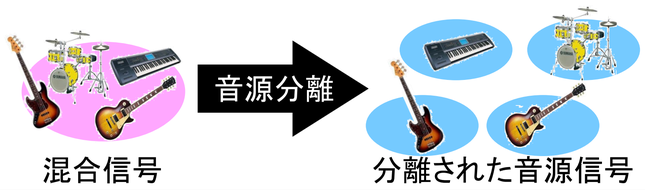
2-R2-13

# スペクトログラム無矛盾性を用いた独立低ランク行列分析

☆豊島直(香川高専), 北村大地(香川高専), 矢田部浩平(早稲田大)

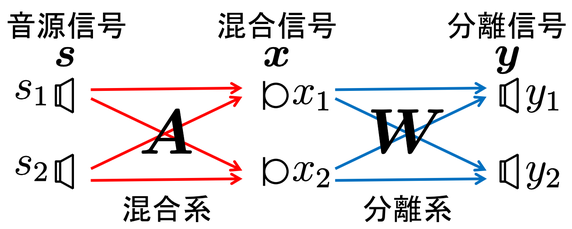
## 1. 研究背景

- **音源分離**：複数の音声や楽器音の混合音から個々の音源を推定する技術




**応用例**
- 自動採譜の前段処理
- 音声認識の精度向上etc.

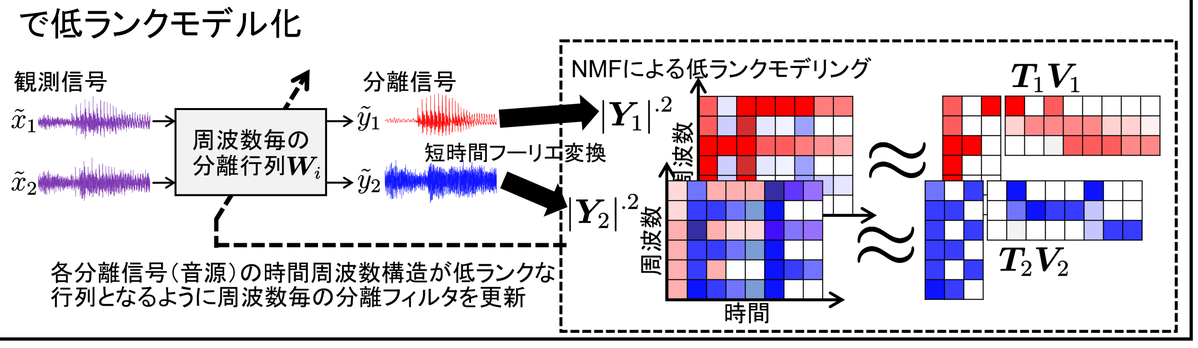
- **ブラインド音源分離（BSS）**：マイクや音源の位置，混合系が未知の音源分離




**独立成分分析（ICA）**[Common, 1994]に基づく手法が一般的

分離音$\boldsymbol{y}$を求める為の分離行列$\boldsymbol{W}$を推定

- **独立低ランク行列分析**（independent low-rank matrix analysis: ILRMA）:[Kitamura+, 2016]
周波数ビン毎のICA＋各音源の時間周波数構造を非負値行列因子分解（NMF）で低ランクモデル化




