

1 はじめに

自動採譜は、音楽を扱う上で重要な技術である。本研究では、演奏データでピッチ情報と時間情報が与えられているとき、音価(音符)、すなわち、リズムパターンの推定を行なう。これは例えば、器楽演奏をSMF(Standard Midi File)形式で記録したデータから自動採譜を行なう場合を想定している。ここでは、多声部を持つ楽曲のリズムパターンの認識を目的とする。

人間は演奏過程において、意識的、または無意識に音長を変動させて演奏するために、個々の音の音符の種類を音長から一意に決めることができない。このため、リズムパターンを認識する技術が必要である。齋藤ら[1]は、リズムパターンの認識と音声認識が同一の構造を持つ問題であることに着目して、隠れマルコフモデル(HMM: Hidden Markov Model)を用いて、拍もテンポも未知の単旋律の演奏に対して音符の種類や拍子を推定する方法を提案した。この方法を、浜中ら[3]は多声部をもつ楽曲の場合に応用したが、一定のテンポで演奏し拍位置が既知であることが条件であった。一方、我々は齋藤らの方法を拡張し、テンポに対して不変である特徴量を用いてリズム認識を行う手法を提案し、単旋律の演奏に対して有効であることを報告した[2]。

今回、我々は先行提案[2]を応用し、多声部を持つ楽曲のテンポも拍も未知である演奏から演奏曲の楽譜を復元する手法を提案する。

2 単旋律のリズムパターン認識

先行提案[2]における単旋律のリズムパターンの認識方法は、演奏データのIOI(Inter-Onset Interval)の時系列 $X = \{x_1, \dots, x_T\}$ から、音価の時系列 $Q = \{q_1, \dots, q_T\}$ を推定するものであった。人間の演奏を、音価列 Q を意図してIOIの時系列 X を生成する確率過程とモデル化することにより、解くべき問題は演奏 X に対して事後確率を最大化する Q を

$$\hat{Q} = \operatorname{argmax}_Q P(Q|X) = \operatorname{argmax}_Q P(X|Q)P(Q)$$

として得ることとして定式化される。よって、リズムパターン Q の生起確率 $P(Q)$ とリズムパターン Q をあるIOIの時系列 X で演奏する確率 $P(X|Q)$ を考える。

$P(Q)$ については n -gram モデルを用いる。モデルのパラメータ(遷移確率)は楽譜データから学習する。 $P(X|Q)$ は、音長比 $r_t = (r_t^{(1)}, r_t^{(2)}, r_t^{(3)})$ を特徴量(リズムベクトル)とした確率モデルを使用する。

$$r_t^{(i)} = \frac{x_{t+i}}{x_t + x_{t+1} + x_{t+2}}$$

r_t は正規分布に従うとする。さらに演奏の局所的なテンポを表す特徴量として3つの連続するIOIから求まる音価1.0当たりの時間長

$$\tau_t = \frac{x_t + x_{t+1} + x_{t+2}}{q_t + q_{t+1} + q_{t+2}}$$

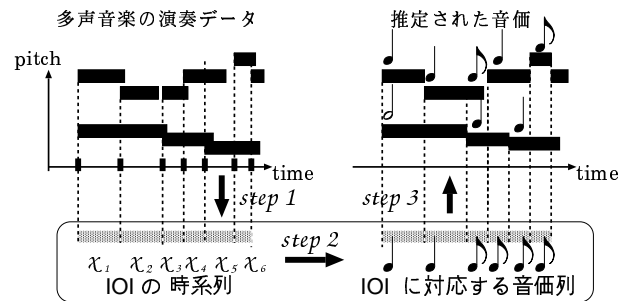


図1: 多声部を持つ楽曲の演奏のリズム推定の方法

を導入し、テンポの変動について $\tau_t - \tau_{t-1}$ が正規分布 $N(0, \sigma)$ に従うとする。

このような枠組みで構築される確率モデルは、HMMと捉えることができる[1, 2, 4]。最適音価列の探索はHMMにおける最尤状態経路を求めることに相当し、効率的な探索法としてViterbiアルゴリズムを用いる。

