

リズムベクトルを用いたリズム認識

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

〒113-0033 東京都文京区本郷 7-3-1

E-mail: {takeda,k-shino,sagayama}@hil.t.u-tokyo.ac.jp

あらまし 本稿では、MIDI キーボードによる人間の演奏から自動採譜を行うシステムについて述べる。鍵盤楽器から MIDI 信号を入力するため、ピッチについての情報は正確に得られるのに対し、リズムについては音の長さが得られるだけなので、音の長さを変換する処理が必要である。市販ソフトではメトロノームなどによりテンポを強制された演奏に対する自動採譜が実現されているが、テンポの指定を行わず自由に弾かれた演奏に対するリズム認識については有効な手法は確立されていない。我々はテンポの指定のない演奏に対するリズム認識を行うために、音楽の時間構造をテンポとリズムパターンに分けて考え、リズムパターンを表す特徴量である、3つの連続する音の長さの比を用いて表す特徴量である「リズムベクトル」に注目する。リズムベクトルとテンポ変動は、隠れマルコフモデルを用いてモデリングを行い、リズム認識の問題を推定の問題として扱う。本稿では、リズムベクトルに基づくリズム認識のための確率モデルを提案する。さらに、提案されるモデルを用いた実験として、3人の被験者の MIDI キーボードによる単旋律の曲の演奏について認識実験を行い、有効性を示唆する実験結果として認識率 93.2% が得られたことを報告する。

キーワード リズム認識、隠れマルコフモデル、リズムベクトル、テンポ

A Rhythm Recognition Method using Rhythm Vectors

Haruto TAKEDA, Koichi SHINODA, and Shigeki SAGAYAMA

Graduate School of Information Science and Technology, University of Tokyo

7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan

E-mail: {takeda,k-shino,sagayama}@hil.t.u-tokyo.ac.jp

Abstract This paper proposes a rhythm recognition method for MIDI signal performed by MIDI keyboard. An usual way of automatic transcription from MIDI signals is to play MIDI keyboard with metronome to perform in constant tempo and quantize the note durations in a resolution level which is given by the user. A new method proposed in this paper, however, does not require performer to obey the beats of metronome and can recognize rhythm pattern for automatic transcription. We define ratio of note durations as a new feature "Rhythm vector". Rhythm Vector and tempo variation are integrated in Hidden Markov Model, which is used in modern speech recognition, and we deal with rhythm recognition as a stochastic estimating problem. Experimental result is also reported.

Key words Rhythm Recognition, Hidden Markov Model, Rhythm Vector, Tempo

1. まえがき

MIDI キーボードによる演奏から自動採譜が行えれば、MIDI による音楽制作において入力作業がより効率的に行えるようになるであろう。この技術は、他の目的、例えば、即興演奏を保存したり、楽譜を書くのに慣れていない人が思い付いたメロディーを楽譜にしたり、あるいは、既にある MIDI データとして保存されている大量の演奏データから 1 度にたくさんの楽譜を得る目的などに用いることも可能であろう。採譜を行うた

めには、演奏のピッチ (音高) とリズムの情報が最低限必要である。MIDI キーボードの演奏から得られる MIDI 信号には、ノートナンバー (note number) としてピッチについては正確な情報が含まれるが、リズムについては演奏された音の長さが与えられるだけなので、採譜を行うためには、音の長さから適切な音符に変換するためのリズム認識の処理が必要になる。本研究は MIDI キーボードによる人間の演奏を楽譜にすることを目標にし、今回は対象を単旋律に限定し、これを楽譜にするためのリズム認識の問題を扱う。

この目的のためのリズム認識の方法としては、市販のソフトで広く用いられている「量子化」する(クオンタイズ、quantize)手法が知られている。これは、メトロノームなどによる決められたテンポに従ってユーザがMIDIキーボードを弾き、ユーザが定めた分解能(リゾリューション: resolution)に従って閾値処理を行い楽譜上の音符の長さに対してはめるといものである。しかしこの方法では、ユーザは厳密に一定テンポで演奏しなくてはならないという制限を受ける。また、MIDIキーボードに慣れていないユーザにとっては意図したリズムを厳密な長さで演奏できないため、意図したリズムの表現された楽譜が得られないことが多い。図1は、実際に市販のソフトを用いて演奏を入力したときの例であるが、ユーザが指定したテンポと分解能に従って演奏データを忠実に量子化した結果であるが、これは演奏者の意図とは異なるものである。

