平成30年度　情報工学コース卒業研究報告要旨

戸田 研究室	氏　　名	大 竹　徹 郎
卒業研究題目	楽曲音源分離における 各種音源抽出ネットワークの統合法	

市販の CD に収録されている楽曲や，インターネット配信楽曲は，ボーカル，ドラム，ベースといった様々な音源信号の混合音として表現される．楽曲音源分離とは，多様な楽器音が混合した楽曲から特定の楽器音を推定し取り出す技術であり，自動採譜や歌手推定，歌詞書き起こしなど，様々な音楽情報処理技術の前処理としての利用が期待される．

楽曲音源分離に対する有効な手法の一つとして，深層学習を用いた U-net convolutional network(U-net) に基づく手法が提案されている．U-net に基づく楽曲音源分離では，楽曲信号の短時間周波数成分の時系列を表すスペクトログラムに対して，目的とする音源成分のみを抽出し，他の成分を抑制するフィルタ (時間周波数マスク) をニューラルネットワークによって推定することで高い分離性能が達成される．楽曲中の各音源に対する U-net を用意し，各々独立に適用することで，個々の音源への分離が実現される．一方で，時間周波数マスクが各音源ごとに独立に推定されるため，必ずしも適切な分離音が得られるとは限らない．

本研究では，楽曲内の個々の音源に対する U-net を同時に適用し，個々のネットワークの時間周波数マスクを統合する手法を提案する (図１)．提案法では，各分離音を再度重畳すると元の混合音が再構成される制約 (sum-to-one 制約) を導入し，個々の U-net で推定される複数の時間周波数マスクに対して，１）正規化に基づく手法，２）目的とする時間周波数マスクへ新たに写像するネットワークを事前に学習し利用する手法により，個々の音源に対する時間周波数マスクが推定される．１）では，各時間周波数マスクを個々の音源に対するマスクの総和で除算することで正規化を行う．２）では，出力層に各ユニットの出力の総和が１となる softmax 関数を用いた小規模な統合ネットワークを用意し，少量の楽曲データを用いて事前に学習する．

提案法の有効性を調査するために，実環境で収録されたボーカル，ドラム，ベースなどの複数音源を含む楽曲データベースを用いて実験的評価を行った．個々の U-net を独立に適用する従来法と比較した結果，提案法による分離性能の改善が確認された (図２)．また，正規化に基づく提案法１と比べ，学習データを活用する提案法２は分離性能が低いことが確認された．以上の結果から，sum-to-one 制約による正規化が有効であることと，学習データを利用する提案法２に課題があることが示された．

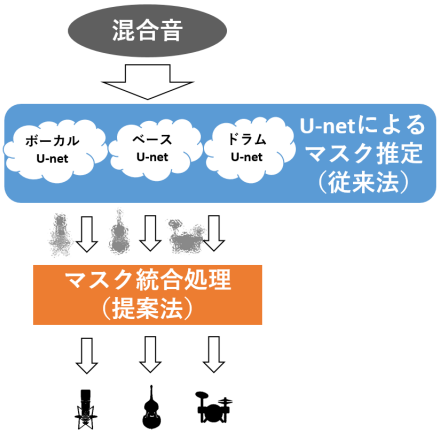

図１提案法の概要

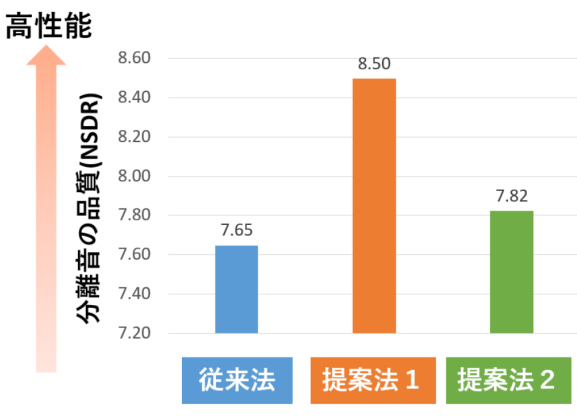

図２分離性能の比較