

確率モデルによる MIDI 演奏のテンポ変化時点の検出*

武田晴登 西本卓也 嵯峨山茂樹 (東大・情報理工)

1 はじめに

我々は、人間の手による演奏(以下、実演奏)を記録した MIDI (Musical Instruments Digital Interface) 信号からの自動採譜の実現を目指して、リズム認識の研究を行っている。

従来のリズム認識の手法の多くは、テンポがほぼ一定であることを仮定していた。例えば、音長をクラスタリングする手法 [1]、発音時刻の位置についての確率モデル [2, 3] を用いる手法が提案され、我々もテンポに不変な特徴量を用いたリズム推定の方法を提案した [4]。また、途中で変化するテンポを推定する手法としては、複数のテンポモデルを用意する手法を提案した [5] が、予め定められた数個のテンポ値から選択するので、テンポを連続値として推定するものではなかった。

本稿では、途中でテンポが変動する曲についてもテンポを連続値として推定する手法を提案する。

2 リズム認識

2.1 音長とリズムとテンポの関係

〔リズム〕 楽譜上の音符の正規の長さは「音価」(time value; 時価ともいう)と呼ばれる。音価 q は、たとえば四分音符を単位長としてそれと整数関係にある離散的な量として扱うことができ、 $= 1[\text{拍}]$ ならば $= \frac{1}{2}[\text{拍}]$ である。音価の並びはリズムパターンとして知覚されるので、ここでは用語として音価の並びを「リズム」と呼ぶことにする。

〔音長〕 一方、音符が演奏され観測された音の物理的長さを「音長」と呼ぶ。音長 $x[\text{秒}]$ は、より正確には音の長さとして認知されるような物理的な時間量であり、ここでは音符の発音時刻の間隔 (inter-onset interval, IOI) として定義する。

〔テンポ〕 音長 $x[\text{秒}]$ は音価 $q[\text{拍}]$ と演奏テンポを反映する。演奏の音価あたりの時間を $\tau[\text{秒}/\text{拍}]$ として、この関係は

$$x[\text{秒}] = \tau[\text{秒}/\text{拍}] \times q[\text{拍}] \quad (1)$$

と書ける。本稿では τ をテンポと呼ぶが、メトロノーム表記のテンポ (毎分の拍数) には反比例する量である。

2.2 不良設定問題としてのリズム認識

リズム認識およびテンポ推定は、式 (1) において音長をリズムとテンポに分解する問題である。音長 X という一変数の時系列からテンポ T とリズム Q という二変数の時系列を推定するという意味で、一種の不良設定問題と捉えられる。テンポとリズムの間には不確実性が存在し、例えば同一のリズムを倍の音価を用いて複数の方法で表現できる。

人間は、音楽を聴いて本来不可分であるリズムとテンポへの分離が行えるのは、よく現れるリズムパターンや、テンポは緩やかに変動するという常識を

持っているからであると考えられる。我々はそのような常識を反映した確率モデルを用いてリズム推定とテンポ推定を行う。

このためには、まずテンポに対して不変である音長比「リズムベクトル」[4] を特徴量に用いてリズムを初期推定した後、推定したリズムと音長をもとにテンポ推定を行い、これを繰り返すことにより事後確率最大のリズム、テンポ分解を行う構想を持っている。今回は、その第一段において、複数のテンポの問題を扱う。

2.3 確率モデルを用いたリズム認識

確率モデルを用いたリズム推定

テンポが未知である演奏の音長の時系列 X からリズム Q を確率モデルを用いて推定するには、与えられた X に対して最も尤もらしい Q として $P(Q|X)$ を最大にする Q を求めれば良い。これは、Bayes の定理を用いると、

$$\hat{Q} = \underset{Q}{\operatorname{argmax}} P(X|Q)P(Q) \quad (2)$$

と表される。 $P(X|Q)$ はリズムが X のように演奏される確率を表し、ここでは特徴量 X としてリズムベクトルを用いる。また、 $P(Q)$ はリズム Q の出現確率を表し、ここでは音価の n -gram モデルを用いる。両者の確率とも、実演奏と既存の楽曲のリズムを用いて事前に統計的に学習できる。これらを組み合わせたモデルは、HMM (Hidden Markov Model) として捉えることができ、(2) は Viterbi 探索問題として解ける [5, 6]。

