

Importance of Regularization in Superresolution-Based Multichannel Signal Separation with Nonnegative Matrix Factorization

Daichi Kitamura, Hiroshi Saruwatari, Kiyohiro Shikano (Nara Institute of Science and Technology)
Kazunobu Kondo, Yu Takahashi (Yamaha Corporation)

1. 研究背景

- 近年、音楽信号分離技術が注目されている

応用例

- 自動採譜
- 音の拡張現実 (AR) etc.

先行研究

- Nonnegative matrix factorization (NMF) [1]**

モノラル信号が対象

各楽器音の調波構造を利用して特定楽器音のみを抽出
分解したスペクトル基底を楽器ごとに分類することが困難

- Multichannel NMF [2]**

NMF をマルチチャンネル信号分解に拡張
各楽器音の調波構造とチャンネル間の位相情報を利用
分解行列の初期値依存性が強く、頑健性に欠ける

- 方位クラスタリング[3]**

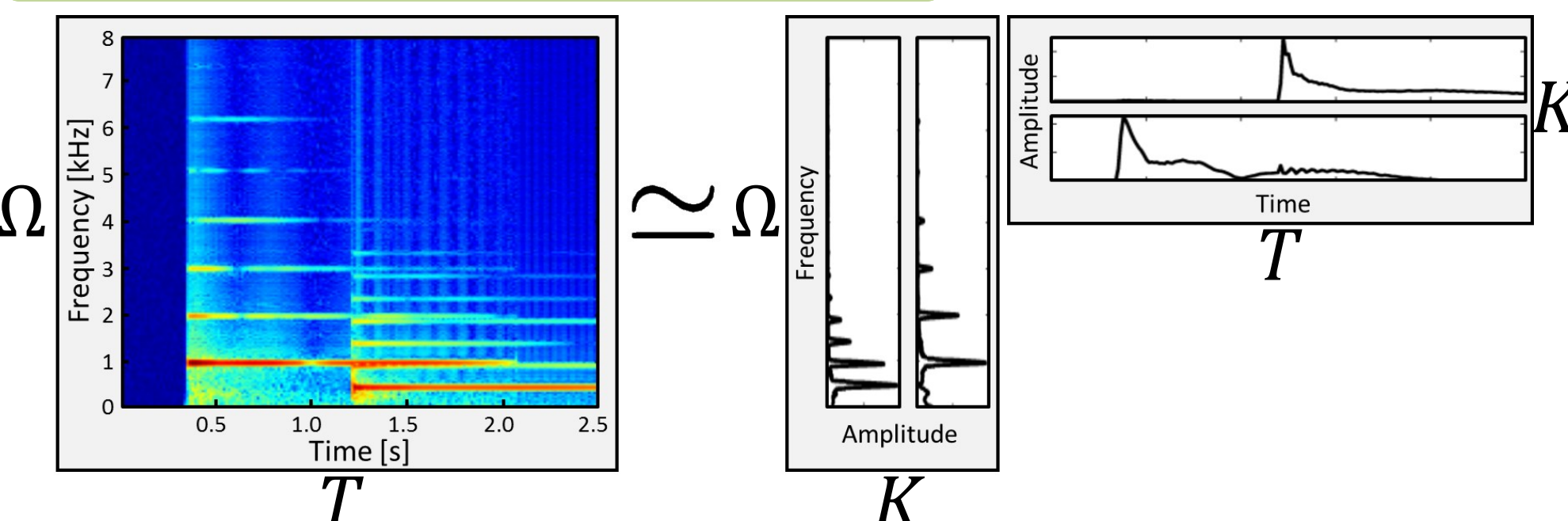
マルチチャンネル音源を対象とし、音源の方位ごとに分離する
同一方位に存在する音源同士は分離できない

