

内97-39

早稲田大学大学院理工学研究科

# 博 士 論 文 概 要

## 論 文 題 目

音楽音響信号を対象とした  
リアルタイムビートトラッキング  
に関する研究

申 請 者

後 藤 真 孝

Masataka Goto

電子・情報通信学専攻 情報構造研究

1997 年 12 月

これまで多くの音楽理解関連の研究がなされてきたが、音楽音響信号を人間のように理解できる計算機システムの構築は依然として難しい課題である。計算機による音楽音響信号の理解へ向けた典型的なアプローチは、音響信号から楽譜のようなシンボル表現を得る自動採譜・音源分離の研究である。このような詳細な採譜技術は重要であるが、これらがコンパクトディスク (CD) などによる人間が通常聞くのと同程度の複雑さを持った音響信号を扱うのは、現時点では大変難しい。音符やコード名を同定する能力は音楽的に訓練された人だけが持っていることからわかるように、採譜のようなシンボル化の能力は、実は人間にとっても獲得するのが難しい高度な技能である。

それに対して本研究では、音楽的に訓練されていない普通の人が、音響信号を楽譜のような表現にシンボル化できなくても、ある程度音楽を理解することができることに着目する。そして、まず最初に訓練されていない人のように音楽を理解するシステムを構築し、その後訓練された音楽家のように音楽を理解するシステムへと拡張するアプローチが重要であると考え、そのようなアプローチの基礎として、まず音楽において重要なビート (拍) や小節の概念の理解、つまりビートトラッキング (beat tracking) の実現を本研究の対象とする。音楽的に訓練された人と訓練されていない人のいずれにとっても、ビートは西洋音楽を理解する上で基本的な概念であり、ビートトラッキングは計算機による音楽理解モデルを実現する上で重要である。さらに、音楽との同期を必要とする多様なアプリケーションにおいても有用である。

従来のビートトラッキングに関する研究の多くは、MIDI などの音符がシンボル化された演奏情報報を対象としていたため、シンボル化の困難な音響信号には適用できなかった。音響信号を対象とした研究も報告されているが、その多くは四分音符に相当する単なるパルス列のビートしか認識できず、CD などによる音響信号をリアルタイムに処理できなかった。

それに対して本研究では、多様な楽器音や歌声の含まれたポピュラー音楽の音響信号に対し、階層的なビート構造をリアルタイムに認識するビートトラッキングシステムを実現する。階層的なビート構造は、四分音符レベル (四分音符に相当するビート時刻)、二分音符レベル、小節レベル (小節線に相当する時刻) の三つで構成される (図 1)。なお本研究では、ポピュラー音楽で一般的な 4/4 拍子でテンポがほぼ一定の曲を対象とし、音響信号中の打楽器音の有無は問わない。

本論文では、このような階層的なビート構造をリアルタイムに推定するビートトラッキング問題の解決法を、以下のような構成で論じる。

第 1 章では、本研究の目的、背景、および意義を明らかにする。本研究は音楽音響信号のビート構造を理解する計算機の処理モデルを、工学的に実現することを目的とする。聴覚的情景分析と音楽情報処理の両研究分野に位置付けられるが、従来の音源分離に基づく音楽理解やシンボル情報を出発点とした理解とは異なり、音響信号からのビートトラッキングという立場からの音楽理解を実現する。本研究は、テーマそのものに意義があることに加え、音楽理解、信号処理、人工知能、並

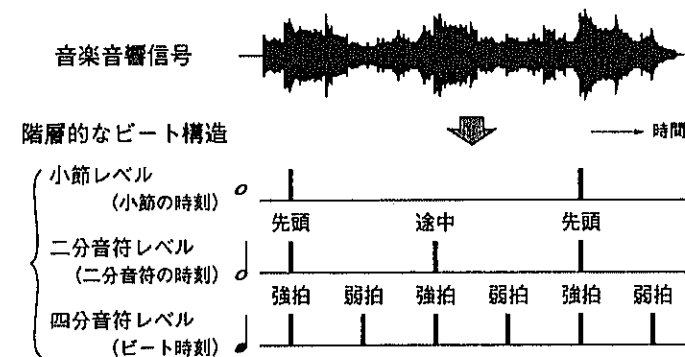


図 1: ビートトラッキング問題

列処理、コンピュータグラフィックス (CG) の各分野においても意義のある研究となっている。

第 2 章では、本研究で解決するビートトラッキング問題が、階層的なビート構造の認識であることを明確にし、この問題がビート提示行為の逆問題となっていることを説明する。順問題に相当するビート提示行為を考えると、ビート構造を特定の音で明示的に (元の構造が一意に決まるように) 表現しているわけではなく、様々な音楽の手がかりの関係の中に非明示的に表している。ビートトラッキングが本質的に難しいのは、音楽中に非明示的にしか表現されていないビート構造を音楽から推定する逆問題であるからである。このビートトラッキングを実現する上で解決すべき主要な課題は、(1) ビートの手がかりの決定と抽出、(2) 手がかりの解釈、(3) 解釈の曖昧性の取り扱いの三つである。

第 3 章では、ビートトラッキングのモデルを考えながら課題の解決法を示す。音響信号に対するビートトラッキングのモデルは、(a) ビート提示行為の逆モデル と (b) 聴覚の順モデル から成る。まず、ポピュラー音楽におけるビート提示行為の順モデルを考え、それから得られる妥当な仮定として、三つの音楽の手がかり (発音時刻、コード変化、ドラムパターン) に関する音楽的知識を導入する。これらを (a) ビート提示行為の逆モデル として用いることで、ビート構造を推定する。次に、これらの音楽的知識に基づいて三つの課題を以下のように解決する。