この例に見られるように、音の長さに「忠実」すぎる音符の変換を行ってしまうことが量子化処理の問題点であり、量子化とは別の方法として、例えば隣接する音の長さの比が有理数になれば安定するエネルギー関数により、安定するまで処理を繰り返す方法[1]などが提案されてきた。

一方で、人間がリズムに対して持っている常識を事前知識として活用する方法が考えられる。先の例では、人間は図1の下段に見られるようなリズムがほとんど現れないことを知識として持っているので、人間がこのようにリズムを認識することはほぼあり得ない。我々は音声認識で用いられている隠れマルコフモデル(ex. [2])をリズム認識に応用し、演奏される音の長さの時系列を確率を用いたモデルからの出力と捉え、事前知識としてモデルのパラメータを学習させてから認識を行う手法を提案し、研究を行ってきた[3][4][5]。このような手法が、テンポが指定された演奏に対して有効な認識性能を持つことが実験によって示されている。また、この方法が多重音を含む演奏に対して適用された例もある[6]。

しかしながら、メトロノームによるテンポの指定のない演奏に対するリズム認識については、確立された有効な標準的な手段は今のところない[7][8]。斎藤ら[3]は、1小節単位のモデルを5種類の異なるテンポについて用意し、それらの間に遷移確率を与えることによってテンポ変化に追従できるとするモデルを提案し、実験で有効性を確認している。

音の長さの変動の扱い方について、斎藤らは、音の1つひとつに単純な確率モデルを当てはめたのに対し、我々は、テンポの指定のない演奏に対しては、テンポの情報を含まない特徴量を使う方が適切であると考え、3つの連続する音の長さの比をとった「リズムベクトル」を用いることにした。音の長さの比をとり認識に用いるという考え方は、旋律のパターン認識においてテンプレートと参照するために入力パターンとの距離を定義するために既に用いられている[9]が、ここでは未知の旋律に対するリズム認識としては用いられていない。本稿では「リズムベクトル」とテンポ変化を持つHMMを提案し、その実験結果について述べる。



意図した音符列

演奏・採譜



一定閾値処理による音符変換 (実例)

図1 閾値処理によるリズムの誤認識の例

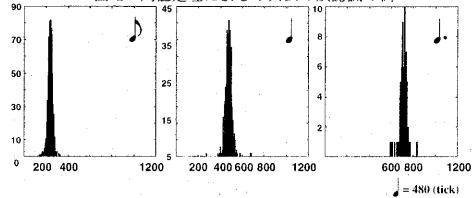


図2 テンポ指定時の演奏の音長分布

2. 音長の変動とリズムベクトル

人間による演奏の音長の変動に対して確率モデルを与える。

2.1 音符に対する音価の定義

楽譜上の音符だけから得られる音の長さの情報として、「音価」を、4分音符の長さを1としたときの理論的な比に従った長さとして定義する。例えば8分音符の音価は、4分音符の半分の長さなので $q = 0.5$ となる。その他の音符の音価は、表1に示すとおりになる。

表1 音符と音価

音符	(音符シンボル n)	音価 q
16分音符	(S)	0.25
8分3連符	(tE)	0.33
8分音符	(E)	0.5
8分3連符 2つ分の長さ	(tEtE)	0.67
付点8分音符	(dE)	0.75
4分音符	(Q)	1.00
付点4分音符	(dQ)	1.50
2分音符	(H)	2.00
8分音符 5つ分の長さ	(HE)	2.5
付点2分音符	(DH)	3.0
8分音符 5つ分の長さ	(DDH)	3.5
全音符	(W)	4.00

