

# HMM を用いたリズムとテンポの反復推定による多声音楽の MIDI 演奏のリズム認識\*

武田晴登, 西本卓也, 嵯峨山茂樹 (東大情報理工)

## 1 はじめに

MIDI 楽器や標準 MIDI ファイルの普及により, 人間の生の演奏を記録した MIDI を対象とした自動採譜や音楽検索に対する要求は多い. 本稿ではこのような音楽情報処理に必要とされる人間の演奏の MIDI 信号を対象としたリズム認識手法を議論する.

## 2 リズムとテンポの反復推定

### 2.1 音長とリズムとテンポの関係

音楽演奏における音の長さ  $x$  秒は, リズムとテンポによって定められる. テンポは通常 bpm (beats per minute) を単位で表記されるが, ここでは時間の単位を秒に統一して「秒/拍」を用いる. また, リズムは音符の持つ長さの情報である音価 (時価, time value)  $q$  により定量的に扱うことができ, ここでは単位に「拍」を用いる. 音価  $q$  の音がテンポ  $r$  で演奏されるとき音長  $x$  は

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

で与えられる. 従って, 実演奏の MIDI ファイルからは音長  $x$  のみが観測されるので, リズム認識は音長  $x$  に対応するリズム  $q$  を求める問題になる.

実演奏には, 演奏に表情付けるために演奏者が意図的にリズムやテンポを変動させ, かつ無意識にもランダムにも音長を変動させるので, 観測された音長は一般に意図した音長  $q \cdot r$  に変動が加わったものである. また, 式 (1) から音長に  $x$  に対応する音価  $q$  とテンポ  $r$  の組み合わせは複数存在する. 従って, 一般には  $x$  から  $q$  を一意に定めることはできず, 音長から適切なリズムとテンポを推定することは容易ではない.

### 2.2 リズムとテンポの事後確率最大化推定

我々はこれまでに, リズムパターンの出現や音長の変動を数理的にモデル化し, 不良設定問題であるリズム認識を確率的に最も確からしいものを求める推定問題として考え, 音声認識の手法を応用してきた [1]. 演奏  $X$  に対して最も確からしいリズム  $Q$  とテンポ  $R$  を求める問題は, 事後確率最大化による推定

$$\begin{aligned} \{\hat{Q}, \hat{R}\} &= \operatorname{argmax}_{Q,R} P(Q, R|X) \\ &= \operatorname{argmax}_{Q,R} P(X|Q, R)P(Q, R) \quad (2) \end{aligned}$$

と定式化される.

以前の研究 [1] ではテンポを離散的なテンポ値を Markov 的に遷移するネットワークモデルでモデル化したが, テンポは時間に対して区分的に連続性を仮定する方がテンポの特性を表現するには適していると思われる. 本稿では, 時間について区分的に連続なパラメトリックな関数によるテンポのモデル化を検討する.

### 2.3 リズムとテンポの反復推定

音楽経験者が演奏を聴いてテンポとリズムを理解できるのは, その人が時間的には緩やかに変化するテンポを感じつつ, そのテンポに基づいて音長を典型的なリズムパターンに当てはめながら聴くことができ, 同時に, ある典型的なリズムパターンからテンポを感じることもできるからであると考えられる. そこで, ここではテンポを仮定したときのリズム推

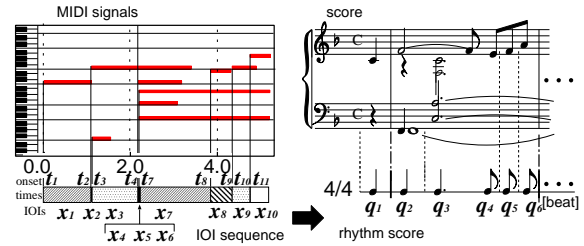


Fig. 1 多声音楽 MIDI 演奏のリズム推定

定と, あるリズムを仮定したときのテンポ推定を用いて, 式 (2) を解く手法を検討する.

ここで仮に, あるテンポ  $R$  が与えられたときにリズムを  $Q \rightarrow Q'$  と更進することで事後確率の増加

$$P(X|Q', R)P(Q', R) \geq P(X|Q, R)P(Q, R) \quad (3)$$

が可能であり, 更に, リズム  $Q'$  が与えられたときにテンポを  $R \rightarrow R'$  と更新することで事後確率の増加

$$P(X|Q', R')P(Q', R') \geq P(X|Q', R)P(Q', R) \quad (4)$$

が可能であるならば, 事後確率  $P(X|Q, R)P(Q, R)$  を単調増加させる更新の繰り返しにより, 事後確率の極大値に対応するリズムとテンポが得られる. この推定により大域的な最適解が得られる保証はないが, 収束は保証される. ここではリズムとテンポは独立であると仮定しテンポの事前知識  $P(R)$  を用いないとすると, 式 (3) は  $P(X|Q, R)P(Q)$  を増加させるリズム推定, 式 (4) は  $P(X|Q, R)$  を増加させるテンポ推定となる. 以下の章でこれらの事後確率を増加させる各ステップについて論じる.

