

本発表の概要

- 空間上の特定の領域中にある目的音源を抽出するスポットフォーミングでは、分散配置したマイクロホンアレイと非負値行列因子分解を用いた手法が提案されている[Kagimoto+, 2022]
- さらなる性能向上および最適化への正則化の導入を目指して分散配置したマイクロホンアレイ非負値テンソル因子分解を用いた手法を提案した
- 2条件での実験より、提案手法はより高い分離性能を得られることを確認した

1. 研究背景

- ビームフォーミング (beamforming: BF)**
 - マイクロホンアレイを用いて所望の方向にある音源を強調する技術
 - スポットフォーミング**
 - マイクロホンアレイを用いて所望の位置にある音源を強調する技術
-
- 干渉音源1 干渉音源0
目的音源
BF 0 = 目的音源 + 干渉音源0
BF 1 = 目的音源 + 干渉音源1
- 分散してマイクロホンアレイを配置し適切な方向にBFを行うことで、すべてのBF出力に共通の目的音源と固有の干渉音源が現れる
 - 共通成分を抽出することで目的音源の抽出が可能
- 複数のBFと非負値行列因子分解 (nonnegative matrix factorization: NMF) を用いた従来法が知られている
 - 本研究では、非負値テンソル因子分解 (nonnegative tensor factorization: NTF) を用いた手法を提案

2. 従来手法

- 複数BFの振幅スペクトログラムを時間方向に結合、得られた行列にNMFを適用

- 係数行列を用いて時間基底バイナリマスクを構成
 - 全BFで係数行列の要素が μ よりも大きい成分を抽出

BF0を表す係数行列 & BF1を表す係数行列
時間基底バイナリマスク
基底行列

- 従来手法の弱点**
 - 異なるBF出力の時間フレームを1次元に結合しているため、分解モデルの解釈性が低くBF毎や音源毎の正則化等が導入しづらい
 - 時間フレームを超えるような時間ズレに対しては性能が低下する

フレームを超える時間ズレによって異なるバイナリマスクが構成される

3. 提案手法

- 複数BFの振幅スペクトログラムから3階テンソルを構成、得られた3階テンソルにNTFを適用

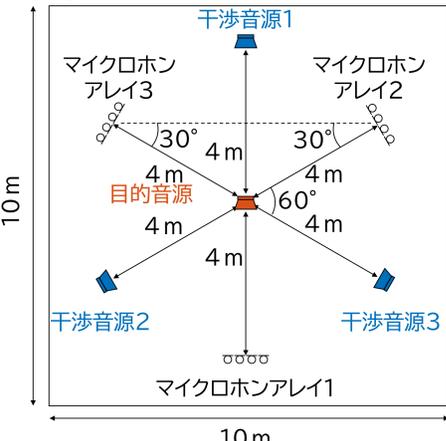
- 分配行列を用いて基底バイナリマスクを構成
 - 分配行列より、共通成分に近い γ 本の基底を抽出
 - 共通成分は全BFに含まれているため、分配行列の列に対して「要素がすべて等しいもの」が理想的な共通成分といえる

基底バイナリマスク
共通成分ベクトル
非共通成分 (BF0固有成分) & 非共通成分 (BF1固有成分)
共通成分に近いものをとっていく

- 基底ごとに共通成分と非共通成分が別れているため正則化を導入しやすい**
 - 例: 分配行列に正則化を導入し、共通成分ベクトルまたは各BFの固有成分ベクトルになるように誘導
- 基底に対してバイナリマスクを適用するので、時間ズレやデータの違いに強い
- 設定値 γ が整数であるため調整が容易

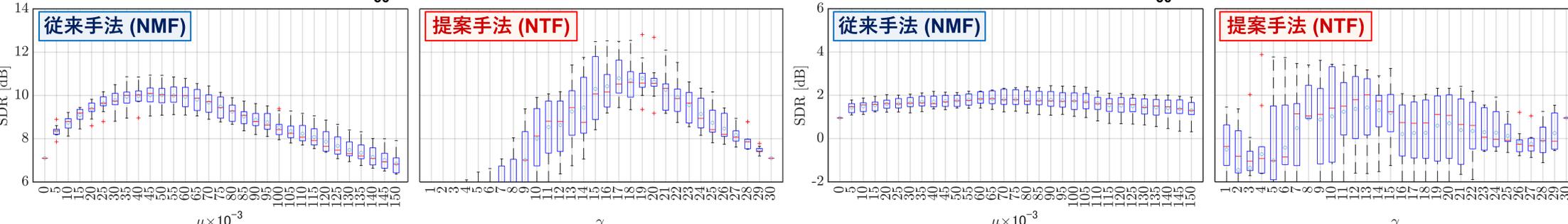
4. 実験

Pyroomacousticsでのシミュレーション



項目	条件
音源	LibriTTSから4話者
残響時間 T_{60}	0/256 ms
標準化周波数	16 kHz
BF手法	MVDR BF
STFT窓長	32 ms
STFTシフト長	16 ms
基底数	30
分配行列初期値	0.5
基底行列係数	乱数
行列初期値	10種類
性能指標	信号対歪み比(SDR)

実験結果



- 提案手法は従来手法よりも高い性能を示す傾向がある
- 提案手法は初期値によるばらつきが大きいことが見られる

- 今後の課題および展望
 - 最適な γ を自動で設定
 - 分配行列の正則化をNTFに導入