2.2 テンポ指定時の演奏の音長の変動

本稿では、演奏データの音の長さとしてはMIDI信号のdurationではなく、発音時刻から次の音の発音時刻までの間の時間(IOI: inter-onset interval)を用いることにする。IOIを用いる理由は、人間は演奏の際に鍵盤を押さえている時間よりも打鍵のタイミングに注意を払っているという知見があり、onset-timeとdurationの両方の揺らぎを同時に扱わなくてはならないという困難な状態を避けるためでもある。

テンポがメトロノームなどによって指定された状態でも、演奏される音の長さは、テンポと音符の音価とから計算される音の長さに対して変動する。図2は、テンポをメトロノームで指

定された状態で演奏された演奏データの中から8分音符、4分音符、付点4分音符を弾いたときの IOI を取り出して度数分布として表示したものであり、テンポと音価から計算される音の長さを中心に分布していることが理解できる。

斎藤、大槻ら [3] [4] [5] のモデルでは、テンポが一定である演奏にみられるこのような音の長さの変動に対して、1つひとつの音の長さに対して確率分布を仮定して用いていた。しかし、テンポを指定されていない状態での演奏に対しては、音の長さの変動はより複雑になるので、新しい確率モデルが必要とされる。

2.3 音楽の時間構造を考慮したモデリング

我々は、テンポの指定のない演奏の音の長さの変動に対して、リズムパターンの変動とテンポの変動の2つの変動に分けて考えることにする。

このように変動要因を2つに分けるモデリングは工学的な見地からのみでなく、音楽心理学の見地からみても妥当であると考えられる。実際、Hulse らも、音楽の時間構造はリズムパターンと拍子 (meter) とテンポによって記述され、リズムパターンと拍子は密接に結び付いた関係にあり、テンポはこれらから独立した別の自由度をもつ特徴量であることを述べ、人間はテンポが変化しても同じリズムパターンを認識することができる事実について論じている [10]。この指摘にあるように、リズムパターンと拍子は相互に関連しているが、今回はリズムパターンのみを扱うことにする。

2.4 リズムベクトル

「リズムベクトル」を3つの連続する音の長さを組にして3つの成分の和が1になるように正規化したものとして定義する。図3に示されるように、音符を3つずつ組にして、各組に対してリズムベクトルを求め、リズムベクトルの時系列を得ることができる。即ち、演奏データを音長系列 $\{d_0, \dots, d_T\}$ で表したとき、この音長系列に対してリズムベクトルの時系列 $\{r_0, \dots, r_{T-2}\}$ を

$$r_t = (r_t^{(0)}, r_t^{(1)}, r_t^{(2)})$$

ただし、

$$r_t^{(0)} + r_t^{(1)} + r_t^{(2)} = 1, \quad r_t^{(0)} : r_t^{(1)} : r_t^{(2)} = d_t : d_{t+1} : d_{t+2}$$

と定義する。 d の添字は、演奏された音の長さに順番に番号を付与したもので、演奏された音の数を $T+1$ 個とする。

2.5 リズムベクトルの観測

リズムベクトルは3つの成分の和を1にするという正規化を行っているので、3次元上の空間において正三角形 $\{(r^{(0)}, r^{(1)}, r^{(2)}) | r^{(0)} + r^{(1)} + r^{(2)} = 1, r^{(0)} > 0, r^{(1)} > 0, r^{(2)} > 0\}$ の上の点となる。このため、リズムベクトルの分布を平面上の分布として観測することができる。図4(A)は、図4(C)に示された楽譜の音価から計算されるリズムベクトルの分布を示したものである。

