

## 論文内容の要旨

博士論文題目      Statistical Approach to the Single-Channel Sound  
Source Extraction

氏 名              井原 瑞希

(論文内容の要旨)

Extracting only the essential sound attributes from sounds is one of the fundamental issues of computational auditory scene analysis. Especially, the estimation of sound source characteristics is difficult since the corresponding physical quantity is not defined. This dissertation proposes two music information retrieval methods: the instrument feature extraction assuming the timbre space and the probabilistic model of sounds considering the source-filter model and dynamics.

In the former part of this dissertation, an instrument feature extraction method with a combination of linear projection methods is developed. For monophonic music instrument identification, various feature extraction and selection methods have been proposed. Although raw power spectra have enough information for accurate instrument identification, their dimensionality is too high and redundant. It is important to find non-redundant instrument specific characteristics that maintain information essential for high-quality instrument identification to apply them to various instrumental music analyses. As such a dimensionality reduction method, a combination of linear projection methods is introduced: principal component analysis (PCA) and local Fisher discriminant analysis (LFDA). Additionally, the reason why linear projection algorithms are suitable for instrument identification is explained by the geometrical analysis of algorithms. After experimentally clarifying that raw power spectra are actually good for instrument classification, the feature dimensionality is reduced by PCA followed by LFDA. The reduced features achieved reasonably high identification performance that is comparable or higher than those by the power spectra and those obtained by existing

studies. These results suggest that the proposed PCA-LFDA can successfully extract low-dimensional instrument features that preserve the characteristic information of instruments.

In the latter part, a probabilistic model that represents our assumption with an extension of the source-filter model is introduced to estimate three elements of sounds: pitch, loudness and instrument-specific characteristics. The source-filter model, originally devised to represent a sound production process, has been widely used to estimate both of the source signal and the synthesis filter. This model suffers from an indeterminacy problem. To resolve it, three constraints are included in the model: harmonics, smoothness and sparseness. In detail, the source signal and synthesis filter contain the time-varying fundamental frequency and amplitude information and time-invariant instrument-specific information, respectively in the source-filter model. A probabilistic model that represents those assumptions with an extension of the source-filter model is constructed. For learning of model parameters, an EM-like minimization algorithm of a cost function called the free energy is introduced. Reconstruction of the spectrum with the estimated source signal and synthesis filter and instrument identification by using the model parameters of the estimated synthesis filter are performed to evaluate the proposed approach, showing that this learning scheme could achieve simultaneous estimation for the source signal and the synthesis filter.

## (論文審査結果の要旨)

音信号から音色や音高などの基本的属性を推定することは、計算機によって音像を理解するうえで重要な課題である。しかし音楽における情報抽出は、音響特性が一定でないことなどの理由で一般には困難である。本論文は、シングルチャンネル音源から情報抽出する方法を二つ提案し、それらの有効性を示したものである。

一方は、線形射影法を応用した楽器判別法である。一般に楽器判別においてはパワースペクトルを用いるが、そのままでは特徴次元が高すぎるためにノイズに対して頑健でない。そこで本論文では主成分分析 (PCA) と局所フィッシャー判別法 (LFDA) を組み合わせることで次元を削減し、より高精度の推定を可能としている。さらに、これらの手法の妥当性が幾何学的見地から議論されている。具体的には、異なる音高の同一楽器のスペクトルがどのような演算に対して不変であるかを考察し、提案法がこれらの演算の元で不変になることが示されている。

もう一方は、ソース・フィルタ・モデルに基づく、音高、音量、音色の3要素の同時推定法である。一般にソース・フィルタ・モデルは音源の生成過程をよく表していると考えられているが、設定すべきパラメータが多いため、データからそのパラメータを推定するのが困難であり、そのため推定精度も高くない。そこで本論文では、ソース・フィルタ・モデルにさらに3つの制約条件、すなわち、倍音、滑らかさ、スパース性を導入する。これにより、従来のモデルでは考慮されていなかった、音高および音量は時間変化するが音色は変化しないという性質がモデルに組み込まれている。また、機械学習におけるEMアルゴリズムと同じアイデアに基づく、自由エネルギーを用いた新たなパラメータ推定法を提案し、精度のよい推定が実現できている。

以上をまとめると、本論文は、機械学習に見られる数理モデルを音楽情報科学に導入することで、モデルに基づいた新しい音楽情報抽出法を提案し、その有効性を実証したものであり、今後の音楽情報科学に新しい道筋を示したものである。したがって、博士(工学)の学位に値するものと認められる。