氏　　名	北村 大地
学位(専攻分野)	博士(情報学)
学 位 記 番 号	総研大甲第 1931 号
学位授与の日付	平成29年3月24日
学位授与の要件	複合科学研究科 情報学専攻 学位規則第6条第1項該当
学 位 論 文 題 目	Effective Optimization Algorithms for Blind and Supervised Music Source Separation with Nonnegative Matrix Factorization
論 文 審 査 委 員	主　査　准教授 小野 順貴 教授 佐藤 いまり 教授 速水 謙 准教授 山岸 順一 教授 猿渡 洋 東京大学

(別紙様式 2)
(Separate Form 2)

論文の要旨
Summary (Abstract) of doctoral thesis contents

In this dissertation, to address a music source separation problem, several optimization algorithms are proposed. Music source separation is a technique to extract or separate specific music sources from an observed mixture signal that contains multiple music instrumental and vocal sounds. There are many feasible applications for this technique, for example, audio remixing by users, automatic music transcription, and musical instrument education. A general audio source separation problem has been investigated for a long time, particularly in the speech signal processing field to reduce background noise and enhance only the speech signal in the observation. Many techniques have been proposed for various recording conditions in the past, and they can roughly be divided into two situations: determined (or overdetermined) and underdetermined cases. In the determined situation, sufficient number of observations (microphones used in the recording) can be utilized for solving the separation problem, whereas the underdetermined situation, which includes monaural observation, basically lacks such multi-dimensional information. Also, presence of external prior information (supervision) such as music scores, source locations, or sound examples of each source in the mixture is another important issue. The source separation techniques without any prior information is often called blind source separation, which is the most difficult but a practical technique.

The objective of this dissertation is to develop an effective optimization algorithm for the music source separation and to achieve satisfactory separation performance. Two main topics are here addressed: determined (and overdetermined) blind source separation and single-channel (underdetermined) semi-supervised source separation. The semi-supervised source separation exploits sound examples of only the target source for the separation, namely, only the target source is extracted from the mixture. In both the topics, an important property of music signals is focused to effectively capture their structures. Since typical music signals consist of limited number of components such as discrete pitches and musical notes and include many reiteration of similar or the same spectral patterns (timbers), the power spectrogram of music signals tends to have a low-rank structure. On the basis of this nature in music signals, for both the topics discussed in this dissertation, a popular algorithm of matrix decomposition called nonnegative matrix factorization (NMF) is exploited for modeling the structure of music signals. By applying NMF to the spectrogram of audio signal, the frequently appearing spectral patterns and their time-varying gains

can be extracted as bases and activations. These components are useful for modeling the audio signals and achieving the source separation.
For the problem of determined blind source separation, independent component analysis (ICA) and its multivariate extension, independent vector analysis (IVA), are traditional and reliable approaches and can provide good separation results particularly for a mixture signal of speech. These approaches estimate spatial demixing filters by assuming that the sources are mutually independent. This assumption is valid in a practical mixture signal and make the separation problem solvable in a fully blind fashion. However, the separation accuracy of ICA and IVA for music signals is not satisfactory. This is because the general music signals frequently contain spectral overlaps and co-occurrences between sources, which result in a harmony of music, and these properties weaken the inherent independence between the sources. Also, the both methods assume only the non-Gaussian source distribution as an unspecific source model and do not utilize any information about the structure in the spectrogram of each source. To solve this problem, in this dissertation, the unified method of NMF and IVA called independent low-rank matrix analysis (ILRMA) is proposed, which performs simultaneous estimation of the spectrogram structure of each source and their spatial demixing filters. The optimization algorithm in ILRMA ensures faster convergence, more stable performance, and better computational efficiency compared with conventional methods including multichannel extension of NMF (MNMF), which is a state-of-the-art method for source separation. Also, theoretical relationships between IVA, MNMF, and ILRMA are revealed, namely, ILRMA is essentially equivalent to MNMF with a constraint for the mixing system, and IVA is also a special case of ILRMA.

For the single-channel semi-supervised source separation task, semi-supervised NMF, which aims to extract only the target source from the mixture, is the most popular approach. In this method, sound examples of the target source are utilized for preparing the supervised bases (spectral dictionary) of the target source. However, when the target source and the other sources in the mixture signal share similar or the same spectral patterns (bases), the separation performance of semi-supervised NMF is degraded because such shared components cannot be separated. This fact means that the supervised bases must be discriminative from the other bases of non-target sources.
On the basis of this fact, in this dissertation, a new training algorithm that provides discriminative supervised bases is proposed for semi-supervised NMF. In this method, other sound examples, which are candidates of the non-target signals in the observed

mixture, are utilized only for learning which spectral components will be frequently shared between the target and non-target sources.