**解決すべき問題**

ILRMAの周波数毎の推定分離行列$\boldsymbol{W}_i$で得られる時間周波数領域の分離信号は**スペクトログラム無矛盾性**が損なわれている

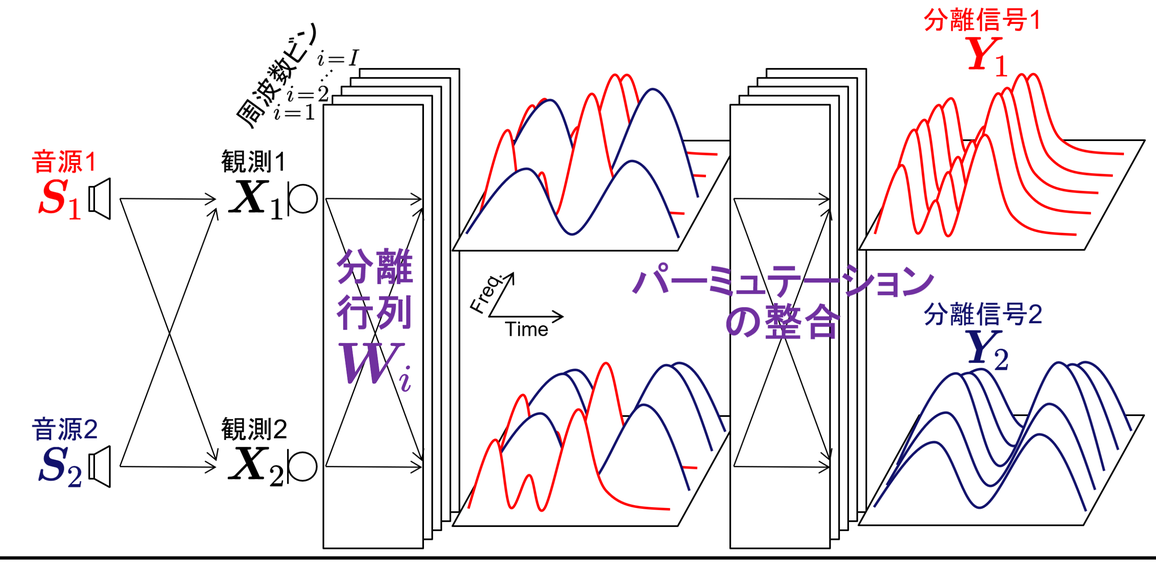
**本研究の目的**

ILRMAの分離行列推定時に**スペクトログラム無矛盾性**を担保した新しいアルゴリズムを提案し，分離性能の向上に寄与するかを実験的に調査

## 2. 従来手法

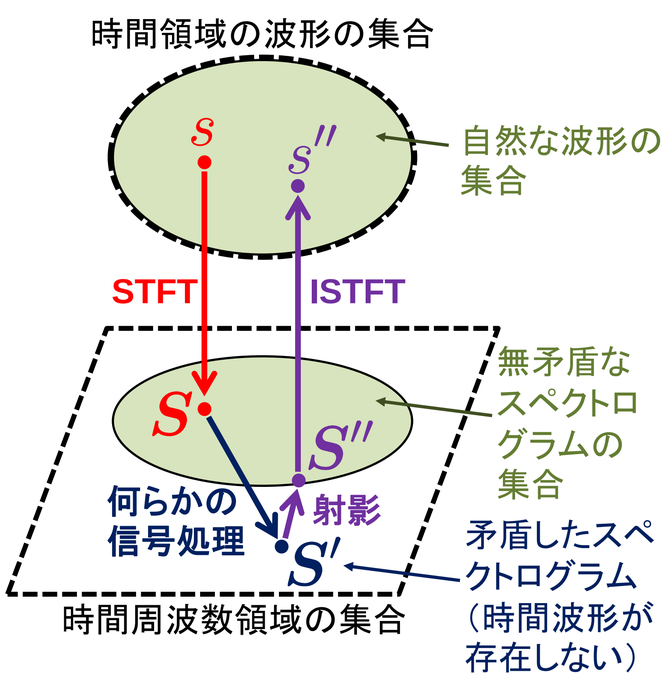
- **BSSにおけるパーミュテーション問題**

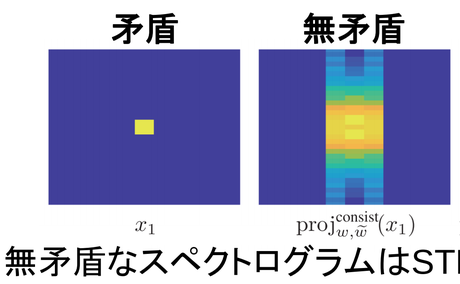
ICAを周波数毎の複素時系列に適用して周波数毎に音源分離することを考える
周波数によって分離信号の音源順序が変わってしまう**パーミュテーション問題**
ILRMAは「分離信号が低ランクな時間周波数構造を持つ」という仮定を導入し回避



