

報告番号	※甲	第	号
------	----	---	---

## 主 論 文 の 要 旨

論文題目 能動的楽曲鑑賞のための楽曲音源分離と  
音源配置・楽曲印象操作系に関する研究  
氏 名 大谷健登

## 論 文 内 容 の 要 旨

インターネットの発達やスマートフォンなどといった携帯型デバイスの普及により、人々は楽曲のデータをダウンロードするだけで、好きなときに好きな場所で音楽を楽しむことができるようになった。一般的な音楽プレーヤでは、ユーザは基本的には再生・停止の操作を行うだけで、手軽に楽曲の受聴を行うことができるが、ユーザ自身が楽曲を個人の好みに合わせて操作することは難しく、受動的なシステムであるといえる。周波数イコライザなどを利用することで楽曲全体の周波数特性を操作し、大まかな楽曲の印象を操作することは可能であることも多いが、ユーザの操作は楽曲全体に影響を及ぼしてしまうため、楽器音の信号を個別に操作するなどといった細かな操作の実現は困難である。一方で、ユーザ自身が楽曲の印象をインタラクティブに操作しながら行う楽曲鑑賞は能動的楽曲鑑賞と呼ばれ、ユーザに従来の受動的な楽曲鑑賞より高度な体験を提示することができる。能動的楽曲鑑賞の例としては、楽曲中の各音源の音色やリズムパターンを操作することで楽曲の音響的印象を操作するものや、各音源の配置を操作することで楽曲の空間的印象を操作するものなどが存在している。楽曲の音響的印象を操作する場合には、非常に柔軟な印象操作を行うことができる一方で、システムに不慣れなユーザでは印象操作の結果がわかりづらいという欠点も存在する。一方で、楽曲の空間的印象を操作する場合には、印象操作の範囲が音源の配置に限定されるが、配置操作の結果受ける空間印象は直感的にもわかりやすい。そこで、本研究では、システムに不慣れなユーザでも比較的扱いやすい能動的楽曲鑑賞系として、楽曲中に存在する各楽器音やボーカル信号といった音源の信号を分離・強調し、それらの信号の配置をユーザ自身が操作することで楽曲の印象を操作することのできるシステムを構築することを目的とする。特に本論文では、本研究で目的とする能動的音楽鑑賞の実現のために必要な基礎技術である、楽曲信号から楽器やボーカルといった各音源の

信号を分離・強調するための音源分離手法と、個別の音源信号をユーザが手軽に操作するためのインタラクション手法に焦点をあて、最後に本論文での成果を統合した能動的楽曲鑑賞系について触れる。

本論文で扱う音源分離手法では、畳み込み雑音除去自己符号化器 (convolutional denoising autoencoder: CDAE) と対数周波数領域振幅スペクトル特徴を利用し、楽曲信号から個々の楽器音信号を強調するための技術について述べる。楽曲信号の音源分離では、従来から非負値行列因子分解 (non-negative matrix factorization: NMF) が広く利用されてきた。しかしながら、NMF では、混合楽曲信号の振幅/パワースペクトルが音源信号の振幅/パワースペクトルの線形結合で表現されることを仮定しているが、実際にはこのような線形混合仮定は完全には成立しないため、目的音の振幅/パワースペクトルを正確に推定できないことがあった。NMF モデルを非線形性を加味した混合モデルへ拡張し、振幅/パワースペクトル推定の精度を高めるために、深層ニューラルネットワーク (deep neural network: DNN) を適用する研究も盛んに行われている。しかしながら、DNN を利用した多くの検討では DNN の構造についてはあまり考慮されていない。本論文では、音源分離性能の向上のために、DNN の構造を従来手法より楽器音に適したものに変更することを考える。そのうえで、楽器音のスペクトルをより高精度に推定するために、対数周波数領域のスペクトル特徴量を CDAE に入力する。楽器音のスペクトルモデルについては、従来から議論が行われているが、本論文では対数周波数領域での楽器音のスペクトルモデルが畳み込み表現で分解されることを示し、提案モデルとの関連についても述べる。実際の楽曲信号を利用して音源分離実験を行ったところ、他のスペクトル特徴量形式や DNN 構造と比較して高精度に音源分離を行うことができることを確認した。また、目的音と雑音間の相補性に着目し、目的音だけでなく、雑音の振幅スペクトル推定を同時に行い、それらを組み合わせたところ、音源分離精度が更に向上した。

続いて、複数の音源の配置をユーザが手軽に操作できるようになることを目的とした、複数音源配置の次元操作系について述べる。音源ごとに音量・配置を操作する場合、音源数の増加に伴って操作が煩雑化し、システムに不慣れなユーザでは操作が困難となることが想定される。従来の能動的楽曲鑑賞系においても、音源ごとの音量や配置を操作する手法は存在するが、音源数の増加による操作の煩雑化の解決などについてはあまり扱っておらず、より直感的かつ手軽に音源配置を操作するためのインタラクション手法について考慮する必要がある。他分野においては従来からも操作インタフェースについての研究は行われてきたが、事前に設定した目的対象物を効率的に探索する手法について議論されることが多かった。しかしながら、本研究で目的とする能動的楽曲鑑賞系においては、ユーザが事前に決定した目的の空間印象を探索する問題ではなく、様々な空間印象を比較しながら最適なもの

のを探索する問題であると考えられる。そのため、従来のアプローチとは異なり、操作の次元数を削減しつつ、音源配置の組合せを網羅的に探索することが可能なインタフェースが必要となる。そこで、本論文では事前に楽曲の空間的印象が互いに大きく異なると考えられる音源配置の組み合わせを抽出しておき、それらを一次元の操作軸で切り替えることができるインタラクション手法を考える。これにより、ユーザはダイヤルノブなどの直感的に操作可能な一次元の操作軸を利用するだけで、空間的印象を大きく変化させながら比較することができるようになる。客観的・主観的評価を通して、提案法により幅広い空間印象を持った音源配置の組み合わせを選択可能であることを確認した。

最後に、能動的楽曲鑑賞系の実現のために必要となる基礎技術について述べ、実際に本論文の成果と組み合わせることで能動的楽曲鑑賞のデモシステムを作成した。本論文では、ヘッドホンでの受聴を前提とし、空間印象の提示には頭部伝達関数 (head related transfer function: HRTF) を利用する。事前に複数の方向から収録された HRTF を分離後の各音源信号に畳み込むことで、各音源信号に空間印象を付与する。HRTF をリアルタイムに切り替えながら畳み込むため、畳み込み処理には重畳加算法を利用する。これらの基礎技術と本論文で議論した音源分離手法及び空間印象の一次元手法を組み合わせることで、実際に能動的楽曲鑑賞系のデモシステムを作成したところ、体験者からは多くの好意的な意見を得ることができた。また、能動的楽曲鑑賞系の発展の一例として、楽曲信号だけでなく演奏映像を同時に付与するシステムについても併せて開発した。映像情報を付与することで、あたかも実際のライブ会場にいるかのような、より臨場感のある楽曲鑑賞体験を提示することを可能にした。