

本発表の概要

背景

- 音楽演奏の録音時は各音源に近接マイクロホン配置すると、録音対象の音源以外の音(被り音)が混入
- 時間チャンネル非負値行列因子分解(従来TCNMF)は被り音抑圧に利用可
- 我々は被り音のゲインを正則化する手法(提案TCNMF)を提案済

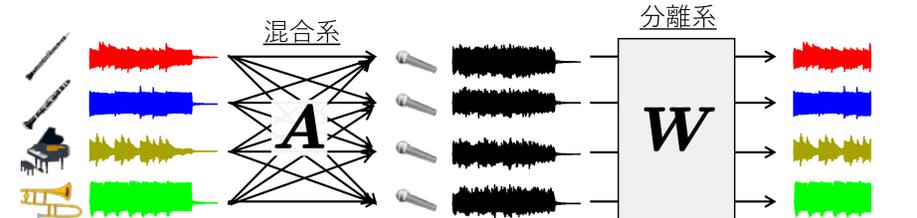
実験

- 提案TCNMFはその定式化から従来TCNMFより最適化が頑健と予想
- 実験の結果、提案TCNMFの被り音抑圧性能は従来TCNMFと比較して、最適化変数の初期値の影響をほぼ受けないことを確認

1. 研究背景

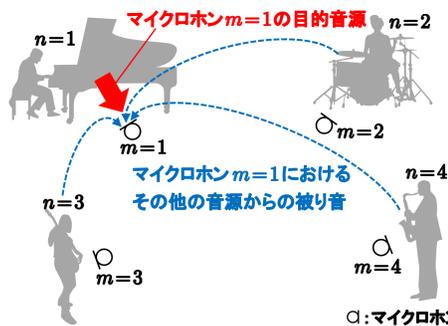
ブラインド音源分離 (blind source separation: BSS)

- 音源やマイクロホンの空間的配置(混合系 A) が不明な状態で分離系 W を推定する技術



被り音 (bleeding sound)

- 録音時、目的の音のみを録音するために楽器等にマイクを近接
- 実際には目的音以外の音が混入
 - ミキシングの質や演奏の質低下の原因

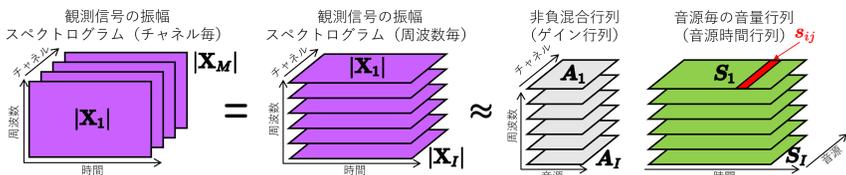


本研究では、音楽信号における被り音の抑圧を取り扱う

2. 従来TCNMF

時間チャンネル非負値行列因子分解 [Togami+, 2010] (time-channel nonnegative matrix factorization: TCNMF) (従来TCNMF)

- 観測信号の振幅スペクトログラムの周波数毎の時間チャンネル行列 (X_1, X_2, \dots, X_M) のそれぞれにNMFを適用
- 周波数毎のゲイン行列 A_i と音源時間行列 S_i を推定
- 音源時間行列 S_i の時間軸方向ベクトルにスパース正則化を導入



目的関数

$$\text{Minimize}_{A, S} \sum_i D_{KL}(X_i | A_i S_i) + \mu \sum_{i,j} \|s_{ij}\|_{0.5} \quad \text{s.t. } a_{imn}, s_{inj} \geq 0 \quad \forall i, m, n, j$$

$$D_{KL}(X_i | A_i S_i) = \sum_{m,j} \left(x_{imj} \log \frac{x_{imj}}{\sum_n a_{imn} s_{inj}} - x_{imj} + \sum_n a_{imn} s_{inj} \right)$$

- μ : 正則化重み係数, s_{ij} : S_i の時間軸方向ベクトル, $\|\cdot\|_{0.5}$: $L_{0.5}$ ノルム
- 最適化には補助関数法 [Hunter+, 2004] を用いる

3. 提案TCNMF

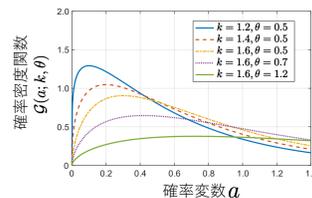
最大事後確率 (maximum a posteriori: MAP) 推定に基づくTCNMF (提案TCNMF)

- 従来TCNMFと同様に、 X_i にNMFを適用
- ゲイン行列 A_i に事前分布を導入

Diracのデルタ分布 (対角要素を1に)

$$a_{imn} \sim \begin{cases} \delta(a_{imn} - 1) & (m = n) \text{ 対角要素} \\ \mathcal{G}(a_{imn}; k, \theta) & (m \neq n) \text{ 非対角要素} \end{cases}$$

