

審査の結果の要旨

氏名 呉 軍

本学位請求論文は、「Musical Instrument Identification Based on Weighted Timbre Features and Probabilistic Model (音色特徴重みづけと確率モデルに基づく楽器認識の研究)」と題し、5章から成る。

第1章「Introduction (序論)」では、本論文が扱う分野の説明と応用領域などについて説明されている。従来から機械に人間のように音を認識させることに関する研究は多くなされているが、その多くは音声認識などに重点を置いていたため、音楽などのそれ以外の音の認識に関する研究はそれほど多くはなされてこなかったことを指摘し、音楽信号を認識するための基礎技術の確立が必要であるという研究動機を述べている。また、従来からソロ演奏に関する採譜などの手法がいくつか提案されている一方で、多数の楽器を含むような、より現実的な音楽には従来法では限界があると指摘し、それを解決するための技術の確立するという研究目的について述べている。この問題が解決された際には(1)多重音のピッチ解析 (2)楽器音認識などが可能となり、さらには(3)音楽の属性解析などへの応用も可能になるといった、応用面への展望も述べている。

第2章「Single note model by joint modeling of sustained and attack sounds (サステイン音とアタック音の結合による単音モデル)」では、音楽信号を個々の音ごとに分離するために、楽音のモデルのひとつであるアタック・ディケイ・サステイン・リリース (ADSR) モデルと、楽器の音色の特徴や性質の分析に基づいた生成的アプローチにより単音モデルを定式化している。この単音モデルは多重音解析にも利用でき、かつ楽器音の音色的特徴を表現するような表現力もある。このモデルは、アタック部とそれ以外のハーモニック部の結合により表現されている。アタック部は周波数領域の混合ガウスモデルと時間領域のガウス分布としてモデル化でき、ハーモニック部は、時間周波数領域上での3次元混合ガウスモデルとしてモデル化できることを論じている。またこれらのモデルのパラメータを推定するために、EM アルゴリズムによる推定方法を導出している。このモデルに基づいた手法により、合成音と実データの両方の場合について、従来のモデルよりも高い精度で実音楽を各単音に分離できることが実験により検証されている。このモデルは、多重音解析のみならず、後段の楽器識別における特徴量として利用することもできるものである。

第3章「Weighted timbre features for instrument identification (楽器音識別のための重み付き音色特徴量)」では、特徴量の「次元の呪い」に対処するための次元圧縮に関して述べている。次元圧縮には、第2章で定式化した単音モデルのモデルパラメータに対して主成分分析(PCA)を適用するという旨が述べられている。ただし、この問題においては単純な PCA は必ずしも有効ではないことを指摘し、この原因として、音楽信号では

多数の信号が同時に発せられるために、ここで抽出される特徴量の信頼性がそれほど高くはないことを挙げている。そこでさらに、音の重なりを検出するための指標を設計し、単なる PCA ではなく重み付きの主成分分析が有効であると議論している。この重みは、それぞれの特徴量の信頼性の高さに基づいて決められる。また、学習データにノイズが加わっている場合に対する対策も述べられている。この特徴量を用いた識別実験の結果、従来用いられた特徴量である MFCC やソースフィルタ特徴量の場合と比べて、合成音の場合はパーセンテージにしてそれぞれ平均で 20.6 ポイント、6.6 ポイントの改善があり、実世界の音楽データの場合は平均でそれぞれ 19.1 ポイント、5.9 ポイント、性能が向上することが示されている。

第 4 章「New boosting algorithm with probabilistic decisions (確率的決定則による新しいブースティングアルゴリズム)」では、楽器音の分類器に関する議論がなされている。具体的には、第 3 章で抽出した特徴量を効果的に用いるブースティングアルゴリズムを議論している。このブースティングアルゴリズムは、従来のブースティングのように決定論的にパラメータを決定するものではなく、特徴量ごとの信頼性に応じて確率的に行うように改良したものである。このブースティングアルゴリズムは、単一の識別器のみを用いた場合の誤りのリスクを低減し、実環境の音楽に対する頑健性を向上させるために、特に信頼性が低い特徴量のためには楽器音識別器をアンサンブルさせることができる。検証の結果、性能は、SVM や AdaBoost を識別器として利用した場合について、合成音ではそれぞれ 9.1 ポイント、4.2 ポイントの性能の改善があり、実世界の音楽データの場合はそれぞれ 6.1 ポイント、3.3 ポイント性能が改善することが示されている。

第 5 章「Conclusion (結論)」では、楽器認識について議論されている。第 2 章では音楽信号を分離するための単音モデルが定式化されている。第 3 章では単音モデルからの特徴量抽出手法と、楽器音識別のための特徴量の変換について述べられている。第 4 章では信頼度の必ずしも高くない特徴量を扱うためのロバストなブースティングアルゴリズムについて述べている。人間の聴覚では、この論文で対象とした音楽よりもさらに複雑な音楽を扱えると考えられるため、本論文の手法は、人間の聴覚には及ばないが、人間のような聴覚を実現するための一里塚となったと考えられる。

本論文の重要な貢献は、(1) 音楽信号を分離するための単音モデルの定式化、(2) 単音モデルからの特徴量の抽出 (3) 特徴量から最も本質的な情報を取り出すための後処理 (4) 楽器音識別のためのための分類器 (5) システムの頑健性のためのブースティングアルゴリズムなど、さまざまな創意工夫を含む手法により、従来難問とされ研究例が少なかった、複数種類の楽器音を含む音楽の多重音解析問題を、解決に向けて一歩前進させたことである。この研究は、実世界の音響信号を、人間が理解しているように機械に理解させるという当該分野の究極目標を目指す上でも重要な意味がある。

以上により、本論文は博士 (情報理工学) の学位請求論文として合格と認められる。