Furthermore, a new efficient initialization scheme for NMF is proposed. Since an optimization in NMF requires initial values for bases and activations, all the results of applications based on NMF always depend on the initialization. The proposed initialization is based on a maximization of mutual independence between the activations using nonnegative ICA algorithm. The efficacy of the proposed method for several source separation tasks including ILRMA and semi-supervised NMF with discriminative basis training is experimentally confirmed.

(別紙様式 3)
(Separate Form 3)

博士論文審査結果の要旨

Summary of the results of the doctoral thesis screening

本博士論文は、非負値行列因子分解（Nonnegative Matrix Factorization; NMF)という手法に基づく音楽信号の音源分離のための効果的な最適化アルゴリズムについて論じたものである。

第１章ではまず、背景として音源分離という信号処理技術の意義と、本論文の主題である NMF という手法のこれまでの研究の流れ、ならびに本論文の貢献が簡潔に述べられている。

第２章では、複数の音源信号が混合された観測から元の音源信号を推定するという、音源分離問題の定義と定式化がなされ、続いて独立成分分析を用いた初期の研究からディープニューラルネットワークを用いた最新の研究まで、これまでの音源分離の先行研究を、マイク数と音源数の大小関係、事前情報の有無などの観点から整理している。次に本論文の動機として、音楽信号の音源分離の難しさと音源モデルの重要性が述べられ、楽音のスペクトログラムの低ランク性に着目するというアイディアが具体例とともに示されている。

第３章では、優決定（マイク数が音源数と同じ、もしくは多い）多チャンネル観測でのブラインド音源分離に対して、従来の独立成分分析、独立ベクトル分析における音源モデルを拡張し、楽音スペクトログラムの低ランク性を積極的に活用する、独立低ランク行列分析（Independent Low-Rank Matrix Analysis; ILRMA)という新しいブラインド音源分離手法を提案し、空間フィルタの推定と NMF 音源モデルの推定を交互に反復する効率的なアルゴリズムを導出している。次に、実際の音楽信号の音源分離に適用し、提案法のアルゴリズムの高速性と分離性能の高さを、従来法との比較実験により示している。また、スペクトログラムのランクはフレーム分析の窓長にも依存することから、窓幅を様々に変えた場合の分離性能や、音源数より多くのマイクが利用できる場合の拡張手法についても述べられている。

第４章では、シングルチャンネルでの教師あり音源分離の新しい学習アルゴリズムを提案している。まず従来法の問題点として、NMF の目的関数は混合音スペクトログラムをどれだけよく近似できるかを表すものであり、分離性能のよさを直接的に表現しているわけではないことを指摘し、これを改善するために、適当な妨害音データを前提に分離性能を評価する目的関数を定義し、これを最小化する新しい基底学習アルゴリズムを導出している。また実験により音楽信号の分離において分離性能が改善できることを示し、最適な点で反復計算をとめるための手法については今後の課題としている。

第５章では、NMF に基づく音源分離に共通する重要な問題として、NMF の反復計算における初期値依存性を取り上げ、これに対して独立成分分析を用いた２つの初期値決定法を提案している。また、どちらの方法も乱数初期値よりよい結果を与えることを実験的に示している。また、第３章、第４章で提案した手法にこの初期値決定法を適用した場合の効果について論じている。

第６章では、本論文全体がまとめられ、今後行うべき課題が列挙されている。

(別紙様式 3)
(Separate Form 3)

審査会では出願者から論文全体について発表がなされ、その後の質疑においても適切な回答がなされた。本博士論文では、音楽信号の音源分離という問題に対して、楽音のスペクトログラムを非負値行列因子分解によりモデル化するという一貫したアプローチに基づき、多チャンネルブラインド条件、ならびにシングルチャンネル教師あり条件において、新規で効果的な音源分離アルゴリズムを導出しており、当該分野に大きく貢献している。また、従来手法との比較実験により提案法の有用性を示すにとどまらず、スペクトログラムの低ランク性という着眼点の確認から分離音の品質の主観評価まで、実験的検証が非常に豊富に行われており、論述にも説得力がある。また、本論文の内容は当該分野のトップジャーナルに

Daichi Kitamura, Nobutaka Ono, Hiroshi Sawada, Hirokazu Kameoka and Hiroshi Saruwatari, "Determined Blind Source Separation Unifying Independent Vector Analysis and Nonnegative Matrix Factorization," IEEE/ACM Trans. Audio, Speech and Language Processing, vol. 24, no.9, pp. 1626-1641, Sept. 2016.

という査読付きジャーナル論文としてすでに掲載されている他、出願者が主著者である 4 編の査読付き国際会議論文がすでに掲載されている。

以上より本論文は、博士学位を与えるに十分な水準に達していると、審査委員全員一致で認められた。