

音源分離を用いた所望信号と外乱信号の分離技術

藤巻葵, 大島風雅, 中山仁史(広島市大院)

1. はじめに

独立低ランク行列分析(ILRMA)[1]

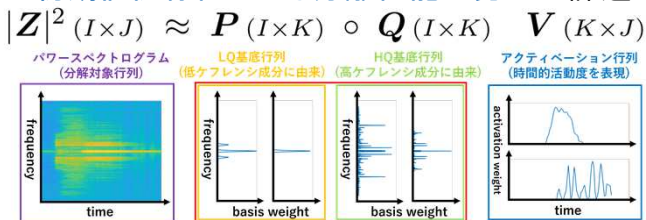
周波数領域独立成分分析(FDICA)[2]に非負値行列因子分解(NMF)[3]によるモデリングを取り入れた、ブラインド信号源分離(BSS)手法

- 楽器信号のような定常的な周波数構造を持つ信号の分離で安定かつ高性能
- 音声信号のようなスペクトルパタンが非定常的な信号の分離で初期値依存性による分離性能の劣化が課題

cepstral-basis-decomposed ILRMA(CBD-ILRMA)[4]

NMFの代わりにCBD-NMFを導入したBSS手法

- 音源モデルにケプストラム基底分解を導入
- 音声分離において安定かつ高性能
- ILRMA同様、外れ値を含む信号の分離に対して性能が劣化



t-ILRMA[5]

ILRMAの音源モデルの分布を複素Studentのt分布へ拡張

- 外れ値に対して特に頑健な分離
- t分布の自由度νにより、ロバスト性を調整可能

→ CBDによる初期値依存性の低減と、t分布による外れ値に対するロバスト性の両立を目指す

2. 提案手法

CBD-ILRMAに複素Studentのt分布を導入したCBDt-ILRMAを提案

- ✓ 自由度νを調整することでロバスト性を制御
- ✓ 通常のNMF基底にケプストラム分析とリフタリングを適用
- 低ケフレンシ成分のみを抽出して得られる低ケフレンシ基底 $p_{ik,n}$ のみを用いたモデリング

- ✓ 音声スペクトルを包絡成分と微細構造成分に分解
- ✓ 包絡成分は低ケフレンシ成分に対応

音源モデル: $\sigma_{ij,n}^2 = \sum_k p_{ik,n} v_{kj,n}$ (LQ model)

$$\text{目的関数: } \mathcal{L}_{CBDt} = -2J \sum_i \log |\det W_i| + \sum_{i,j,n} \left[\left(1 + \frac{\nu}{2}\right) \log \left(1 + \frac{2}{\nu} \frac{|y_{ij,n}|^2}{\sigma_{ij,n}^2}\right) \right] + 2 \log \sigma_{ij,n}$$

n: 音源番号 i: 周波数ビン
j: 時間ビン W: 分離行列
σ: 音源モデル
ν: Studentのt分布の自由度パラメータ

3. 実験

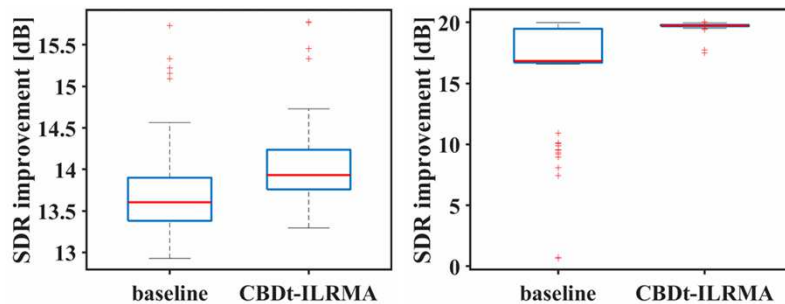
実験条件

SiSEC2011[6]より女性-女性及び男性-男性の混合音声を用いて分離する決定可能条件を想定

分離性能の評価指標には総合的な音源分離品質を示すSDR(Source-to-Distortion Ratio)の改善量を使用

Parameter	Value
Sampling frequency	16,000 Hz
FFT length	4,096 samples
Overlap length	1,024 samples
Number of iterations	100 times
Initial value of spacial model	$N \times N$ identity matrix
Initial value of source model	Uniform random value (0, 1)
Source model for CBD	LQ model
Degree of freedom	1,000

実験結果



	baseline (CBD-ILRMA)	CBDt-ILRMA
female-female	13.70 ± 0.51	14.02 ± 0.45
male-male	16.78 ± 3.82	19.70 ± 0.31

- ✓ 最大で約+3dBの改善を確認
- ✓ 外れ値の数や標準偏差も減少

音声信号に対して、より頑健な分離手法が実現

4. おわりに

まとめ

CBD-ILRMAに複素Studentのt分布を導入したCBDt-ILRMAを提案

- CBDによる初期値依存性の軽減と、t分布による外れ値成分へのロバスト性の両立を実現
- CBDt-ILRMA分離性能および分離性能が極端に低下する分離失敗を回避する傾向を確認

今後の展望

- 自由度νを固定としない、動的νを用いたCBDt-ILRMAの検討
- 話者や残響条件に依存しにくいより頑健な音声分離を目指す

参考文献

- [1] D. Kitamura et al., "Determined blind source separation with independent low-rank matrix analysis," Audio Source Separation, pp. 125–155, Springer, 2018.
- [2] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," Neurocomputing, vol. 22, pp. 21–34, 1998.
- [3] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, Vol. 401, No. 6755, pp. 788–791, 1999.
- [4] F. Oshima and M. Nakayama, "Speech source separation avoiding initial value dependency by cepstral-basis-decomposed nonnegative matrix factorization," Acoust. Sci. & Tech., vol. 44, no. 2, pp. 137–140, 2023.
- [5] S. Mogami et al., "Independent low-rank matrix analysis based on complex Student's t-distribution for blind audio source separation," in Proc. IEEE International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1–6, 2017.
- [6] S. Araki et al., "The 2010 signal separation evaluation campaign (SiSEC2010): Audio source separation," Proc. ICA/ICA, pp. 114–122, 2010.