本研究の目的

より頑健なマルチチャンネル信号分解手法として、
前段に方位情報に関する分解、後段に教師あり NMF による分解を行う音源分離手法を提案する

2. NMF

音楽音源分離における NMF



Y : 観測信号行列

A : 基底行列

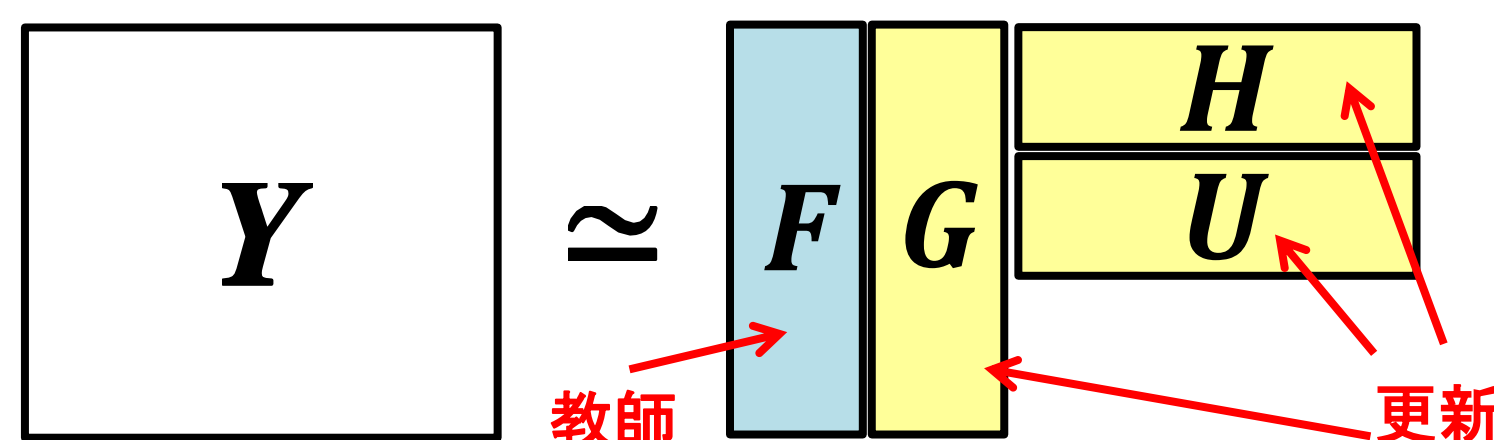
B : アクティベーション行列

Ω : 周波数ビン数

T : フレーム数

K : 基底数

罰則条件付き教師あり NMF (Penalized Supervised NMF: PSNMF) [4]



F : 目的音信号に含まれるスペクトルパターンに対応する基底行列 (教師情報)

G : 教師基底 F に対応するアクティベーション行列

H : 非目的音信号に含まれるスペクトルパターンに対応する基底行列

U : H に対応するアクティベーション行列

F と H に同じスペクトルパターンが現れないよう、 F と H の間に互いが無相関になるような罰則条件を導入する。

コスト関数

$$\mathcal{J}_{\text{PSNMF}} = \mathcal{D}(Y|FG + HU) + \mu \|F^T H\|_{\text{Fr}}^2$$

subject to $g_{k,t} \geq 0, h_{\omega,l} \geq 0, u_{l,t} \geq 0$ for all ω, k, l, t ,
where $g_{k,t}, h_{\omega,l}$, and $u_{l,t}$ are entries of the matrices F, G, H, U .

PSNMF の問題点

- 分離対象信号に含まれる楽器数が少ない場合 (2~3 つ程度) は精度良く分離できる
- 分離対象信号に含まれる楽器数が多い場合 (4 つ以上) は分離精度が極端に落ちる
- 楽器数が多い場合、楽器間で似たスペクトルパターンが多く現れることが原因

