リズムベクトルを用いたHMMによる単旋律MIDI演奏の楽譜化*

◎武田晴登, 篠田浩一, 嵯峨山茂樹 (東京大学大学院 情報理工学系研究科)

1 はじめに

本研究は、MIDIキーボードの演奏から自動採譜を行うことを目的にしている。自動採譜を行うためには、音楽演奏の一つ一つの音に対して音高(ピッチ)と音符の種類の情報を得なければならない。音高は弾いたMIDIキーボードの鍵板から正確に得られる。一方、人間は意識的、あるいは、無意識に様々な要因で音長を変動させて演奏するので、一般に音長から音符の種類を一意に決定できない。このため、音長から楽譜にふさわしい音符の種類を決めるリズム認識が必要である。

市販のソフトでは、音長から音符を決定するために、音長を「量子化する(quantize)」方法が従来用いられている。この手法においては、音符は音長を閾値処理することにより決定されるが、メトロノームなどでテンポを強制したり、演奏後に拍の位置を指定するなど、拍が既知であることが条件である。しかし、テンポの強制は演奏者の自由な演奏を妨げ、また、自由に演奏した演奏データに拍を与えるのが常に容易とも限らない。さらにこのような条件のもとでも、実際に意図したリズムを表現した楽譜が得られないことが多い。例えば、図1に見られるように、人間がリズムパターンについて持っている常識に反する出力が得られることがある。

本稿では、テンポの指定のない演奏に対して拍の情報を与えることなくリズム認識を行う手法について述べる。

2 HMMによるリズム認識

我々は音声認識で用いられる統計的確率モデル(HMM: Hidden Markov Model, 隠れマルコフモデル)を用いたリズム認識の手法を提案してきた[1][2][3]。この手法では、演奏者がリズムパターン Q を意図して音長系列 X を演奏する演奏過程を確率モデルとしてモデル化する。そして、リズム認識を、与えられた演奏 X に対して、$P(Q|X)$ を最大にする $\hat{Q}$ を推定する問題として扱う。ベイズの定理より $P(Q|X) \propto P(X|Q)P(Q)$ であるので、

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

と定式化することができ、確率的逆問題となる。

このリズム認識の手法は、連続音声認識の手法と表1に示す対応関係にある。リズムパターンのモデルとしては、音符の n-gram モデル[1]や一小節単位のリズムパターン[3]を提案した。また、音長の変動については、固定テンポの演奏に対し1つの音の長さが確率的に変動するとした。そして、テンポの変動を許す演奏には、演奏に適したテンポの固定テンポのモデルを一小節単位に対応させることにより対処したが、複数の固定テンポのモデルを用意する必要があった。

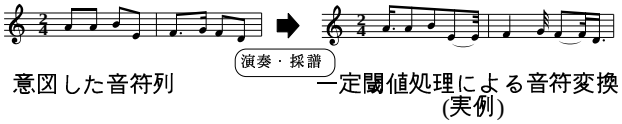

図 1: 閾値処理の誤認識例: 意図した音符列ではない

表 1: 音声認識とリズム認識の対応

	連続音声認識	リズム認識
入力単位	文音声	楽曲
語彙	単語	リズムパターン
単位モデル	音素	音符
隠れ状態	音響イベント	
観測値	スペクトル列	音長系列

3 リズムベクトルを用いたHMM

本研究では、音長の変動を表す新しい特徴量を提案する。即ち、従来のように固定テンポでの音長の変動をもとに考えるのではなく、音長の変動の要因をリズムパターンの揺れ、テンポの変動の2つに帰し、各々に対して新しい特徴量を導入する。これらの特徴量をHMMによってモデル化し、リズム認識を行う。

3.1 リズムベクトル

リズムの特徴を捉えるために、従来のように1つの音の長さの変動に注目するのではなく、連続する複数の音の相対的な関係である音長の比を特徴量とする。これはテンポの変動に対して不変な特徴量である。人間はテンポが変動してもリズムパターンを認識できるが、これは音長の比を手掛りにリズムを認識しているからであると指摘されている[4]。

音長系列 $\boldsymbol{d} = \{d_0, \cdots, d_T\}$ に対して、連続する3つの音長 d_t, d_{t+1}, d_{t+2} の比を成分とし、成分の和を1に規格化したベクトルをリズムベクトル $\boldsymbol{r}_t$ と定義すると、リズムベクトルの系列は $\{\boldsymbol{r}_0, \cdots, \boldsymbol{r}_{T-2}\}$ となる。人間の演奏のリズムベクトルは、楽譜に書かれている音符の音価のリズムベクトルに対して変動する。ここでは、この変動を正規分布に従うとし、確率密度を $b(\boldsymbol{r})$ で表す。なお、本研究では音長として発音時刻の間隔(IOI: inter-onset interval)を用いる。