2.6 人間の演奏に見られるリズムベクトルの変動

一方、図4(B)に見られる図形は、図4(C)を実際に人間が MIDI キーボードで演奏したデータのリズムベクトルの分布で

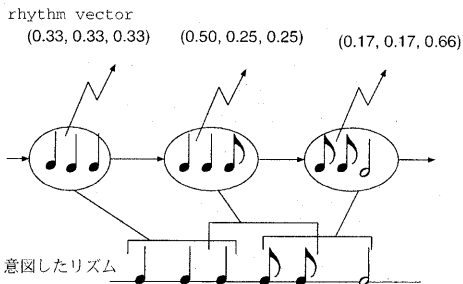


図3 リズムベクトルの出力の図

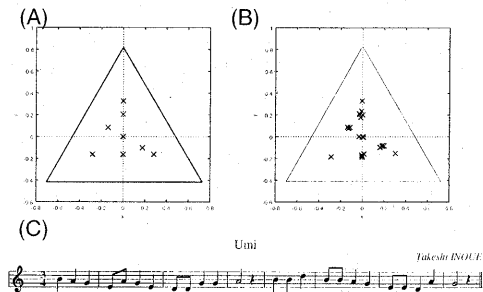


図4 リズムベクトルの様子

あるが、人間の演奏によるリズムベクトルが、意図した音符の音価から計算されるリズムベクトルに対して変動していることが分かる。

2.7 リズムベクトルの出力確率

人間の演奏したデータのリズムベクトルの変動を、確率を用いて表すことにする。人間の演奏のリズムベクトル r が得られる確率は

$$b(r) = b(r^{(1)}, r^{(2)}, r^{(3)}) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma|}} \exp\left(-\frac{1}{2}(r - \mu)\Sigma^{-1}(r - \mu)\right) \quad (1)$$

という正規分布に従うと仮定する。ここで μ はリズムベクトルの確率分布の平均を表す3次元ベクトル、 Σ は共分散行列を表す。

2.8 テンポの定義

人間は音楽の演奏において無意識にあるいは意識的にテンポを変動させる。このテンポの変化を捉えるために、テンポを計算する必要がある。音長 (d_t, d_{t+1}, d_{t+2}) が音価 $q_t^{(0)}, q_t^{(1)}, q_t^{(2)}$ に対応する音符を演奏したときのテンポを

$$Tp_t(d_t) = \frac{d_t + d_{t+1} + d_{t+2}}{q_t^{(0)} + q_t^{(1)} + q_t^{(2)}} \quad (2)$$

と定義する。1つの音符に対してテンポを定義せず3つの音からテンポを算出しているため、計算されるテンポは3つの音の平均化された値になる。このことは、例えば、付点音符を含むリズムパターン等を正確な長さで弾けなかったときなど、極端に異なる長さで弾いた音符の長さが平均化されるので、局所テンポの大きな変動を抑えることが可能となる。

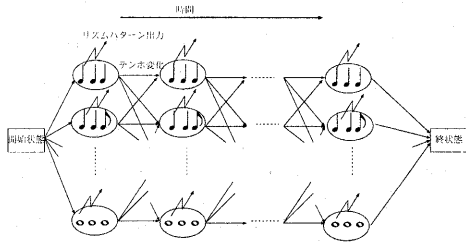


図5 確率モデルの全体構造

2.9 テンポの変動の確率モデル

演奏のテンポの変動も、確率を用いて考える。ただし、曲の中で急激なテンポの変化が指定されている曲は対象にしないこととする。ある時刻のテンポは、その直前のテンポに依存して確率的に起きるとし、その確率は

$$v(\text{Tp}_i(d_t), \text{Tp}_j(d_{t+1})) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left[-\frac{(\text{Tp}_i(d_t) - \text{Tp}_j(d_{t+1}))^2}{2\sigma_i^2} \right] \quad (3)$$

であるとする。直前のテンポと同じテンポである確率ももっとも高く、テンポの変化が大きくなる程、低い確率が与えられる。

斎藤らの5つの固定テンポのモデルを用意する方法[4]では、小節の中でのテンポの変化や、固定された5つのテンポ以外のテンポでの演奏を明示的にモデルの中に入れていなかった。これに対して、ここで提案されるモデルでは、3つの音符からのテンポを計算し、テンポを連続値として扱っているため、一小節より細かい範囲のテンポや、連続値としてのテンポの変化もモデルの中に含まれている。