3. 方位クラスタリング

ステレオ信号における推定

- ステレオチャンネル入力信号

$$X(\omega, t) = [X_R(\omega, t), X_L(\omega, t)]^T$$

- 空間代表ベクトル

$$H_n(\omega, t) = [H_{nR}(\omega, t), H_{nL}(\omega, t)]^T$$

ω : 周波数ビン, t : 時間フレーム

- 音楽信号に含まれる各成分 $X(\omega, t)$

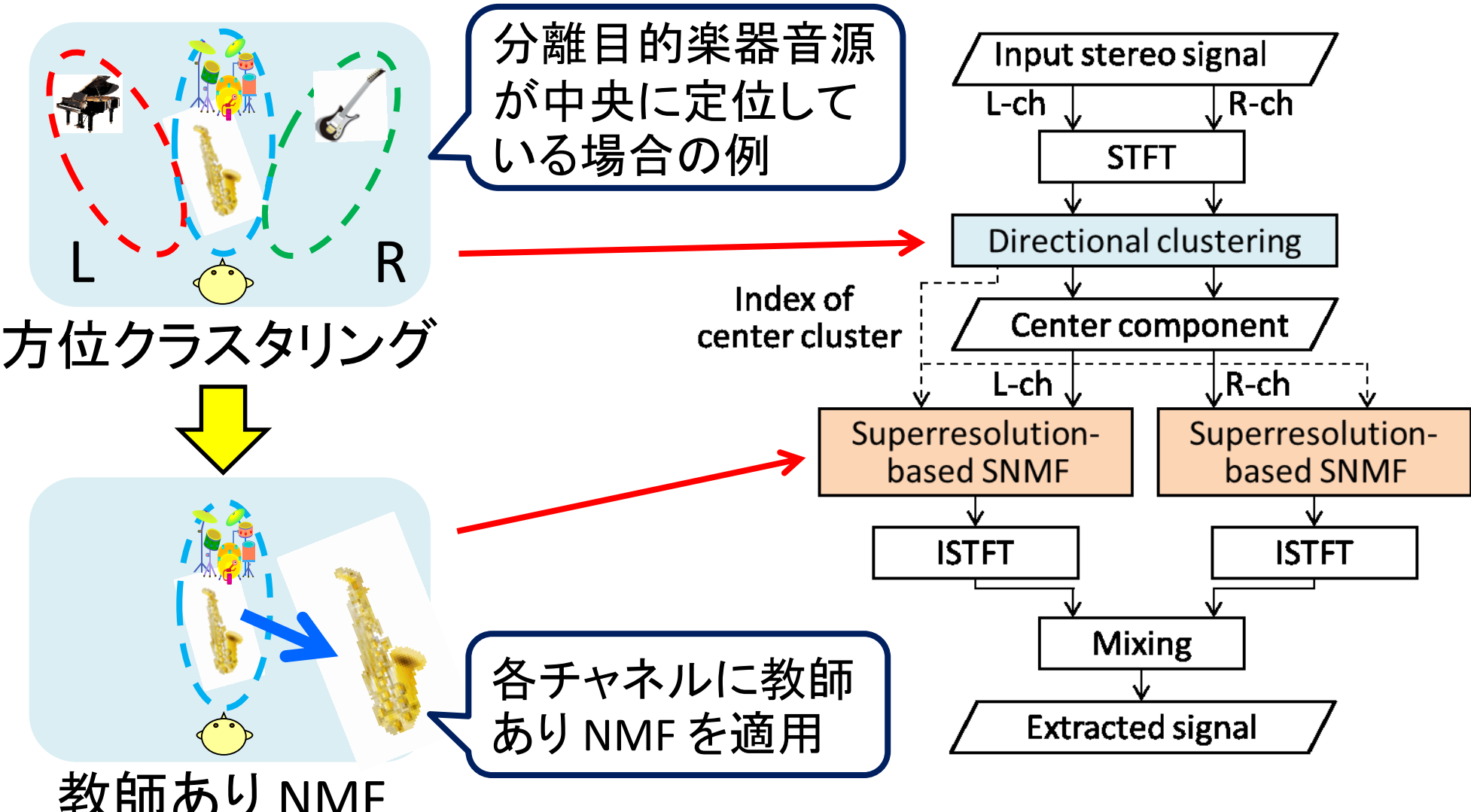
空間代表ベクトル $H_n(\omega, t)$ ($n = 1, \dots, N$: N は空間代表ベクトル数)