ところで、リズムベクトルから音符を一意に決定することはできない。例えば、音長比が $1:1:1$ であるという情報のみでは、対応する音符が八分音符3つであるのか、四分音符3つであるのか判別できない。

3.2 テンポの局所的変動

従来のように固定テンポのモデルを複数用意するのでなく、テンポも特徴量として扱う。1つの音からテンポを計算するとテンポの変動が大きくなるので、連続する3つの音からテンポを求める。即ち、音長 d_t, d_{t+1}, d_{t+2} が音価 $q^{(0)}, q^{(1)}, q^{(2)}$ である音符に対応する演奏であるとき、これら3つの音に対するテンポを

$$\tau_i(d_t) \equiv \frac{d_t + d_{t+1} + d_{t+2}}{q^{(0)} + q^{(1)} + q^{(2)}}$$

とし、これを局所テンポと呼ぶことにする。

曲の途中から2倍速のテンポで音価が半分になる等の誤推定を防ぐために、局所テンポの変動に確率を設ける。局所テンポの差分 $\Delta\tau = \tau_i(d_t) - \tau_j(d_{t+1})$ が正規分布に従うとし、局所テンポの差分が $\Delta\tau$ である確率を $v(\Delta\tau)$ と表すことにする。

3.3 状態遷移確率

リズムベクトルが3つの連続する音符からの出力であるのに対応して、3つの連続する音符を1つの状態

* "Automatic transcription of MIDI performance using HMM with rhythm vectors" by Haruto TAKEDA, Koichi SHINODA and Shigeki SAGAYAMA (Graduate School of Information Science and Technology, The University of Tokyo).

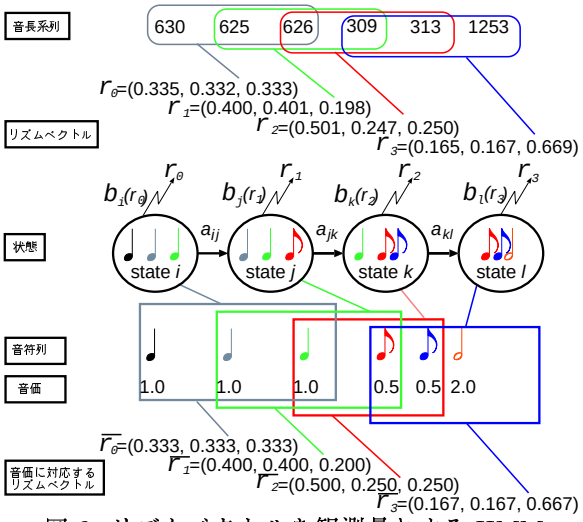

図 2: リズムベクトルを観測量とする HMM

表 2: 提案手法の HMM の枠組み

出力信号	リズムベクトル $\boldsymbol{r}_t$
状態	音符 3 つ組 $\{s\}$
状態遷移確率	a_{ij}, $v(\tau_i(d_t) - \tau_j(d_{t-1}))$
出力確率	$b_i(\boldsymbol{r}_t)$

とし、この状態系列に注目する (図 2)。ある時刻の状態 i から次の時刻の状態 j への遷移には、状態遷移確率 a_{ij} を設ける。遷移する状態の間では 2 つ音符が共有されており、ある時刻の状態において前の時刻の状態と共有していない新しい音符が、前の時刻の状態の 3 つの音符に依存して出現する。従って、この状態系列のモデルは音符の時系列を 4-gram とモデル化したことと等価である。3 つ以上の音符から成るリズムパターンは状態系列として表されるので、あるリズムパターンの出現確率は、そのリズムパターンに対応する状態系列における状態遷移確率の積で与えられる。したがって、状態遷移確率の値は、典型的なリズム、常識に反するリズムなど、人間がリズムパターンについて持っている知識を表現していると言える。状態遷移確率 $\{a_{ij}\}$ は、楽譜データの統計からの学習で定めることができる。

3.4 HMM

リズムベクトル、局所テンポの変動、状態系列を合わせて、HMM として捉えることができる。(表 2)

状態の系列が $\boldsymbol{s} = \{s_0, \cdots, s_{T-2}\}$ で出力音長系列が $\boldsymbol{d} = \{d_0, \cdots, d_T\}$ であるときの尤度 $l(\boldsymbol{s}|\boldsymbol{d})$ は、

$$\underbrace{\pi_{s_0}}_{\text{初期状態}} \cdot \prod_{t=1}^{T-2} \underbrace{a_{s_{t-1},s_t}}_{\text{状態遷移}} \cdot \underbrace{b_{s_{t-1}}(\boldsymbol{r}_t)}_{\text{出力}} \cdot \underbrace{v\left(\tau_{s_{t-1}}(d_{t-1}) - \tau_{s_t}(d_t)\right)}_{\text{テンポ変動}}$$

