

審査の結果の要旨

氏名 ラチンスキ スタニスワブ アンジェイ

本論文は、音楽音響信号の多重音分析について論じている。多重音分析とは、自動採譜技術の中核技術であり、録音された音楽音響信号に混在する様々な楽器から生じた複数の音符の音から、其々の音符の基本周波数、オンセット時間と音価を自動的に推定することである。多重音解析技術は実用価値が高く、音源分離、自動採譜、構造的オーディオ圧縮 (structured audio coding)、ロボット聴覚のための分析手段、音楽コンテンツの自動メタデータ化とそれによる高機能な音楽検索システムなど、極めて広範囲にわたるアプリケーションへの応用が期待される。

多重音分析は、学際的な新興分野の音楽情報検索(Music Information Retrieval, MIR)の重要かつ古くからの問題である。多くの手法が提案されたが、いまだ解決に至っていない。十分な音楽訓練を受けた人間は、(オーケストラ曲は無理としても)複数の楽器で演奏された音楽を何度も聴いて楽譜にすることができる。品減の場合は、この問題をトップダウンで解決しているように考えられる。つまり、最初はジャンル、使われている楽器、拍子構造等、調、ベースライン、和音の順に理解し、その情報や音楽論的な知識を応用し、音符に変換して楽譜を作成している。似た分野である音声認識では、言語モデルなどの事前知識を活用した統計的なトップダウンアプローチが成功し、現代の音声認識技術に発展したという歴史がある。以上の二つのヒントから、本論文では多重音解析の問題を解くため、確率的なトップダウンシステムを企て実現することを目的としている。現在までの採譜システムは対照的にほとんどがボトムアップ方式で、音楽理論がほとんど利用されていなかった。

まず第2章では、入力信号から非負行列近似(Nonnegative Matrix Approximation, NNMA)に基づく特徴抽出手法が提唱されている。NNMAは過去にも多重音分析として用いられ、フレーム毎に音符が存在する可能性を定量的に扱えるが、NNMA自体は音楽に特化しない一般的な方法なので、誤りが多く、事後処理と事後解析が必要であった。本論文の筆者は、音楽信号の特質を活かすため、各音に調波制約(harmonicity)を課した調波的非負行列近似(Harmonic NNMA, HNNMA)と呼ぶ手法を考案し、多重音(多声音楽)からそれらのパラメータを学習推定する新たな学習方法を提案した。すなわち、旋律楽器の合奏など、単音が整数倍音構造を持っているような調波性を仮定できる場合、それをNNMAの学習推定アルゴリズムの反復計算に組み込んで制約付きの推定法を提案した。音楽に使われる単音は、一般に途中で急に大きくなったりしない。また、楽譜化する場合は、同じ聞こえがする最少の音符で記述するであろう。このような原則を数理的に表現して、音符強度の時間包絡の滑らかさ(continuity)や、音符ができるだけ疎である(sparsity)などの正則化を含めたNNMAを提唱した。

この手法を評価した結果、過去の手法と比較して優位であり、論文は世界で引用されている。

第3章では、ダイナミック・ベイジアン・ネットワーク(DBN)に基づく史上初の統計的な階層音楽論モデルが提唱されている。提案モデルには、拍子、リズム構造、調、和音構造など、いわば音符を時間(横軸)と和音(縦軸)の2次元に配置する2次元の言語としての文法構造が確率的にDBNにより表現されており、その確率パラメータは様々なレベルの音楽データを用いて学習抽出される。この音楽理論モデル(musicological model)を評価するため、シンボリック(音符単位)な情報処理では、自動作曲実験や和声分析の実験結果により妥当性が示された。しかし、さらに音響入力にも対応するには、現時点で入手可能な限られた学習データではモデルが大きすぎてパラメータ学習精度が得られないため、簡略化した実用的なモデルが必要である。それを作成して、周辺クロスエントロピーと条件付きクロスエントロピーを用いたモデルの音符予測力評価では、提案モデルでは、音楽理論を用いない一様モデルより大幅に性能が向上した。

第4章では、提案した特徴抽出手法と音楽理論モデルを確率的な枠組みで多重音分析システムとして統合し、それを音響入力データに適用し、推定精度を評価した。すなわち、複数声部楽曲の音楽信号を入力として、自動的に楽譜を出力するシステムの評価を行った。提案手法を、ボトムアップアプローチ、即ち音楽理論モデルを含まないシステムと比較した結果では、平均して50%もの改善があった。

本論文の大きな成果は、かつて音声認識分野に革命をもたらした音声言語モデルに相当する、多重音認識の音楽理論モデルが作成できることを証明しながら、MIRのいくつかのタスクに応用し、広範囲な用途がありながら解決が難しかった多重音分析の問題の解決策となる独創的な方法論を築き上げた点にあり、今後の計算能力の増加に伴ってこのような音楽理論モデルを含むトップダウン構造のシステムが一般的になるであろうと期待される。

よって本論文は博士(情報理工学)の学位請求論文として合格と認められる。