クラスタリング結果から生成されるバイナリマスクをスペクトログラムにかけることで、特定の方位成分のみを分離できる

4. 提案手法

提案手法の概要

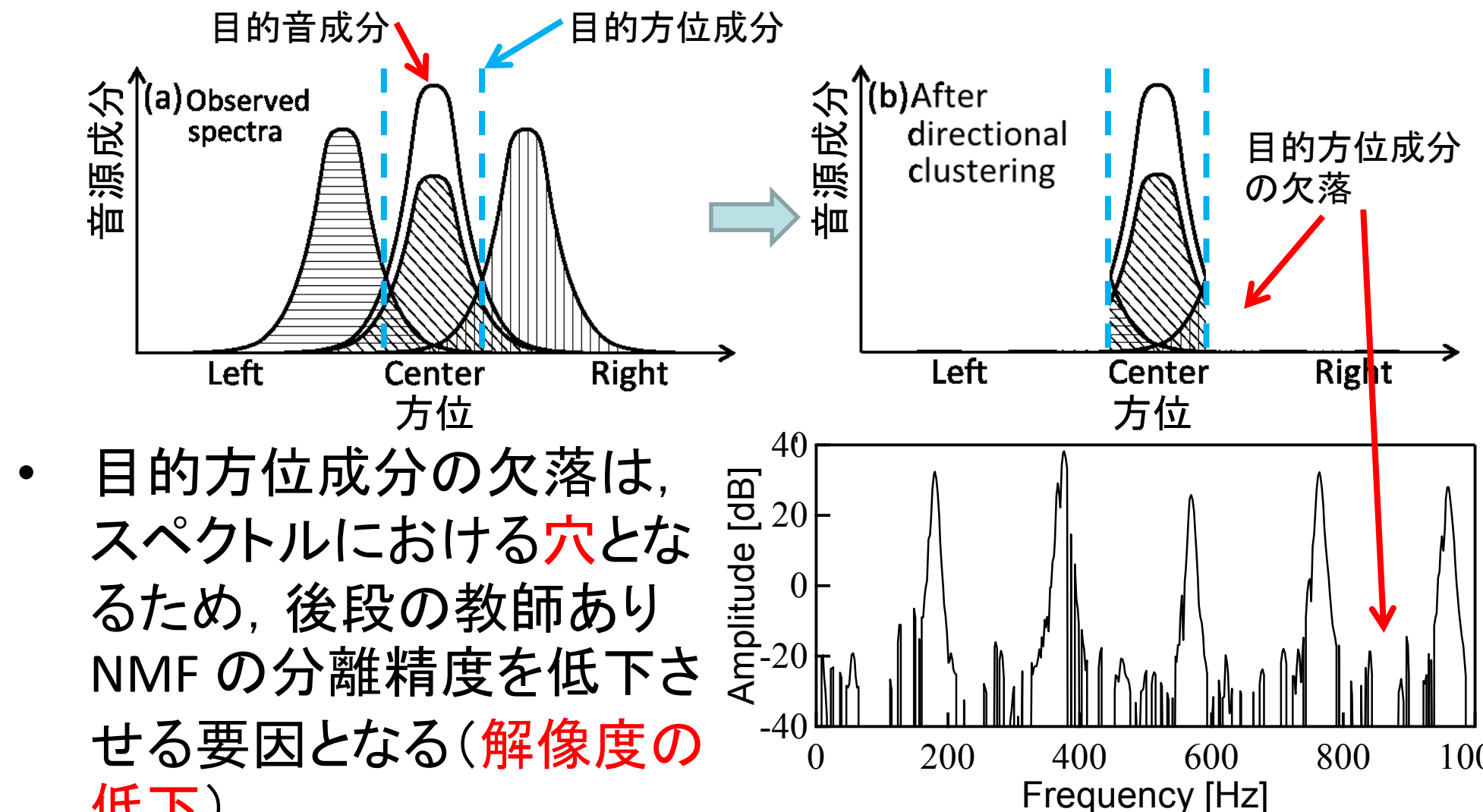
- 教師あり NMF と方位クラスタリングを組み合わせ、より確実に動作するマルチチャンネル信号分解手法を提案する
- 前段の方位クラスタリングにより、分離目的楽器音源が含まれる方位のみを抽出
- 分解された同一方位の音源成分に対して、後段に教師あり NMF による分解を行い目的楽器音成分を抽出