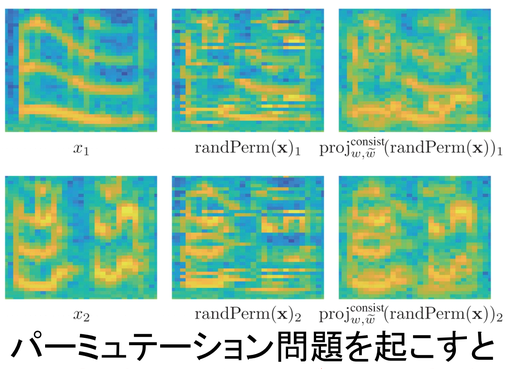

- **BSSにスペクトログラム無矛盾性を導入** [K.Yatabe, 2020]

時間波形を短時間フーリエ変換（STFT）するとスペクトログラムが得られ，スペクトログラムを逆STFT（ISTFT）すると時間波形に戻る（完全再構成条件を満たす場合）
スペクトログラムを信号処理で加工すると，**矛盾したスペクトログラム**となる
矛盾したスペクトログラムをISTFTすると無矛盾な領域に射影され時間波形に戻る
BSSの分離行列の最適化の反復毎に
「ISTFT→STFT」という処理を挿入することで
パーミュテーション問題を少し回避しながら推定




無矛盾なスペクトログラムはSTFTの窓掛けやオーバーラップシフトで時間周波数の両方向に**<u>滲んでいる</u>**




パーミュテーション問題を起こすと**<u>矛盾したスペクトログラムになる</u>**

## 3. スペクトログラム無矛盾性に基づくILRMA

- **ILRMAへのスペクトログラム無矛盾性の適用**

Consistent ILRMAのアルゴリズム

**Input:** $\{\boldsymbol{x}_{ij}\}_{i=1,j=1}^{I,J}$, maxIter
**Output:** $\{\boldsymbol{y}_{ij}\}_{i=1,j=1}^{I,J}$

Initialize $\{\boldsymbol{T}_n\}_{n=1}^{N}, \{\boldsymbol{V}_n\}_{n=1}^{N}, \{\boldsymbol{W}_i\}_{i=1}^{I}$
**for** iter $= 1, 2, \cdots,$ maxIter **do**
$\boldsymbol{Y}_n \leftarrow \mathrm{STFT}_{\omega}(\mathrm{ISTFT}_{\tilde{\omega}}(\boldsymbol{Y}_n))$ (21)
$t_{ikn} \leftarrow t_{ikn}\sqrt{\dfrac{\sum_j |y_{ijn}|^2 \left(\sum_{k'} t_{ik'n} v_{k'jn}\right)^{-2} v_{kjn}}{\sum_j \left(\sum_{k'} t_{ik'n} v_{k'jn}\right)^{-1} v_{kjn}}}$ (19)
$v_{kjn} \leftarrow v_{kjn}\sqrt{\dfrac{\sum_i |y_{ijn}|^2 \left(\sum_{k'} t_{ik'n} v_{k'jn}\right)^{-2} t_{ikn}}{\sum_i \left(\sum_{k'} t_{ik'n} v_{k'jn}\right)^{-1} t_{ikn}}}$ (20)
$\boldsymbol{U}_{in} \leftarrow \dfrac{1}{J}\sum_j \dfrac{1}{\sum_k t_{ikn} v_{kjn}} \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{H}}$ (15)
$\boldsymbol{w}_{in} \leftarrow (\boldsymbol{W}_i \boldsymbol{U}_{in})^{-1} \boldsymbol{e}_n$ (16)
$\boldsymbol{w}_{in} \leftarrow \boldsymbol{w}_{in}\left(\boldsymbol{w}_{in}^{\mathrm{H}} \boldsymbol{U}_{in} \boldsymbol{w}_{in}\right)^{-\frac{1}{2}}$ (17)
$y_{ijn} \leftarrow \boldsymbol{w}_{in}^{\mathrm{H}} \boldsymbol{x}_{ij}$ (18)
$\boldsymbol{w}_{in} \leftarrow \boldsymbol{w}_{in}\lambda_{inm_{\mathrm{ref}}}$ (23)
$y_{ijn} \leftarrow \boldsymbol{w}_{in}^{\mathrm{H}} \boldsymbol{x}_{ij}$ (24)
$t_{ikn} \leftarrow t_{ikn}|\lambda_{inm_{\mathrm{ref}}}|^2$ (25)
**end for**