以上により、テンポの指定のない条件での人間の演奏の音の長さの変動のモデルとして、リズムベクトルとテンポ変動の確率モデルが得られた。

3. リズムベクトルを用いた隠れマルコフモデル

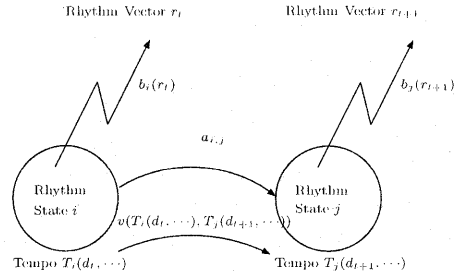
リズムベクトルとテンポの変動を隠れマルコフモデルの枠組みの中に入れ、音長系列を生成する確率モデルについて述べる。

3.1 確率モデルの基本的な構造

連続する3つの音符を組にしたものを隠れ状態に対応させ、この状態がリズムベクトルを出力する隠れマルコフモデルを考える。さらに状態の間を遷移するときのテンポの変化する確率も、このモデルの尤度計算に含める。式(2)に表されているようにテンポは3つの音の長さとして状態の属する3つの音符の音価から計算され、このテンポをもとにテンポの変化する確率を式(3)に従って計算するので、隠れマルコフモデルに特徴量でも遷移確率などのモデルのパラメータでもない量をモデルの中で扱うことになる。この点が、厳密に言えば通常用いられている隠れマルコフモデルと若干異なる。図5にあるように、この状態がマルコフ過程として時系列となる。

3.2 状態遷移確率

状態 i の次の時刻の状態が j となる確率を、状態遷移確率 a_{ij}



$$\text{Likelihood: } b_i(r_t) \cdot a_{i,j} \cdot v(T_i(d_t, \dots), T_j(d_{t+1}, \dots)) \cdot b_j(r_{t+1})$$

図6 リズムベクトルを用いた確率モデルにおける状態出力の概念図

とする。状態遷移確率は時間に依存しないが、音符列に対応しないダミーの状態0を用意し、 a_{0s} は時刻0における状態 s の出現確率(初期状態の確率分布)、 a_{s0} は状態遷移の時系列の最終時刻 $T-2$ における状態 s の出現確率を表すことにする。また、遷移確率 a_{ij} は論理的にありえないリズムパターン(3連符ひとつが4分音符に挟まれるなど)の出現を禁止する。さらに、音符を3つの組にしているため、前後の状態で共有されている2つの音符は同一のものでなくてはならないという制約も必要である。この状態遷移確率 a_{ij} は楽譜データから学習し、学習データに対して最尤推定値を求めることができる。

3.3 状態出力確率

状態からのリズムベクトルの出力確率は、隠れ状態のもつ音符に対応した音価に対応して式(1)で与えられるとする。ただし時刻0に対応する時刻では、リズムベクトルではなく、音の絶対長を用いて出力確率を計算する。このときの確率分布は、各音符に対してあらかじめ定めておいた音長を中心とする確率分布を用いる。

3.4 尤度計算

以上のモデリングから、状態 i から状態 j に遷移するときの尤度は、図6で表すように状態からのリズムベクトルの出力確率と状態遷移確率とテンポの変動確率を掛け合わせたものとなり、 $b_i(r_t) \cdot a_{ij} \cdot v(\text{Tp}_i(d_t), \text{Tp}_j(d_{t+1})) \cdot b_j(r_{t+1})$ となる。したがって、隠れ状態の系列が $\{s_0, \dots, s_{T-2}\}$ で出力音長系列が $\{d_0, \dots, d_T\}$ であるときの尤度は、

$$a_{0s_0} \prod_{t=1}^{T-2} a_{s_{t-1}, s_t} b_{s_t}(r_t) v(\text{Tp}_{s_{t-1}}(d_{t-1}), \text{Tp}_{s_t}(d_t)) a_{s_T, 0}$$