具体的な処理内容

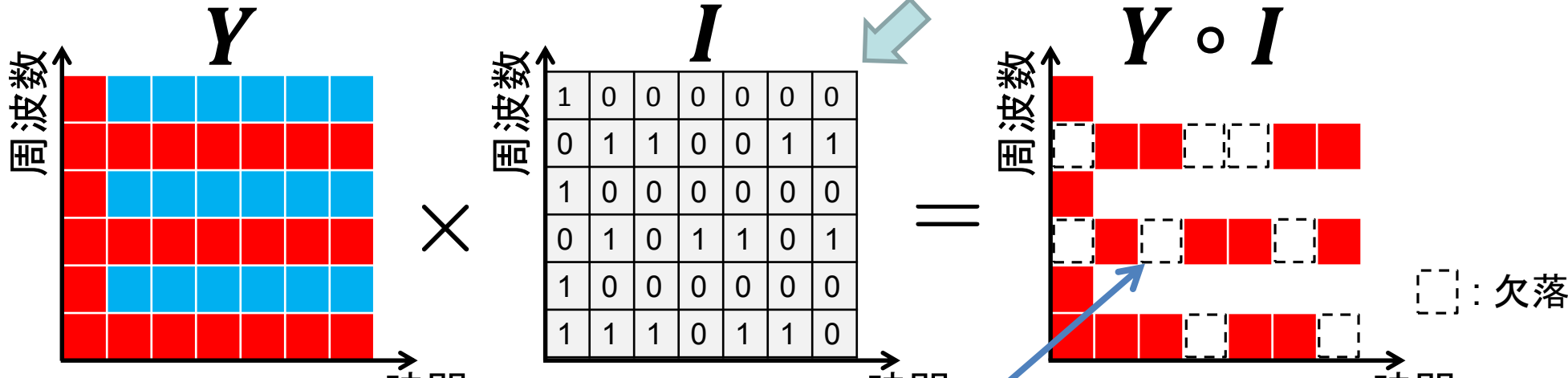
(1) 方位クラスタリングステップ

- 方位クラスタリングはバイナリマスク的に分解するため、目的方位の成分の欠落と非目的方位成分の混入が生じる

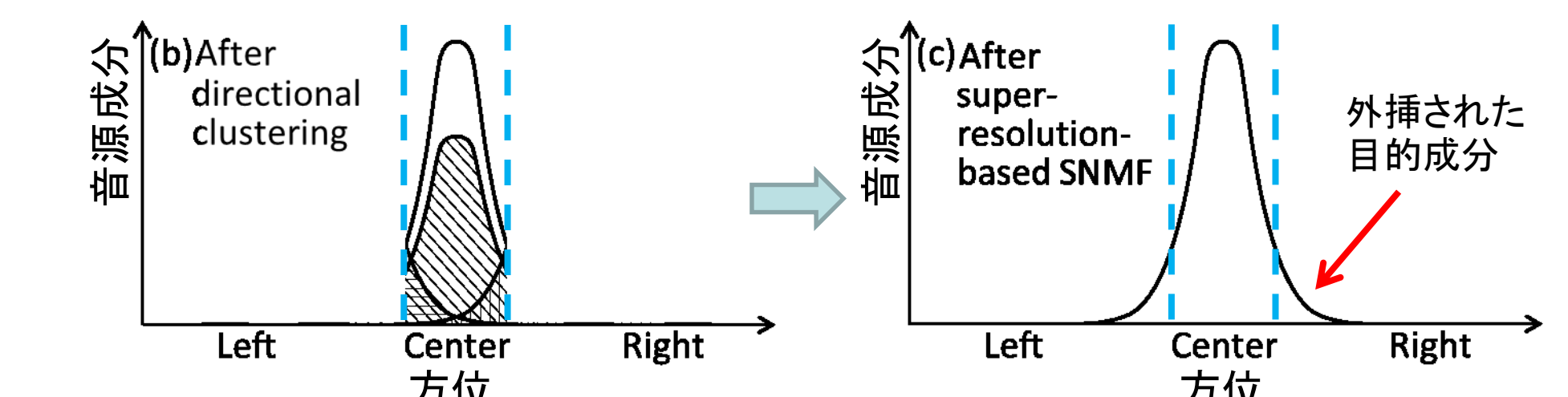


(2) 教師あり NMF による超解像ステップ

- 後段の NMF においては、欠落した成分は見えなかったものとし、観測可能な成分のみで教師基底のマッチングを行う



- 選ばれた教師基底によって、欠落した目的音成分は外挿される
- 方位クラスタリングによって劣化していた目的音成分の解像度が、本教師あり NMF によって復元される (これはスペクトログラムの超解像に相当する)



コスト関数

$$\mathcal{J}_1 = \mathcal{D}(I \circ Y | I \circ (FG + HU)) + \mu \|F^T H\|_{\text{Fr}}^2$$

subject to $g_{k,t} \geq 0, h_{\omega,l} \geq 0, u_{l,t} \geq 0$ for all ω, k, l, t .