3 提案手法

単旋律のリズム認識では、演奏者の意図した各音の音長をIOIとして扱うことができた。しかし、多声部楽曲の演奏では、各音の意図された音長とIOIを同一視することはできない。例えば、長い間伸ばす音を演奏しても、そのすぐ後に別の声部の音が演奏されれば、音の継続時間は長くてもIOIは短くなる。また、単旋律のように演奏されている音価の系列を一次元の時系列での表現は、同時に複数の音が鳴る多声部楽曲の演奏ではできない。したがって、前述の単旋律のリズムパターンの認識方法を直接多声部楽曲に対して適用することはできなかった。

これに対して、我々は以下に、前処理と後処理を行なうことにより多声部楽曲に対するリズムパターンの認識が2.で述べた手法を用いることにより可能であることを述べる。図1の流れに従って、以下の3つの処理を行う。

Step 1: 同時発音の同定(前処理)

多声部楽曲の演奏に置いて、異なる声部で同時に発音された音はIOIが0となり、IOIの系列を得ることができる。実際の演奏では、同時に発音することを意図された音は実際には僅かな時間差をもって発音され、その僅かな時間差がIOIとして観測されることがある。同時発音を意図された音であるかどうかは、IOIが閾値 δ 以下であるかにより検出することにした。なお、前打音やトリルなどの裝飾音についてはここでは扱わない。これにより、IOIの時系列 $X = \{x_1, \dots, x_T\}$ が得られる。

Step 2: IOIに対応する音価を推定

Step 1で得られたIOIの時系列 X に対して、前述の単旋律のリズムパターンの認識の方法を適用する。ここで用いられるモデルのパラメータは、多声部楽曲の楽譜や演奏の統計から得られるものを用いる。その結果、IOIの時系列 X に対応する音価の時系列 $Q = \{q_1, \dots, q_T\}$ が得られる。

*“Automatic Music Transcription from Multi-voice MIDI Signals Using HMMs” by Haruto TAKEDA, Takuya NISHIMOTO, Koichi SHINODA and Shigeki SAGAYAMA (The University of Tokyo).

Step 3: 各音の音価を推定 (後処理)

Step 2 で推定した音価の時系列とそれぞれの音の継続時間の情報から、全ての音の音価を求める。具体的には、推定した局所的なテンポに基づいて継続時間を量子化し、音価を決定する。

4 評価実験

クラシック音楽のピアノ曲の演奏を用いて、提案手法の評価を試みた。多声部を含む楽曲では、8分音符や16分音符で刻む IOI が等間隔でつながる単純なリズムパターンの曲が多く、そのような曲は市販ソフトのクオンタイズ手法により比較的良好に認識できる。しかし、楽譜の指示や自然な表情づけなどでテンポの変動を含み、IOI のパターンも複雑である曲の場合は容易でない。例えばシューマン作曲「トロイメライ」(図 2) は、演奏を市販の楽譜作成ソフトの採譜機能を用いて音価の推定をしたところ、に対する部分の結果が図 3 となるように、全く認識できなかった。



図 2: 評価データのオリジナル楽譜の冒頭部



図 3: 市販ソフトによる評価データの採譜の冒頭部

評価データは、電子ピアノ (YAMAHA Clavinova CLP-970) による演奏者 1 名による「トロイメライ」の演奏の MIDI データを用いた。

Step 1 においては、 $\delta = 0.04$ (秒) に設定した。Step 2 においては、テンポ変動の分散、リズムベクトルの出現確率の平均と分散は、演奏データから学習した。また、状態遷移確率は、評価データの曲と異なるクラシックのピアノ曲の 7 曲楽譜データから学習したものの (open) と、対象曲も含めて学習したものの (closed) の 2 種類で行なった。出現頻度の少ない遷移は、統計的に信頼性のある確率値が学習できないので、線形補間 (重みは、unigram 0.3, bigram 0.1, trigram 0.1, quadgram 0.5) を行なった。Step 1, Step 2 における正解率 (IOI の音価正解率) は、IOI の音価正解率を n を正解音価列の総シンボル数としたとき

