

氏名	澤田 隼
学位名	博士（システム情報科学）
学位記番号	第53号
学位授与年月日	令和3年3月22日
学位論文題目	音楽の記号創発に向けた信号と記号の相互最適化フレームワークの構築
論文審査委員	主査 平田 圭二 副査 鈴木 恵二 副査 三上 貞芳 副査 竹川 佳成 副査 高木 清二 副査 寺井 あすか

論文要旨

本論文では、音楽情報処理における記号創発を実現する枠組みである信号と記号の相互最適化フレームワークを提案し、その枠組みを実現したシステムに関する研究について述べる。

音楽情報処理の分野では楽譜などを対象とした記号処理と音響信号を対象とした信号処理の分野がそれぞれ独立して発展してきた。しかし、システムがより人間らしく振舞うためには、記号の領域と信号の領域が相互に作用する枠組みを構築し、適切な記号接地を実現させる必要があると考えている。人工知能の知識表現において、そこで使われる記号を実世界の実体をもつ意味に結び付けられるかという問題が記号接地問題である。一方で、静的な記号主義的な記号システムを前提とする記号接地問題はそもそも問題の前提がずれているとした新しい問題設定として、記号創発問題が掲げられている。本論文では、音楽を対象とした性質の異なる3つの問題を、創発的な記号と信号の相互最適化フレームワークという観点から解決を試みる。

第3章では、重要な要素の1つである拍節構造に着目し、拍の位置を調節することで、創発的に音楽のグルーピング構造における記号を得る。従来の音楽理論 GTTM の階層的なグルーピング構造を音楽のスペクトログラムから抽出する試みでは、スペクトログラムを拍の位置で分割し、そのセグメント毎に特徴量を抽出し、階層的クラスタリングを行っていた。しかし、スペクトログラムを分割する位置によって、そのセグメントが持つ特徴量の値が変化するが、スペクトログラムを分割する段階まで戻って修正する枠組みが無か

った。本章では楽曲の構造の距離（類似度）の持つ性質に基づいた目的関数を設計し、その目的関数の値を最小化するように、スペクトログラムを分割する段階へフィードバックすることで、適切な分割位置を獲得する枠組みを提案する。その結果、スペクトログラムの分割位置によって精度が変わり、適切な分割位置では期待する階層的なグルーピング構造が得られ、大域的構造を考慮した分割位置の修正が有用であることが示された。

第4章では、音楽における重要な要素の1つであるリズムに着目し、ドラムパターンの大域的な構造に基づいて、ドラムパターンの特徴量のハイパーパラメタを調節することで、創発的にリズムにおける記号を得る。従来のドラムパターンの差し替えシステムは、差し替えるドラムパターン同士のマッチングの際に、小節単位のリズムパターンの認識という局所的なリズムパターンの認識のレイヤーと、リズムパターン間の遷移確率に着目した大域的なリズムパターンの認識という異なる二つのレイヤーが存在している。しかし、局所的なパターンの認識の精度によって、大域的なパターンの認識の精度も変化するが、局所的なパターンの認識のレイヤーまで戻って、その認識を変化させる枠組みが無かった。本章では、一般的にドラムパターンが有する性質に基づいた目的関数を設計し、その目的関数の値を最小化するように、記号処理によるドラムパターン構造の認識結果を信号処理による小節単位のドラムパターン認識へとフィードバックすることで、ドラムパターンの認識性能を改善する枠組みを提案する。本枠組みを検証する実験を行い、実際にドラムパターンの認識性能を向上できることを確認した。