Index I によって、NMF は低解像度データに引きずられず、最終的に信号を外挿して解像度を復元

超解像における正則化

基底外挿による超解像における問題

- 観測可能な成分が極端に少ないフレームにおいては、いかなる教師基底もマッチングできるが、間違った教師基底によってスペクトルが外挿される可能性がある
- このような外挿誤りを避ける為に、何らかの正則化が必要

ノルム最小化による正則化

穴が極端に多いフレーム → 本来目的音成分が存在していなかった

- 各時間フレームにおいて、欠落グリッドの数が多いほど出力音が小さくなるような教師基底を選ぶ正則化項を付与する

コスト関数

$$\mathcal{J}_2 = \mathcal{D}(I \circ Y | I \circ (FG + HU)) + \mu \|F^T H\|_{\text{Fr}}^2 + \lambda \|\bar{I} \circ (FG)\|_{\text{Fr}}^2$$

subject to $g_{k,t} \geq 0, h_{\omega,l} \geq 0, u_{l,t} \geq 0$ for all ω, k, l, t .

5. 乗法更新式

補助関数法による更新式の導出

- 解析的に最適な変数を導出することは困難
- コスト関数の上限を定める補助関数を設計し、反復型の更新式を導出

補助関数 (EUC-distance)

$$\mathcal{J}_2 \leq \mathcal{J}_2^+ = \sum_{\omega, t} i_{\omega, t} \left(y_{\omega, t}^2 + \sum_k \frac{f_{\omega, k}^2 g_{k, t}^2}{\alpha_{k, \omega, t}} + \sum_l \frac{h_{\omega, l}^2 u_{l, t}^2}{\beta_{l, \omega, t}} + 2w_{\omega, t} \right) + \mu \sum_{k, l, \omega} \frac{f_{\omega, k}^2 h_{\omega, l}^2}{\gamma_{\omega, k, l}} + \lambda \sum_{k, l, \omega} \frac{f_{\omega, k}^2 g_{k, t}^2}{\delta_{k, \omega, t}}$$
$$w_{\omega, t} = (\sum_k f_{\omega, k} g_{k, t}) (\sum_l h_{\omega, l} u_{l, t}) - y_{\omega, t} (\sum_k f_{\omega, k} g_{k, t} + \sum_l h_{\omega, l} u_{l, t})$$

$\alpha_{k, \omega, t}, \beta_{l, \omega, t}, \gamma_{\omega, k, l}, \delta_{k, \omega, t}$ は補助変数

- 上記の補助関数をそれぞれの変数で偏微分し 0 とおいた結果に対して、全ての補助変数の等号成立条件を代入することで、各変数の更新式が以下のように得られる

更新式 (EUC-distance)

$$g_{k, t}^{[p+1]} \leftarrow \frac{\sum_{\omega} i_{\omega, t} y_{\omega, t} f_{\omega, k}}{\sum_{\omega} i_{\omega, t} f_{\omega, k} (\sum_{k'} f_{\omega, k'} g_{k', t} + \sum_{l'} h_{\omega, l'} u_{l', t}) + 2\lambda \sum_{\omega} i_{\omega, t} f_{\omega, k} \sum_{k'} f_{\omega, k'} g_{k', t}}$$
$$h_{\omega, l}^{[p+1]} \leftarrow \frac{\sum_t i_{\omega, t} y_{\omega, t} u_{l, t}}{\sum_t i_{\omega, t} u_{l, t} (\sum_{k'} f_{\omega, k'} g_{k', t} + \sum_{l'} h_{\omega, l'} u_{l', t}) + 2\mu \sum_{k} f_{\omega, k} \sum_{k'} f_{\omega, k'} g_{k', t}}$$
$$u_{l, t}^{[p+1]} \leftarrow \frac{\sum_{\omega} i_{\omega, t} y_{\omega, t} h_{\omega, l}}{\sum_{\omega} i_{\omega, t} h_{\omega, l} (\sum_{k'} f_{\omega, k'} g_{k', t} + \sum_{l'} h_{\omega, l'} u_{l', t})}$$