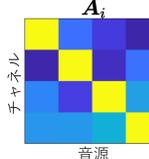
ガンマ分布: $\mathcal{G}(a; k, \theta) = \frac{1}{\Gamma(k)\theta^k} a^{k-1} e^{-a/\theta}$



目的関数

$$\text{Minimize}_{A, S} \sum_i D_{KL}(X_i | A_i S_i) + \sum_{i,m,n \neq m} \left[-(k-1) \log a_{imn} + \frac{1}{\theta} a_{imn} \right]$$

s.t. $a_{imn}, s_{inj} \geq 0 \quad \forall i, m, n, j$ and $a_{imn} = 1 \quad \forall m = n$



- 最適化には補助関数法 [Hunter+, 2004] を用いる

3. 実験

動機

- TCNMFのコスト関数には一般化Kullback-Leibler (KL) ダイバージェンスが用いられている

- KLダイバージェンスに基づくNMF (KLNMF) は、 A_i 及び S_i のいずれか片方を定数とみなすと凸最適化となり、唯一解が存在するという良い性質がある
- 従来TCNMFは、 $L_{0.5}$ ノルムの正則化項を追加したことでこの性質を失った
- 提案TCNMFは、KLNMFの生成モデル(ポアソン分布)の共役事前分布(ガンマ分布)を仮定しているため、前述の性質が引き継がれている
 - そして、正則化項で A_i の対角成分を1に固定していることから、初期値頑健性が期待

実験条件

ドライソース

- DSD100 [Rafii+, 2015] から20曲を選出
 - 開発データ10曲, テストデータ10曲
 - vocals, bass, drums, otherの4音源

評価尺度

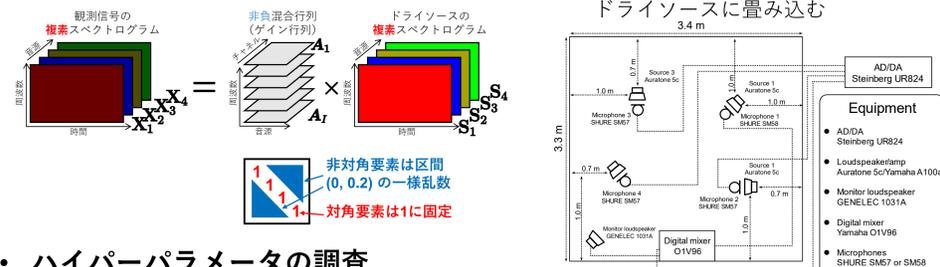
- SDR [Vincent+, 2006] の改善量

混合モデル

- シミュレーション信号の実験のモデル
 - ドライソースに非負混合行列を乗じる

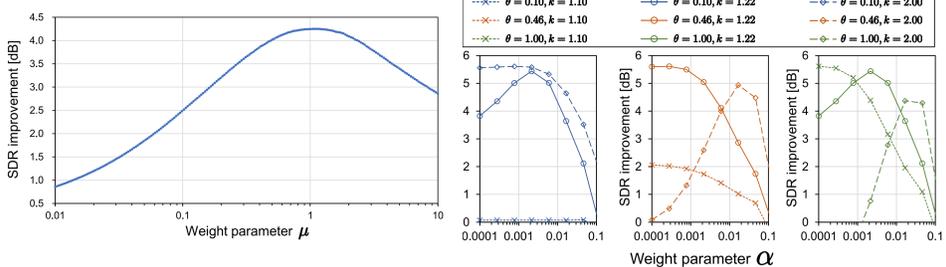
その他の条件

サンプリング周波数	44100 Hz
短時間 Fourier 変換の窓関数	Blackman 窓
短時間 Fourier 変換の窓長	92.9 ms
短時間 Fourier 変換のシフト長	46.5 ms
最適化アルゴリズムの反復回数	200 回



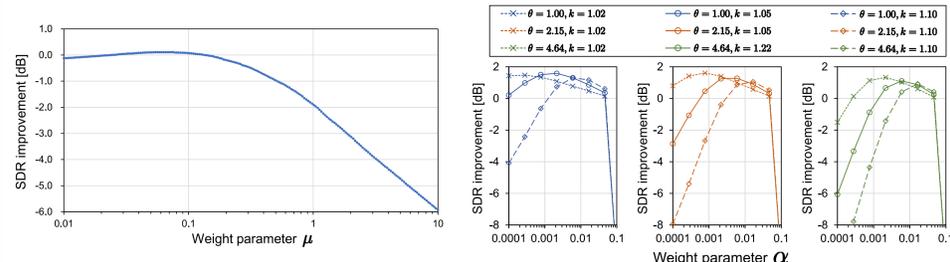
ハイパーパラメータの調査

- 開発データを用いて各手法の最適なパラメータを調査
- シミュレーション音源の被り音抑圧実験
 - 各ハイパーパラメータにおける被り音抑圧性能の変化
 - 従来TCNMF (左), 提案TCNMF (右)



実音源の被り音抑圧実験

- 各ハイパーパラメータにおける被り音抑圧性能の変化
 - 従来TCNMF (左), 提案TCNMF (右)