矛盾した推定分離スペクトログラム$\boldsymbol{Y}_n$を無矛盾な領域へ射影

**原稿のAlgorithm 1に誤植あり**
(15)~(18)と(19)~(20)の順番が逆になっています

**従来のILRMA**
4，5行目でNMF音源モデルの最適化
6～9行目で分離行列の最適化

毎反復においてプロジェクションバックを適用

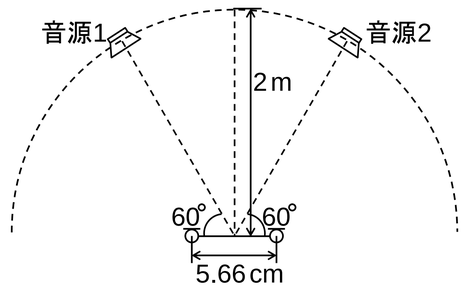
GitHub: https://github.com/d-kitamura/ILRMA/blob/master/consistentILRMA.m

$\boldsymbol{w}_{in}$は$\boldsymbol{W}_i$の行ベクトル，$y_{ijn}$は$\boldsymbol{Y}_n$の$(i,\ j)$要素，$t_{ikn}$は$\boldsymbol{T}_n$の$(i,\ k)$要素，$v_{kjn}$は$\boldsymbol{V}_n$の$(k,\ j)$要素，$\lambda_{inm_{\mathrm{ref}}}$は$m_{\mathrm{ref}}$にプロジェクションバックして求まる補正係数
$i$は周波数ビンインデクス，$j$は時間フレームインデクス，$k$は基底ベクトルインデクス

## 4. 比較実験

- **実験条件**

| | |
|---|---|
| 窓関数 | ハン窓 |
| 窓長 | 64, 128, 256, 512, 768, 1024 ms |
| シフト長 | 窓長の1/16 または 1/2 |
| 基底数 | 音楽10，スピーチ2 |
| 初期値 | $\boldsymbol{W}_i$：単位行列<br>$\boldsymbol{T}_n$ and $\boldsymbol{V}_n$：乱数行列 |
| 反復回数 | 100 回 |
| 試行回数 | 乱数シードを変えて5回 |
| 参照マイクチャネル | $m_{\mathrm{ref}} = 1$ |



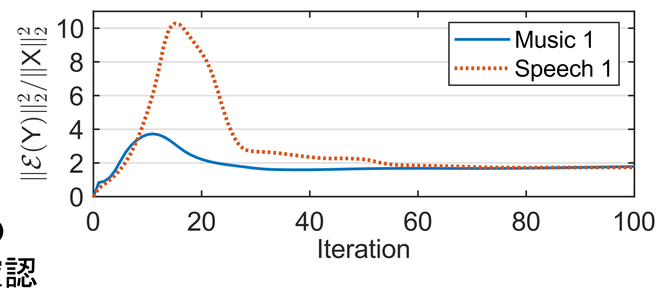

JR2インパルス応答（RWCP）
(残響時間: $T_{60}$ = 470 ms)

- **実験結果**

右図は観測パワーで正規化された反復毎の矛盾成分量（二乗L2ノルム）

$$\mathcal{E}(\mathrm{Y}) = \mathrm{Y} - \mathrm{STFT}_{\omega}(\mathrm{ISTFT}_{\tilde{\omega}}(\mathrm{Y}))$$
$$\mathrm{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2],\quad \mathrm{Y} = [\boldsymbol{Y}_1, \boldsymbol{Y}_2]$$

一度大きな値に上昇するが，最適化の後半では一定の値に収束することを確認




右下図は従来手法と提案手法（前章のアルゴリズムの3行目の有無）の比較

**<u>音楽信号の考察</u>**
STFTの窓長が長い場合提案手法が明らかに従来手法を上回る

シフト長の違いによる傾向差はあまり確認できず

**<u>音声信号の考察</u>**
STFTの窓長が512 msの場合に提案手法が従来手法を上回る

シフト長の違いによる傾向差はあまり確認できず

**<u>全体的な考察</u>**
音源分離が成功する程提案手法の有効性が顕著になることを確認

**分離が成功した場合，推定スペクトログラム$\boldsymbol{Y}_n$は無矛盾なスペクトログラムに近づくため**と推測される