で表される。

4. HMMとリズムベクトルによるリズム認識

リズム認識を、前章で述べたリズムベクトルを用いたHMMを用いて定式化する。

4.1 確率的逆問題としての定式化

リズム認識の問題は、入力された音長系列(演奏データ)に対して、リズムパターンの系列を探すことにある。これは、先に

述べた確率モデルにおける最尤推定に相当し、確率的逆問題として考えることができる。

人がリズムパターン Q を演奏しようという意図を持って音長系列 X を演奏したとき、リズム認識とは、観測された X から Q を正しく推定することであり、推定は最尤原理に基づけば $P(Q|X)$ を最大にする Q^* を求めることにある。 $(P(Q|X)$ は X という事象が起きたときに Q が起きるという条件付確率を表す。) ここで、ベイズの定理によれば

$$P(Q|X) = \frac{P(X|Q)P(Q)}{P(X)}$$

であるので、 $P(Q|X)$ を最大にする Q^* を求める問題は $P(X|Q)P(Q)$ を最大にする Q^* を求める問題

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

と同値である。 $P(Q)$ はリズムパターン Q の現れる確率であり、 $P(X|Q)$ は Q というリズムパターンに対して X が演奏される確率である。我々の提案した HMM について適用すると、演奏データ $\{d_0, d_1, \dots, d_T\}$ に対して、演奏者が意図したリズムパターンに対応した状態系列 $\{s_0^*, \dots, s_T^*\}$ は、

$$\{s_0^*, \dots, s_T^*\} = \operatorname{argmax}_{s_0, \dots, s_T} \left[a_{0s_0} \prod_{t=1}^T a_{s_{t-1}, s_t} b_{s_t}(r_t) v(Tp_{s_{t-1}}(d_{t-1}, \dots), Tp_{s_t}(d_t, \dots)) a_{s_T, 0} \right]$$

と推測される。

5. リズムベクトルを用いた HMM の認識実験

認識実験とその結果について述べる。

5.1 モデルの生成

出力される音符の種類を 12 種類を表 1 に従って定め、これらの音符を組み合わせて隠れ状態を作った。音符の種類が 12 種類でこれらを 3 つ組にして状態をつつたので、状態の総数は $12^3 = 1728$ となった。

遷移確率の学習には、戦前の唱歌 [11]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、クラシック音楽より 11 曲の旋律を用いた。

認識実験に用いるデータとしては、3 人の被検者 A, B, C による MIDI キーボードの演奏を用いた。被検者 A はピアノ経験者で、被検者 B, C はピアノ以外の楽器経験者であり、演奏は各曲とも二回ずつを採取した。演奏条件は、MIDI 信号の設定は basetime を 60 にし tick を 480 に設定し、メトロノームの音は鳴らさず、テンポについての指定は全く行わず、被検者には楽譜のみを渡し、被検者が自由に弾けるように設定した。演奏曲は、旋律としてのまとまりがあるように、歌唱曲の旋律を用いた。

また、今回は隠れ状態の数がとても多く、ある状態に対しては十分な学習データを用意できず、頑健なパラメータの推定が困難であるため、出力確率については、平均と分散はあらかじめ与えておくことにした。平均は、音符の音価から計算されるリズムベクトルを用い、分散には対角成分が平均リズムベクトルの 0.1 倍になるように設定した。

表 2 隠れ状態

状態	音符	中心リズムベクトル	IOI 中心
1	Q Q Q	(0.33, 0.33, 0.33)	480
⋮	⋮	⋮	⋮
1728	W W W	(0.33, 0.33, 0.33)	1920

5.2 認識結果

認識率は、入力楽譜と同じ音符を出力したら正解とし、異なる音符を出力したら誤りとした。表 3 は認識率を表したものである。対象とする曲がいずれも単純なリズム構造をしたものであるが、どの対象曲に対しても高い認識率が得られた。全体の認識率は、93.2%となった。

表 3 演奏に対する認識率 [%]

曲名 (総音符数) \ 演奏	A1	A2	B1	B2	C1	C2
蛍の光 (55)	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	96.6	100	98.3
若者たち (42)	97.6	97.6	80.1	73.8	100	97.6
喜びの歌 (61)	98.4	98.4	98.4	98.4	100	98.4

5.3 誤認識の例

誤認識の例をいくつか挙げてみる。

(1) テンポの誤推定

演奏 A2 の「蛍の光」の認識率は 0% であるが、これは、図 7 に示すように、8 分音符にあたる音の長さが 4 分音符として認識されるなど、ほとんど全ての音価が正解の音価の半分として認識されたためである。これはリズムとしては正しいので、正解とみなすことも可能であるが、今回は正解には含めなかった。