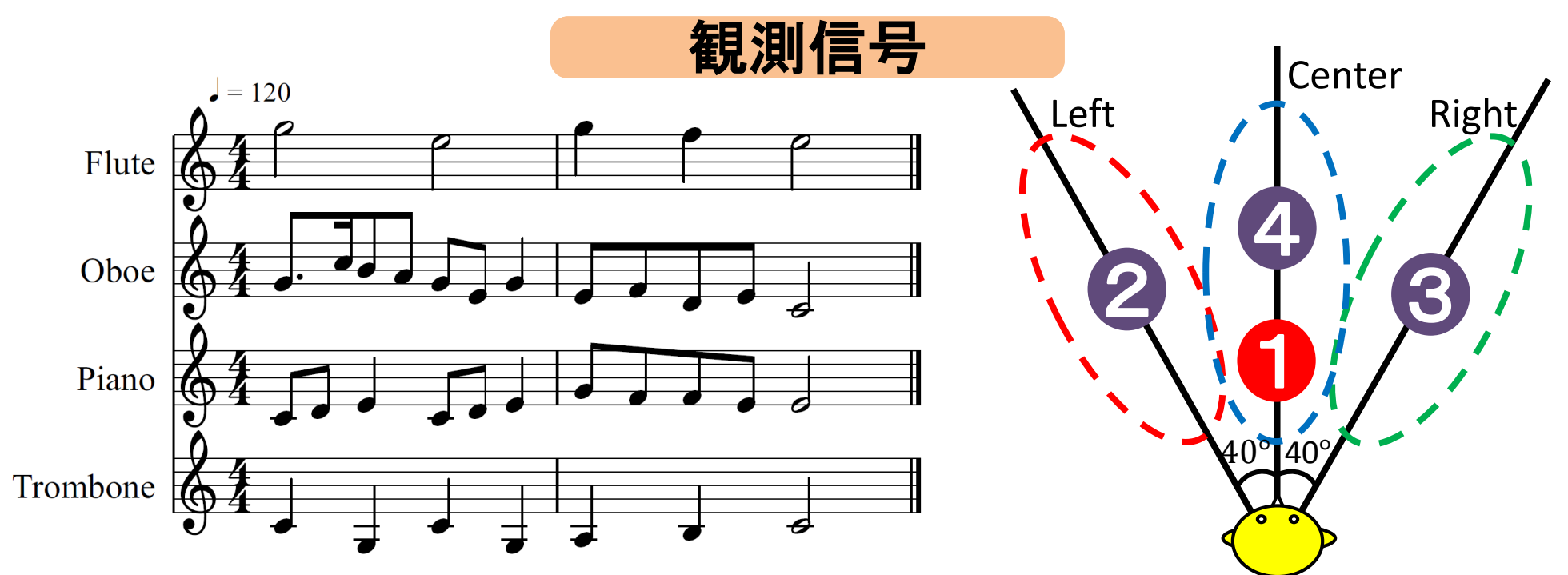
$f_{\omega, k}$ は事前学習により得られるため固定

6. 実験

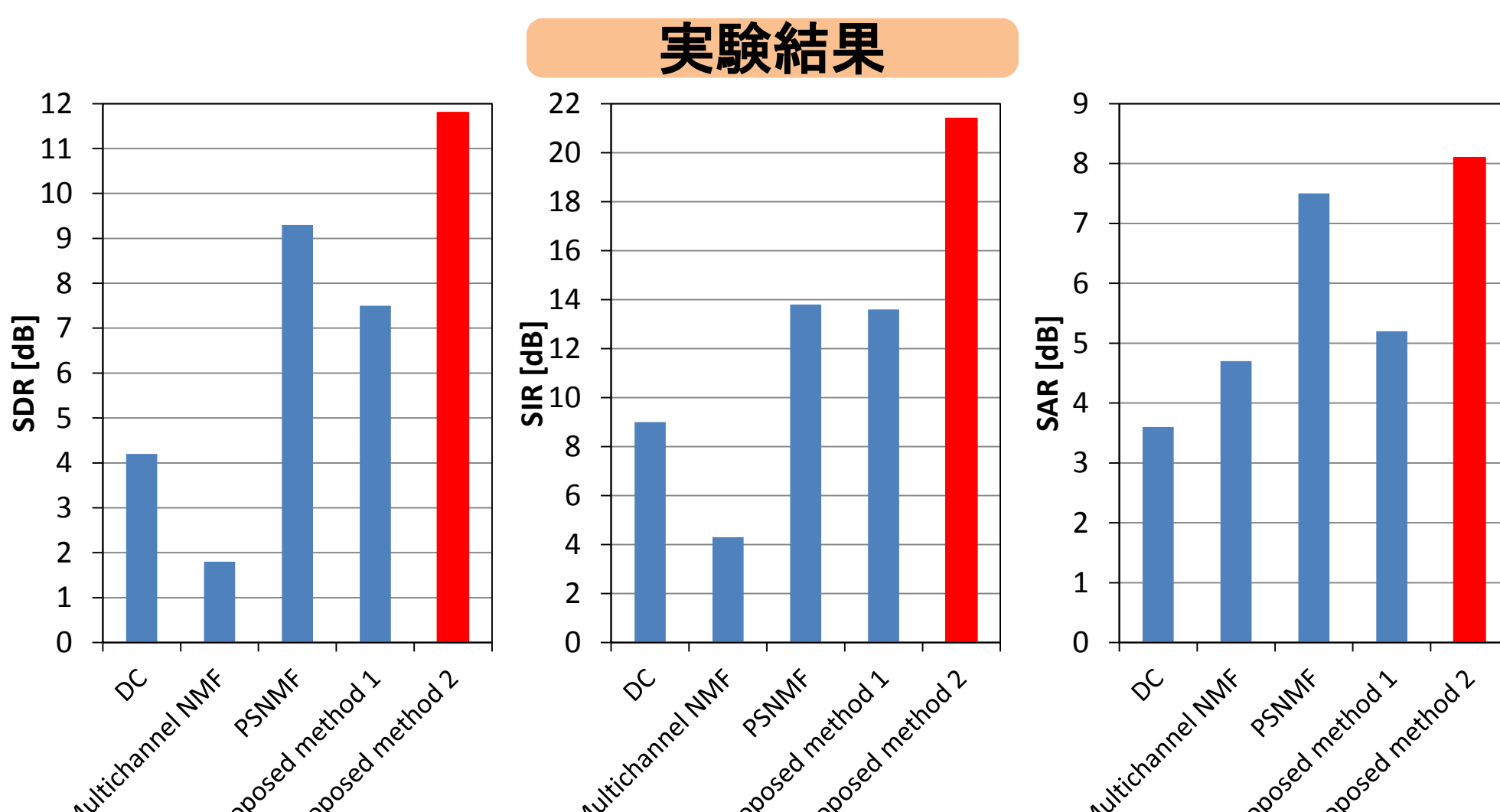
- MIDI で作成したステレオ音源から目的楽器音を抽出
- マルチチャンネル NMF, PSNMF, 方位クラスタリング, 提案手法 (正則化無し: Proposed method 1), 提案手法 (正則化あり: Proposed method 2) を用いた場合の分離精度を比較する

