DNNの出力値（確率値）をパーミュテーション行列の係数として利用
  - 確率値は，観測信号の各周波数における自身の音源成分の割合を示す



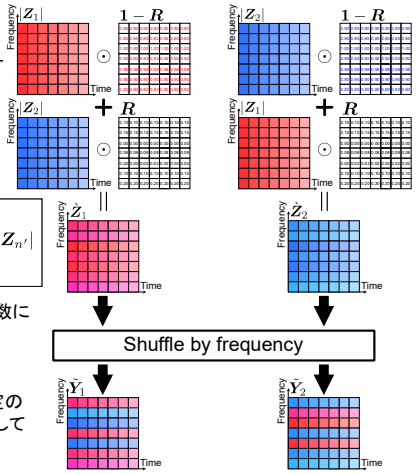
- **テストデータに対する多数決処理**
  - 各局所時間スペクトログラムに対してDNNで予測し，推定パーミュテーション行列を作成
  - 予測した行列の各要素に対して多数決処理を施す



3. 提案DPSにおける学習データ

- **FDICAで生じる周波数毎の分離誤差をシミュレーション**
  - 推定誤差量の相対的な割合を表す時間周波数行列 R を用意
  - R と完全分離信号 $(Z_{n'})_{n'=1}^{N}$ を用いてFDICAの推定信号に含まれる分離誤差を模倣した信号を作成

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



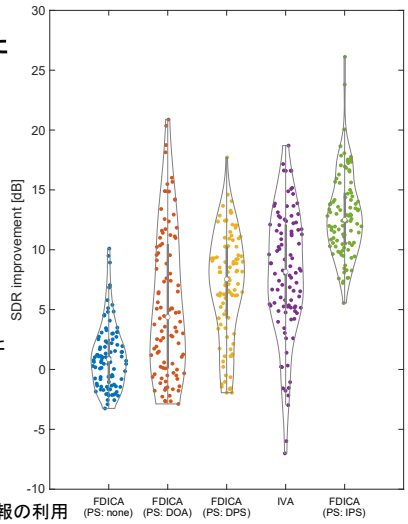
  - 分離誤差を模倣した後に，各周波数において成分を不揃いにすることで学習データを生成
  - FDICAを適用した推定信号は各周波数において分離誤差が一定のため R の要素は周波数方向に対して統一

4. 実験

- **従来手法とsource-to-distortion-ratio（SDR）の改善量を比較**
  - PSを用いないFDICA
  - 音源の到来方向（direction of arrivals: DOA）情報によるPSを用いたFDICA [Saruwatari+, 2006]
  - 提案DPSを用いたFDICA
  - 独立ベクトル分析（independent vector analysis: IVA） [Ono, 2011]
  - 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（これはFDICAに基づくBSSの上限性能を示す）
- **実験条件**
  - 学習データには，SiSEC2011のドラムとギターの11秒程度の2つの音楽信号を使用
  - 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
  - 2個のマイクロホンを用いており，間隔は5 cm
  - 音源とマイクロホンの配置は高さのみ固定（1.5 m）で，高さ以外は乱数値で設定
  - Pyroomacousticsで100部屋分のシミュレーション
  - 2つの音源とマイクロホンがなす角は必ず30°以上になるように設定
  - テストデータにはJVSコーパスの男女の100セット分の音声信号を使用

FFT長	256 ms（ハミング窓）
シフト長	128 ms
エポック数	500
T_{60}	220 ms

- **実験結果（テストデータ100セットにおける各手法のバイオリン図）**
  - 提案DPSのSDR改善量の中央値は7.5 [dB]
  - IVAのSDR改善量の中央値は8.2 [dB]
  - FDICA（PS: IPS）のSDR改善量の中央値は12.6 [dB]
  - 提案DPSのSDR改善量の最小値は−2.2 [dB] 程度であり，従来手法と比べると，ばらつきが少ない
  - 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た
  - FDICA（PS: IPS），IVAと比較すると提案DPSの性能が劣っている



- **今後の展望**
  - 残響長を変更した際の性能調査
  - スペクトログラムに含まれる位相情報の利用
  - DNN学習時に，信号を3次元方向に結合してBiLSTMに入力
  - IVAと提案DPSを組み合わせたパーミュテーション解決法の実装