

音源分離を用いた合唱音声のパート抽出

重定 美咲[†] 川波 弘道[†]
[†] 津山工業高等専門学校

1. はじめに

複数の音が混在した信号から個々の音を分離・抽出する技術を音源分離という。音源分離技術を用いて複数のパートで歌われた歌唱音声から特定のパートを抽出できれば、全体合唱をした際の自身の発声を抽出し練習に生かしたり楽譜落としなどで有効である。本研究では、多チャンネル収録の合唱音声を音高の異なるグループに分離する方法を検討する。

2. 提案手法

音源分離は音源の方向推定の問題と深く関係している。古典的には音源のある方向によって複数のマイクへの到達時間差が異なることを利用した遅延和ビームフォーマがある。また、特定パートのみの音源を教師とした非負値行列因子分解も使用して精度を高めるシステムを作成する。本研究では CD 音源 [1]を使用して仮想空間にて混合音を作成し、それに対して音源分離を実行する。本実験でのプログラムの概要を図 1 に示す。

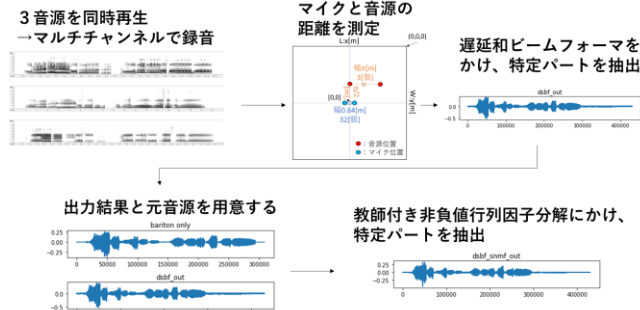


図 1. 音声分離プログラムの流れ

3. 遅延和ビームフォーマ[2]

遅延和ビームフォーマ (Delay-and-Sum Beam Forming, 以下 DSBF) ではマイクロホン x_{ik} と目的音のステアリングベクトル a_k との内積を取った信号を計算する。ステアリングベクトルとは、音源からマイクロホンまで音が伝わる時の、減衰・遅延を表すベクトルのことである。

$$\hat{s}_{ik} = a_k^H x_{ik} \quad (\text{式 } 1)$$

4. 教師付き非負値行列因子分解

非負値行列因子分解は、非負の行列を別の2つの非負行列の行列積に分解する数値アルゴリズムである。楽譜データを用いた音源を教師とする際は教師付き非負値行列因子分解 (Supervised Non-negative Matrix

Factorization, 以下 SNMF) という[3]。音声 Y を基底行列 H とアクティベーション行列 U の内積で近似できるものとし、以下に示す式 2 でユークリッド距離を計算する。

$$\frac{1}{2(Y^2 + (H \cdot U)^2 - 2Y \cdot (H \cdot U))} \quad (\text{式 } 2)$$

5. 実験結果と考察

実験では、DSBF のみで音源分離した場合と SNMF も併せて音源分離した場合それぞれの SNR を比較した。図 2 はバリトン音声を目的音源とした場合のバリトンとソプラノとアルトの混合音声と DSBF で分離後の音源、SNMF も実行したときの音源、それぞれのスペクトログラムである。

また、各音源に注目した際の SNR 向上率はバリトンは -9.5、ソプラノは +4.63、アルトは +4.81 であった。

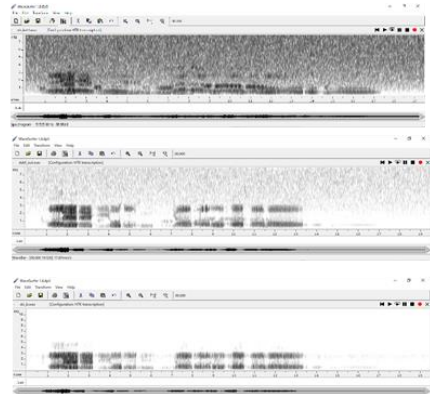


図 2. 各音源のスペクトログラム

(上: 混合音声, 中: DSBF 実行後, 下: DSBF+SNMF 実行後)

6. まとめ

DSBF と SNMF を用いた合唱音声の音源分離パート抽出実験を行った。今回は簡易化のために1パート1人で合唱したが、今後は、1パート複数人でのパート抽出、さらにパート内の個人の歌唱を抽出する方法を検討する。

参考文献

- [1] 中山一郎, “日本語を歌・唄・謡う”, 日本音響学会誌, no.59, vol.11, pp. 688-693 (2002)
- [2] 戸上真人, Python で学ぶ音源分離, 株式会社インプレス, pp.160-183(2020).
- [3] 半教師あり NMF による音源分離の実装実験, <https://keik.github.io/battlefield-acoustics/bss-with-ssnmf.html> (2022-01-28 最終参照)