確率モデルを用いたテンポ推定

リズム認識により推定された音価の時系列 $\hat{Q} = \{\hat{q}_1, \dots, \hat{q}_t\}$ と観測された音長の時系列 $X = \{x_1, \dots, x_N\}$ から、テンポの時系列 $T = \{\tau_1, \dots, \tau_N\}$ を計算できる。ここでは、式 (1) から音長 x_k の区間 $[t_{k-1}, t_k]$ に対応するテンポを $\tau_k = x_k / \hat{q}_k$ とする。これを「観測されたテンポ」と呼ぶことにする。

各時間間隔 $[t_0, t_1], [t_1, t_2], \dots, [t_{N-1}, t_N]$ の演奏者の意図したテンポが $T' = \{\tau'_1, \dots, \tau'_N\}$ であり、 τ_1, \dots, τ_N が観測される確率は、継続時間 $\Delta t_k = t_k - t_{k-1}$ を考慮して

$$P(T|T') \approx \prod_{k=1}^N p(\tau_k | \tau'_k)^{\frac{\Delta t_k}{T}} \quad (3)$$

で表される。観測されたテンポに対して最も尤もらしいテンポを推定するために、式 (3) を最大にする。ここでは右辺の対数を目的関数とし、テンポ推定を

$$\hat{T} = \underset{T}{\operatorname{argmax}} \sum_{k=1}^N \log P(\tau_k | \tau'_k)^{\frac{\Delta t_k}{T}} \quad (4)$$

と定式化する。

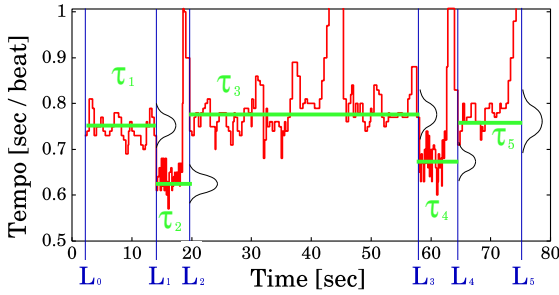
3 テンポ変化時点の検出

3.1 複数のテンポを持つ楽曲

曲の途中でテンポを変える楽曲がある。例えば、ピアノ曲「Fürchtenmachen (こわいぞ)」[†] は、A(1-8) –

*“Detection of Tempo Changes in Performed MIDI Music Using Probabilistic Models” by Haruto TAKEDA, Takuya NISHIMOTO, and Shigeki SAGAYAMA (The University of Tokyo).

[†]R. Schumann 作曲 Kinderszenen (子供の情景) Op. 15, no. 11.



$$F = F[L_0, L_1] + F[L_1, L_2] + F[L_2, L_3] + F[L_3, L_4] + F[L_4, L_5]$$

図 1: 複数のテンポを持つピアノ曲の実演奏のテンポ

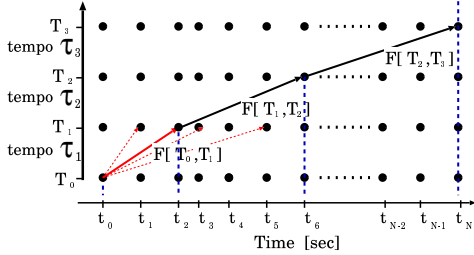


図 2: 二段 DP を用いたテンポ変化時点の推定

B(9-12) – A(13-20) – C(21-28) – A(29-36) – B(37-40) – A(41-48)*のような 3 部形式であるが、B の部分は “Schneller(早く)” 演奏するよう楽譜に指示されている。この曲の実演奏から抽出されたテンポを観測すると、図 1 に見られるように、実際に図中の区間 B でテンポ τ [秒/拍] の値が小さくなっている。リズムとテンポの分解のためには、このようなテンポ変化も正しく検出しなければならない。

3.2 テンポ変化時点の検出方法

このように区間によってテンポが切り替わる曲について、テンポの変化時刻を推定する手法について述べる。Segmental k -means 法のような区間分割クラスタリングの効率的なアルゴリズムを用いることもできるが、初期値の与え方によっては極値に収束する恐れがある。そこで、ここでは以下のように 2 段 DP アルゴリズムを用いて最適解を求める。

時間区間 $[L_{m-1}, L_m]$ で意図されたテンポが τ_m であるときのこの区間の目的関数への寄与は式 (4) より、

$$F[L_{m-1}, L_m] = \sum_{t_k \in [L_{m-1}, L_m]} \log P(\tau_k | \tau_m) \frac{\Delta t_k}{T} \quad (5)$$