で表される。演奏データの音長系列 $\boldsymbol{d}$ に対して、上式で表される尤度を最大とする状態の系列 $\boldsymbol{s}$ を求めることが、リズム認識である。尤度を最大とする状態系列は Viterbi アルゴリズムによって効率良く求めることができる。

4 評価実験

単旋律の曲を対象に行った認識実験とその結果について述べる。評価データは、3 人の被験者 A, B, C による MIDI キーボード演奏である。演奏曲は、旋律としてのまとまりがあり、被験者にとっては既知である歌唱曲の旋律である。テンポの指定は行わなわず、演奏者は各曲について 2 回ずつ演奏を行った。また、学習データには、良く知られた単旋律として、戦前の唱歌 48 曲、「みんなのうた」より 19 曲、中学校の音楽の教科書より 6 曲、クラシック音楽の旋律より 11 曲を用いた。使用したモデルは、12 種の音符を組み合わせた $12^3 = 1728$ の状態を持つ HMM で、遷移確率は学習データから学習した。状態出力確率である 3 次元正規分布は、音符の音価列により決まるリズムベクトルを平均とし、平均の 0.1 倍を対角成分とする共分散行列を分散とした。

表 3: 各演奏の認識率

	演奏者					
	A		B		C	
曲名 (総音符数)	1 回目	2 回目	1 回目	2 回目	1 回目	2 回目
蛍の光 (55)	100.0	*100.0	96.4	98.2	98.2	98.2
こいのぼり (52)	98.1	98.1	92.3	96.2	88.5	82.7
ふるさと (44)	97.7	97.7	84.1	93.2	97.7	100.0
早春賦 (58)	98.3	96.6	100.0	96.6	100.0	98.3
若者たち (42)	97.6	97.6	80.1	73.8	100.0	97.6
喜びの歌 (61)	98.4	98.4	98.4	98.4	100.0	98.4

(* リズムとしては正しいが、音価は半分と認識された。)

表 3 にそれぞれの演奏に対する認識率を示す。認識率は、演奏データから推定された全音符数に対する楽譜に一致した音符数の割合として計算した。平均認識率は 96.0% である。

誤認識の多くは、音価を微小量だけ長く、或いは、短く推定してしまうものであった。しかし、例えば 1 つの音を 8 分音符 1 つ分の長さのだけ長く誤推定した場合、以後の音符が正しく認識されたとしても拍としては常識と合わないものになるので、このような微小な音価の伸縮の誤推定は、拍の知識をモデルに含めることにより改善が可能であろう。

5 おわりに

MIDI キーボードによる演奏を自動採譜するために必要となるリズム認識の一手法として、テンポの指定のない演奏に対してリズムベクトルを用いた HMM を用いる方法を提案した。テンポ不変量である「リズムベクトル」とテンポの確率変動を出力信号とした隠れマルコフモデルを、リズム認識に用いた。被験者 3 人による単旋律曲の MIDI キーボード演奏に対して認識率 96.0% を得た。

今後の課題として、より複雑なリズムを含んだ曲を対象に評価実験を行い、有効性を調べたい。また、提案手法の原理を拡張して多重音を含む演奏を対象としたリズム認識を行いたい。さらに、モデルのパラメータを演奏者や演奏曲に応じて設定することにより、認識性能の向上も目指したい。例えば、ある演奏者の実演奏から出力確率のパラメータを学習させることにより、その演奏者の演奏に対して高い性能を示すパラメータ設定が可能であろう。更に、特定のスタイルを持つジャンル、作曲家の曲だけを用いて遷移確率の学習を行えば、そのスタイルの曲に対して高性能となるパラメータの設定が可能であろう。

謝辞

本研究の一部は、科学技術振興事業団戦略的基礎研究推進事業 (CREST)(「脳を創る」聴覚脳研究プロジェクト) の支援を受けて行われた。また、MIDI キーボード演奏データを提供された被験者の方々に感謝する。

参考文献

[1] 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演奏からの音符列の推定,” 平成 11 年情報処理学会音楽情報科学研究会, Vol. 99-MUS-33, pp.27–32, Dec 1999.

[2] 大槻知史, 齊藤直樹, 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.

[3] 大槻 知史, 中井 満, 下平 博, 嵯峨山 茂樹, “HMM と音符 n-gram を用いた音楽リズム認識,” 情報処理学会音楽情報科学研究会, 2001.

[4] S. H. Hulse, A. H.Takeuchi, R. F. Braaten, “Percptual Invariances in the Comparative Psychology of

Music," Music Perception, vol. 10, No. 2, pp. 151-184, 1992.