第5章では、音楽における重要な要素の一つである旋律に着目し、旋律における記号創発問題に取り組み、教師なしの旋律分割を可能にした。音楽の旋律に対して、統計的性質を用いて自然言語の「単語」のような基礎的な単位の導出を目指す。音楽情報処理の分野において、旋律を意味のある単位に分割することは楽曲構造の分析や自動作曲などの応用タスクにおいても重要な課題の一つである。音符単位のNグラム言語モデルと、フレーズ単位のNグラム言語モデルを統合した確率モデルから楽曲の旋律（音符系列）が生成されているとみなし、旋律の基礎的な単位（フレーズ）を推定する。旋律から言語モデルを学習し、その言語モデルから生成されたモチーフの遷移に関する統計的な情報を用いて言語モデルを更新するというサイクルを経て、創発的な旋律分割とその言語モデルが得られた。

本研究では、独立して発展してきた記号処理と、信号処理の研究を今一度見つめなおし、信号と記号の相互最適化フレームワークを提案した。

これは、音楽情報処理のさらなる発展に留まらず、音声認識、自然言語処理、画像認識等の発展に貢献するものである。また、時系列データの一つである音楽を対象として、コンピュータを用いて情報処理を行うことにより、その音楽自体の根底に迫ると同時に、人間の認知や知性を解明する手がかりとなり得る。そのため、本研究は認知科学、人工知能などの分野にも大きく寄与するものであると考えられる。

keyword: 音楽情報処理, 記号創発, 記号と信号の相互最適化

審査結果の要旨

本学位論文では、音楽情報処理における記号創発を実現する枠組みとして、信号と記号の相互最適化フレームワークを提案し、当該枠組みに基づく3種類の音楽識別システムの事例を示した。

音楽という非言語的なメディアには、言語におけるような明確な意味を持つ記号が存在しないため、適切な記号接地を行うのは一般的に困難である。人工知能分野において、静的な記号を想定した記号接地問題の次の段階として、記号の動的な側面を取り込んだ記号創発が提唱されたように、音楽情報処理においても、音楽の共時的かつ通時的な意味を理解するには、音楽における記号創発問題に取り組む必要がある。

1つめの音楽識別システムは、音楽理論GTTMの階層的なグルーピング構造を音楽のスペクトログラムから直接抽出するものである。創発的に記号としてのグルーピング構造を得るため、階層的に拍の位置を調節する必要がある。グルーピング構造どうしの距離（類似度）が持つべき特徴に基づく目的関数を設計し、その目的関数の値を最小化するように、スペクトログラム分割にフィードバックすることで、創発的に適切なグルーピング分割位置を獲得する。

2つめの音楽識別システムは、ドラムパターンの遷移に着目して記号としてのドラムパターン列を抽出するものである。信号処理による小節単位のドラムパターン認識から記号レベルのドラムパターンの大域的な構造を生成する処理と、一般的なドラムパターンが有する特徴に基づいて設計された目的関数を最小化するように、記号処理によるドラムパターン構造の認識結果を信号処理による小節単位のドラムパターン認識へとフィードバックする処理を組合せることで、創発的にリズムにおける記号を得る。

3つめの音楽識別システムは、教師なしで旋律分割を行い、記号としての旋律（モチーフ）の創発的抽出をするものである。音符単位のNグラム言語モデルと、フレーズ単位のNグラム言語モデルを統合した確率モデルから楽曲の旋律（音符系列）が生成されているとみなし、フレーズを推定する。旋律からこの階層的な言語モデルを学習し、その言語モデルから生成されたモチーフの遷移に関する統計的な情報を用いて言語モデルを更新するというサイクルを経て、創発的な旋律分割とその言語モデルを得る。

本学位論文で示された研究成果は、時系列非言語メディアの代表格である音楽情報処理における創発的な信号処理と記号処理の相互最適化を実現するだけでなく、音声認識、自然言語処理、画像認識等の広範囲のメディア処理発展に寄与する。また、当該研究成果によって音楽というメディア・コンテンツの本質に迫ると同時に、人間の認知や知性を解明する手がかりにもなり得るという意味において、システム情報科学にも大きく寄与する。

本学位論文で提案された内容は、学位授与に値すると思われるので、審査合格と判定する。