である。区間の数を M とし、各区間を区分する時刻を L_m で表す (L_0 は演奏開始時刻 t_0 、 L_M は終了時刻 t_N) とき、テンポ変化時点を推定するには、

$$\{\hat{L}_1, \dots, \hat{L}_{M-1}\} = \operatorname{argmax}_{L_1, \dots, L_{M-1}} \sum_{m=1}^M F[L_{m-1}, L_m] \quad (6)$$

を解けば良い。各区間のテンポ τ_m は式 (5) の計算を通して連続値として求まる。式 (6) の計算は、図 2 に示すように二段 DP(Dynamic Programming) アルゴリズムを用いることができる。この方法は計算量は多いが、最適解を求めることができる。なお、One-pass DP 法の適用は検討中である。

3.3 評価実験

ピアノ奏者 2 名 (A,B) による “Fürchtenmachen” の演奏に対して、提案したテンポ変化時点の検出方

* () は小節番号を表す。

表 1: 実験条件: HMM とテンポ確率のパラメータ

音価の種類	19
HMM 状態数	6859
n -gram の次数	4 (リズムベクトルの 3 次元)
遷移確率の学習 (楽譜データ)	13 曲のクラシックピアノ曲の 4355 個の音価
出力確率の学習 (演奏・楽譜データ)	2 曲のピアノ演奏 (2 名、6 演奏) 1288 個の音価
評価データ	1 曲ピアノ演奏 (2 名 2 回)
テンポ変動確率分布	分散: 0.01(固定)

表 2: 実演奏のテンポの切り替わる区間の検出率 [単位 %]

演奏データ	A1	A2	B1	B2
検出率 1	66.2	51.6	25.6	51.6
検出率 2	24.2	11.6	30.0	20.7
F 値	35.4	18.9	27.6	29.2

法の評価実験を行った。異なるテンポの区間の数として $M = 5$ を与え、テンポが早くなる区間 B の検出を試みた。テンポの確率分布 $p(\tau|\tau')$ には、補正した[†]正規分布 $N(\tau|\tau', \sigma^2)$ を用いた。評価は次式で定める F 値で評価した。

$$\text{再現率 } R = \frac{\text{正しく検出した区間 } B \text{ の長さの合計}}{\text{正解の区間 } B \text{ の長さ}}$$

$$\text{適合率 } P = \frac{\text{正しく検出した区間 } B \text{ の長さの合計}}{\text{検出した区間 } B \text{ の長さ}}$$

$$F \text{ 値} = \frac{2PR}{P + R}$$

検出結果を表 2 に示す。各区間では厳密にはテンポは一定でなく、特にテンポ変化時点の直前にテンポが遅くなる傾向がある。このため、テンポ変化時点厳密に正しく求めることはできていない。小節構造から得られるリズムの情報を用いることにより、より厳密なテンポ変化時点を求めることが可能であると思われる。

4 おわりに

本稿では、曲の途中でテンポが変わる場合の変化時点とテンポ推定の両方について、確率モデルの推定の観点からアルゴリズムを提案した。テンポ変化数の推定を含め、今後は、リズムとテンポを統合した数理モデルについて検討したい。

参考文献

- [1] S. Dixon, “Automatic Extraction of Tempo and Beat from Expressive Performances,” *Journal of New Music Research*, 30, 1, 2001, pp 39-58.
- [2] A. Cemgil, P. Desain, B. Kappen, “Rhythm Quantization for Transcription,” *CMJ*, pp60-76, 1999.
- [3] C. Raphael, “Automated Rhythm Transcription,” *Proc. of ISMIR2001*, pp. 99-107.
- [4] 武田, 篠田, 嵯峨山, “リズムベクトルの概念に基づくリズム認識,” *情処研報*, MUS-46, pp. 23-28, 2002.
- [5] 大槻, 齋藤, 中井, 下平, 嵯峨山: “隠れマルコフモデルによる音楽リズムの認識,” *情処論*, Vol. 43, No. 2, pp. 245-255, 2002.
- [6] H. Takeda, T. Otsuki, N. Saito, M. Nakai, H. Shimodaira, S. Sagayama, “Hidden Markov Model for Automatic Transcription of MIDI Signals,” *Proc. MMSP*, 2002.

[†] $|\tau - \mu| > 0.1$ のとき $p(\tau) = 10^{-7}$ とした。