(2) 細かい音の長さの違いについて

図 8 は、「早春賦」における誤認識の例である。8 分音符 5 個の長さを 4 個分の長さで誤認識している。この例に見られるように、細かい音の長さの違いについては、今回提案した確率モデルのように局所的な音符列の履歴と音の比だけでなく、拍子(ビート)の情報も含めて総合的に処理することによって解決される可能性がある。

図 7 音価が半分と誤認識された例
誤認識の例



図 8 8 分音符単位の音の長さの誤認識の例
誤認識



6. ま と め

本稿では、MIDI キーボードによる演奏を自動採譜するため必要となるリズム認識の方法について、従来では認識が困難であった、テンポの制限のない演奏を対象にして行った研究について述べた。はじめに、音の長さの変動はテンポの変動とリズムパターンの変動で表されると仮定し、連続する3つの音の長さの比を用いた「リズムベクトル」として、リズムベクトルとテンポの2つの変動を用いて、音の長さの変動の確率モデルを与えた。次に、音の長さの変動モデルと、リズムパターンの生成モデルをHMMを用いて統合し、意図したリズムパターンに対して演奏された音長を確率により出力する確率モデルを提案した。最後に、提案されたHMMを用いて、対象曲を単旋律に限定し、3人の被験者のMIDI キーボード演奏に対して行い、認識率93.2%が得られ、モデルの有効性が示唆されたことを述べた。本稿で提案したモデルについて、今後、以下に述べるような課題について取り組みたい。

● 対象音楽に特化したリズム認識モデル

今回の実験では、遷移確率についてのみ学習を行ったが、出力確率のパラメータについても、リズムパターンの平均と分散を演奏データから学習して求めることが可能である。例えば、特定の個人の演奏データを用いて学習を行うことにより、個人の演奏のリズムパターンの特徴を反映したパラメータ値を推定したり、あるいは、不特定多数の演奏者の演奏データを用いて学習を行うことにより、不特定の演奏者に対して頑健であるパラメータ推定も可能であろう。さらに、遷移確率に用いる楽譜データも、対象をジャンルや作曲家などにより区別し、各ジャンル、作曲家別の曲の学習を行い、それぞれの対象とするジャンルや作曲家の作品にに対して頑健になるように学習を行うことも可能である。このように、それぞれ目的に特化した学習を行うことによる認識性能の向上が考えられる。

● 多重音を含む(多声部の音楽の)演奏への対応

今回は単旋律を対象としたが、今後は、多重音を含む演奏に対してもリズム認識を行いたい。本稿で提案されたリズム認識の手法は、単旋律に特化した手法ではなく、基本的に多重音を含む演奏に対しても適用可能である。今後は、本稿で提案したリズムパターンとテンポを組み込んだモデルを活用する方向で研究を進めたい。

7. 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。

文 献

- [1] P. Desain and H. Honing, "The Quantization of Musical Time: a Connectionist Approach," *Comp. Mus. J.*, Vol. 13, No. 3, pp. 56-66, 1989.
- [2] L. Rabiner, and B.-H. Juang: *Fundamentals of Speech Recognition*, Prentice-Hall, 1993.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] M. Hamanaka, M. Goto, H. Asoh, and N. Otsu: "Learning-Based Quantization: Estimation of Onset Times in a Musical Score," *Proc. SCI 2001*, Vol. X, pp. 374-379, 2001.
- [7] R. Curtis: *The Computer Music Tutorial*, MIT Press, Cambridge, 1996.
- [8] 長嶋, 橋本, 平賀, 平田: *コンピュータと音楽の世界, bit 別冊*, 共立出版, 1998.
- [9] E. J. Coyle, I. Shmulevich, "A System For Machine Recognition of Music Patterns," *Proc. of ICASSP98*, 3597-3600, 1998.
- [10] S. H. Hulse, A. H. Takeuchi, R. F. Braaten, "Perceptual Invariances in the Comparative Psychology of Music," *Music Perception*, vol. 10, No. 2, pp. 151-184, 1992.
- [11] 堀内, 井上: *日本唱歌集*, 岩波文庫, 1958.