$$\frac{n - \text{置換誤り数} - \text{挿入誤り数} - \text{脱落誤り数}}{n} \times 100[\%]$$

と定義する。評価結果を表 1 に示す。挿入、脱落誤りは、Step 1 で同時発音が正しく検出できなかったことに相当する。また、音価の音価正解率 (Step 1, Step 2, Step 3 を通しての正解率) は、 N を全音符数として

$$\frac{N - \text{置換誤り数}}{N} \times 100[\%]$$

と定義する。評価結果を表 2 に示す。

今回の評価実験で誤認識を起こした原因は、以下の 2 点が考えられる。ひとつは、Step 1 の同時発音の検出で、対象曲に装飾音が含まれていたため、装

学習方法	正解率	n
closed	93.1%	218
open	70.6%	218

表 1: IOI に対する音価の正解率 (Step 1, 2)

学習方法	復元率	補正正解率	N
closed	81.7%	94.1%	460
open	61.5%	73.3%	460

飾音の分だけ短い IOI が挿入され、これが Step 2 における誤認識と数えられた。同時発音は全て正しく検出できており、装飾音に対する対処が必要である。もう一点は、実際の演奏が楽譜どおりに演奏されていないためである。トロイメライでは物理的に楽譜の音価のとおり鍵盤を押さえるのが困難である箇所があり、後処理 (Step 3) によっても原譜が復元できなかった。例えば図 4 に示すように、鍵盤上での腕の移動や、手の大きさや指の構造などにより、継続時間音が楽譜の指示よりも短く演奏し、誤認識と数えられている。実際の演奏制約のために、楽譜どおりに音が継続しなかった箇所を正解に数えた場合の認識率を、表 2 「補正認識率」に示す。装飾音の演奏と、記譜法と実際の演奏方法との違いについての対処は、今後の課題である。また、トロイメライに現れる IOI 系列のようなパターン (リズムベクトル) は、学習データ (7 曲) での出現頻度が小さいものが多く、そのために open, closed での認識率に大きな差が生じている



図 4: 音価推定の誤認識例 (左: 認識結果, 右: 正解)

5 おわりに

本稿では、多声部を持つ楽曲の演奏について、演奏の時間情報からリズムパターンを推定する方法について提案し、その効果について述べた。今後は、大量の多声部楽曲から IOI の n -gram の学習を行ない、演奏データも増やして性能評価を行ないたい。また、提案手法をクラシック音楽だけでなくポピュラー音楽やジャズ音楽などにも適用し、ジャンルの違いに頑健にリズムパターンの認識が可能であるか、検討を進めていきたい。

参考文献

- [1] 大槻知史, 齋藤直樹, 中井満, 下平博, 嵯峨山茂樹: “隠れマルコフモデルによる音楽リズムの認識,” 情報処理学会論文誌, Vol. 43, No. 2, pp. 245-255, 2002.
- [2] 武田晴登, 篠田浩一, 嵯峨山茂樹: “リズムベクトルの概念に基づくリズム認識,” 平成 11 年情報処理学会音楽情報科学研究会, Vol. 2002, No. 63, MUS-46, pp.23-28, 2002.
- [3] 浜中 雅俊, 後藤 真孝, 麻生 英樹, 大津 展之: “発音時刻の楽譜上の位置を確率モデルにより推定するクオンタイズ手法”, 情報処理学会論文誌, Vol.43, No.2, pp.234-244, February 2002.
- [4] H. Takeda, T. Otsuki, N. Saito, M. Nakai, H. Shimodaira, S. Sagayama, “Hidden Markov Model for Automatic Transcription of MIDI Signals,” Proc. 2002 IEEE Workshop on Multimedia Signal Processing (MMSP2002), 2002.