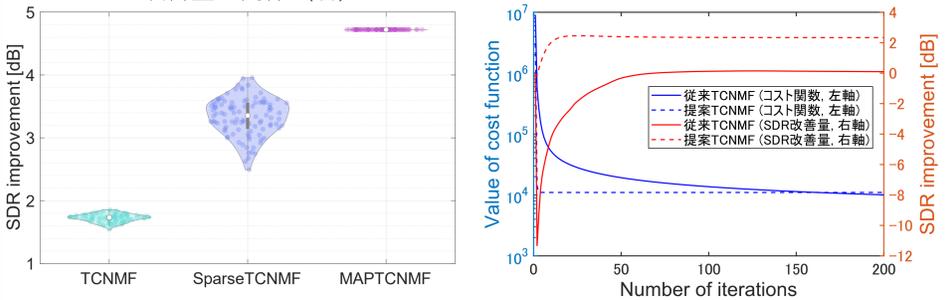


被り音抑圧の性能及び初期値頑健性の比較

- 最適なハイパーパラメータでテストデータにおける被り音抑圧性能を調査
- シミュレーション信号の実験
 - 100種類の初期値に対する曲毎の音源平均SDR改善量の平均と標準偏差 (SD)

Music no.	4	5	19	20	34	70	71	77	79	99	
Method	Mean	1.37	1.31	1.31	1.49	1.36	1.72	1.81	2.01	1.43	1.34
Simple	SD	0.02	0.02	0.07	0.03	0.11	0.06	0.05	0.23	0.05	0.06
Conventional	Mean	3.49	4.53	6.88	4.35	4.23	3.33	5.74	6.63	5.72	4.99
TCNMF	SD	0.09	0.20	0.39	0.14	0.17	0.29	0.28	0.44	0.27	0.24
Proposed	Mean	3.80	5.60	8.17	4.64	5.12	4.72	5.88	9.30	6.51	6.02
TCNMF	SD	2.05 x 10^-8	1.08 x 10^-7	1.14 x 10^-7	5.49 x 10^-8	5.22 x 10^-7	1.18 x 10^-7	1.02 x 10^-7	1.16 x 10^-7	1.17 x 10^-7	3.77 x 10^-7

- 曲番号70のデータにおけるバイオリン図 (左), 反復回数とコスト関数値及びSDR改善量の関係 (右)

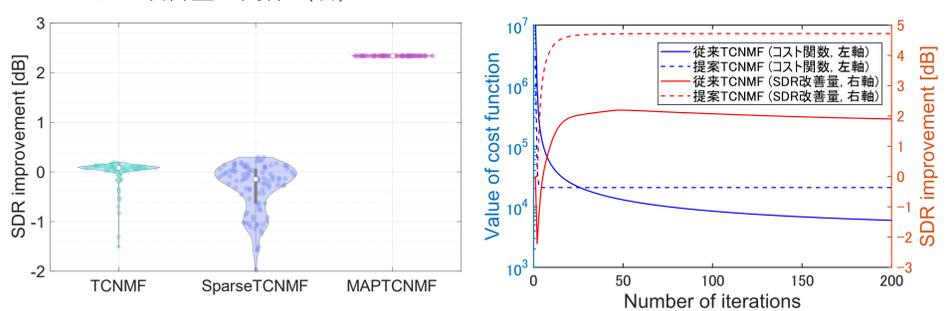


実収録混合信号の実験

- 100種類の初期値に対する曲毎の音源平均SDR改善量の平均と標準偏差 (SD)

Music no.	4	5	19	20	34	70	71	77	79	99	
Method	Mean	0.15	-0.58	-2.46	-0.29	-3.03	0.004	-0.30	-4.34	-2.16	-5.29
Simple	SD	0.08	0.50	0.97	0.54	0.73	0.26	0.82	1.22	0.65	0.66
Conventional	Mean	-0.15	-0.95	-1.93	-0.18	-3.19	-0.32	-0.19	-5.02	-2.52	-4.15
TCNMF	SD	0.23	0.51	1.16	0.61	0.78	0.49	0.81	1.36	0.81	0.56
Proposed	Mean	1.23	2.33	1.61	2.25	0.52	2.34	2.98	1.97	1.69	0.72
TCNMF	SD	5.33 x 10^-7	1.83 x 10^-6	1.42 x 10^-4	7.66 x 10^-6	6.82 x 10^-6	3.57 x 10^-5	8.71 x 10^-7	1.09 x 10^-3	1.22 x 10^-6	2.94 x 10^-3

- 曲番号70のデータにおけるバイオリン図 (左), 反復回数とコスト関数値及びSDR改善量の関係 (右)



両実験において提案TCNMFの被り音抑圧性能は従来TCNMFと比較して、最適化変数の初期値の影響をほぼ受けないことが確認できる。KLNMFの良い性質と、提案TCNMFの推定パラメータの少なさに起因していると考えられる。SDR改善量の収束はいずれのTCNMFも同程度であった。