#### (1) ビートの手がかりの決定と抽出

(b) 聴覚の順モデル に相当する処理によって、上述の三つの音楽の手がかりを抽出する。その際に、コード変化とドラムパターンを、複数種類の楽器音を含む複雑な音響信号からボトムアップに得るのは困難である。そこで、仮に得られたビート時刻をトップダウン情報として用いることで、両者の抽出を可能にする手法を提案する。特に、提案するコード変化検出手法は、音符やコード名にシンボル化せずに変化だけを直接検出できる特長を持つ。

#### (2) 手がかりの解釈

発音時刻を手がかりとして音楽的知識を用いることで四分音符レベルを推定し、コード変化とドラムパターンを手がかりとして二分音符/小節レベルを推定する。打楽器音が含まれた音楽では、ドラムパターンを中心に音楽を聞く傾向があるため、コード変化とドラムパターンに関する音楽的知識は、提案する打楽器音の有無の判定手法の結果に応じて選択的に適用する。

#### (3) 解釈の曖昧性の取り扱い

手がかりの解釈は曖昧性があり一意に決まらないため、様々な解釈の可能性を評価できるようにマルチエージェントモデルを導入する。各エージェントは、相互作用、自己評価、適応の能力を持つ。エージェントは相互作用しながらお互いに異なる戦略で様々な解釈の可能性を調査し、その解釈の確信度を自己評価する。また、解釈する際の戦略を変更することで、現在の入力に適応しようとする。そして、最も確信度の高い解釈に基づいて最終的なビート構造を決定する。

第 4 章では、主要な処理である周波数解析とビート予測を中心に、マルチエージェントモデルに基づく処理モデルを説明する。まず、A/D 変換された音響信号に対して周波数解析をおこなう。ここでは、音響信号から得た周波数スペクトルを出発点として、発音時刻ベクトル、打楽器音 (バスドラムとスネアドラム) の発音時刻候補、打楽器音の有無の判定結果を求める。次にビート予測で

は、各エージェントが発音時刻ベクトルを手がかりとして、自己相関関数・相互相関関数により四分音符レベルの仮のビート時刻を推定する。エージェントと一対一に対応する高次情報抽出器は、この結果を受けとると、そのビート時刻に基づいてコード変化度とドラムパターンを抽出する。エージェントは、これらの結果を手がかりとして音楽的知識に基づき上位のビート構造を推定し、解釈の確信度の自己評価をおこなう。そして最終的には、全体のエージェントの解釈が統合され、確信度に基づいて適切なものが決定される。最後に出力結果であるビート情報が、ネットワークを通じて視覚化や聴覚化のアプリケーションプログラムへと送信される。

第5章では、並列化の手法と並列計算機 AP1000 上への実装方法を説明する。本研究では計算量の多い処理をリアルタイムに実行するために、並列計算機上に処理モデルを実装した。本実装では、従来のデータパラレル型並列処理とは異なり、質が異なる複数の処理を同時に実行する必要がある。そこで、並列計算機を構成するプロセッサをグループに分け、それぞれに処理の質に応じた4種類の並列処理方式(パイプライン処理、データパラレル処理、コントロールパラレル処理、分散協調処理)を適用する手法を提案した。そしてこれらのグループ間でデータを流し、計算機全体ではパイプライン状にデータを処理した。また、AP1000 上でリアルタイム処理を実現するために、時間確認可能な同期機構も提案した。最終的にビート情報をネットワーク上へ出力することで、アプリケーションの実装が容易になるだけでなく、ネットワーク上の様々なプロセスが同時に出力を活用できた。

第6章では、ビートトラッキングの精度を定量的に評価する尺度を提案し、それを用いた評価実験の結果を示す。システムの認識精度は、ポピュラー音楽の CD からサンプリングした 85 曲を対象に評価した。その結果、打楽器音の有無を問わず、階層的なビート構造の各レベルにおいて 86.7 %以上の高い認識率を得ることができた。各レベルのトラッキングの開始時刻は全曲の平均で 10～20 秒の範囲であり、トラッキングの平均誤差は最大でも 64 分音符の時間程度と小さかった。さらに、提案したコード変化検出手法の有効性を確認し、二種類のコード変化度に基づく音楽的判断が効果的であったことも実験により確認した。

第7章では、評価実験の結果を考察し、本論文が提案した処理モデルが、CD による実世界の音楽音響信号に対して有効に機能したことを結論付ける。次に、ビートトラッキングの観点から曲の難易度を評価する平均拍位置パワー差尺度を提案し、評価実験の対象曲の難易度について考察する。そして、従来のビートトラッキングの関連研究を紹介し、音響信号に対して階層的なビート構造をリアルタイムに理解するシステムは実現されていなかったことを指摘する。また処理モデルの限界を議論し、今後の研究課題を示す。

第8章では、音楽との同期を必要とする様々なアプリケーションにおいてビートトラッキングが有用であることを、具体例を挙げて紹介する。そして、本研究が応用例として実現した音楽のビートに合わせて踊る仮想の CG ダンサーを表示するシステムを紹介し、実装したシステムがリアルタイム CG という実際のアプリケーションにおいても有効であることを実証する。

最後に、第9章で本論文のまとめを述べる。本研究により、人間の音楽聴取過程の一段階である音響信号に対するビートトラッキングが、計算機上でリアルタイムに実現可能なことを示すことができ、人間のように音楽を理解する計算機システムの構築に一步近付くことができた。