Experimental condition

目的音信号 (MIDI)	フルート, オーボエ, ピアノ, トロンボーン
観測信号 (MIDI)	4種類の楽器の等パワー混合信号
教師信号 (MIDI)	各楽器音の半音階で2オクターブ上昇する24音
空間代表ベクトル数	3
基底数	教師基底: 100, その他の基底: 30
SNMF の更新回数	教師基底学習時: 500, 分離時: 400
重み係数 μ, λ	評価値が最も高くなる値を実験的に求める
客観評価指標 [5]	Signal to distortion ratio (SDR: 目的音の品質), Source to interference ratio (SIR: 分離度合), Sources to artifact ratio (SAR: 処理歪みの少なさ)



- 各楽器音でメロディは固定
- 分離目的音は常に中央に配置, 同方位に干渉音源が存在
- 左右 40° 方位にそれぞれ別の干渉音源が存在
- 1種の目的楽器音に対し 3パターンの配置を用意, 計12パターンの音源に対して分離した結果の平均を比較



考察

- 方位クラスタリングのみや教師無しのマルチチャンネル NMF では、目的音源の分離が極めて困難
- 方位クラスタリングと超解像に基づく SNMF を用いた提案手法において、正則化無しの手法 (Proposed method 1) では外挿誤りにより著しくスコアが低下した
- 正則化を用いた提案手法 (Proposed method 2) では、従来の PSNMF を大きく上回る精度で分離できることが示された

参考文献

- D. D. Lee, et al., "Algorithms for nonnegative matrix factorization," *Neural Inf. Process. Syst.*, vol.13, pp. 556-562, 2001.
- A. Ozerov, et al., "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE trans. Audio, Speech and Language Processing*, vol.18, no.3, pp.550-563, 2010.
- S. Miyabe, H. Saruwatari, et al., "Temporal quantization of spatial information using directional clustering for multichannel audio coding," *Proc. WASPAA*, pp.261-264, 2009.
- K. Yagi, H. Saruwatari, et al., "Music signal separation by orthogonality and maximum-distance constrained nonnegative matrix factorization with target signal information," *Proc. AES*, 2012.
- E. Vincent, et al., "The 2008 signal separation evaluation campaign: a community-based approach to large-scale evaluation," *Proc. ICA*, pp.734-741, 2009.