## 3 リズムとテンポの確率モデル

### 3.1 テンポと音価に依存した音長の変動確率

リズム推定とテンポ推定の目的関数には, リズムとテンポに依存した音長の変動確率が共通項  $P(X|Q, R)$  として含まれている. 音価  $q$  とテンポ  $r$  に依存する音長  $x$  の確率密度関数  $p(x|q, r)$  を, 音価とテンポ各々について 2 倍と 1/2 倍の変化に対して等しい観測確率を与えるように以下の対数正規分布で定める.

$$\begin{aligned} p(x|q, r) &= \frac{1}{\sqrt{2\pi\sigma_q^2}x} \exp\left(-\frac{(\log x - \log q - \log r)^2}{2\sigma_q^2}\right) \quad (5) \end{aligned}$$

### 3.2 テンポを仮定したときのリズム推定

#### 3.2.1 多声音楽のリズム譜と多声部間 IOI

音声認識手法を用いたリズム認識手法 [1] を応用して, 多声音楽のリズムを推定する. 具体的には, Fig. 1 に示すように多声音楽の演奏の MIDI 信号で全ての声部の発音時刻から IOI (Inter-onset Interval, 発音時刻) を抽出し, これに対応する音価情報である多声音楽の全ての音の発音位置から得られるリズム譜を推定し, 単旋律のリズム推定と同型の問題に帰着させる.

#### 3.2.2 自己遷移を含む HMM を用いたリズム推定

$P(Q)$  はリズム  $Q$  の出現確率を表し, ここでは一小節単位のリズムを単位 (リズム単語 [1]) とした  $n$ -gram

\*Rhythm Recognition of Polyphonic MIDI performance by iterative Estimation of rhythm and Tempo. by Haruto Takeda, Takuya Nishimoto, Shigeki Sagayama (University of Tokyo)

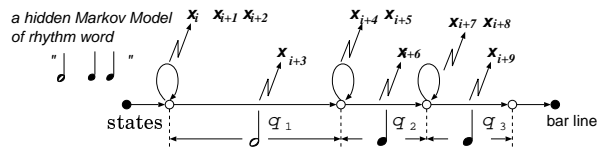


Fig. 2 リズム推定に用いる HMM

モデルを用いる。これは、既存の楽曲のリズムを用いて統計的に学習することが可能である。

また、 $P(X|Q, R)$  はテンポ  $R$  を仮定したときのリズム  $Q$  を音長  $X$  で演奏する確率を表す。ここでは先行研究の単旋律の HMM (Hidden Markov Model, 隠れマルコフモデル) を拡張し、Fig. 2 に示すように音価の境界である拍の位置を隠れ状態とし、多声音楽に含まれる同時に発音される音のほとんど 0 に近い IOI を出力する自己遷移を含む HMM を用いる。音価に対応する遷移出力の確率は式 (5) で与え、同時発音の IOI は 0 を中心とした正規分布に従うとする。

ここで用いた HMM と  $n$ -gram モデルを組み合わせたモデルにおける事後確率を最大化させる状態系列の探索問題は、音声認識では one-Pass DP (時間同期 Viterbi 探索) によって解かれている。これを利用して、与えられたテンポ  $R$  に対して  $P(X|Q, R)P(Q)$  を最大化するリズム  $Q$  が求まる。

### 3.3 リズムを仮定したときのテンポ推定

#### 3.3.1 瞬時テンポとテンポ曲線

音長  $x$  秒の音価が  $q$  拍であるとき、その音のテンポは「瞬時テンポ」 $x/q$  秒/拍として計算できるが、これは、音長の変動を無視しているため、実演奏では Fig. 3 に示すように一音ごとにランダムな変動を含み、テンポに備わっている特性である時間について区分的に連続で緩やかに変動する様子を表していない。そこで、我々はテンポを時間  $t$  について区分的に連続に変化する量として「テンポ曲線」 $\tilde{r}(t)$  としてモデル化する。

#### 3.3.2 瞬時テンポのテンポ曲線からの確率的変動

瞬時テンポのテンポ曲線からの変動は音長変動によって起こるものと考えられるので、ここでは音長の変動確率に基づいて瞬時テンポのテンポ曲線からの変動確率を与える。ここで式 (5) において  $\log x - \log q - \log \tilde{r}(t) = \log x/q - \log \tilde{r}(t)$  であることに注意すると、テンポ  $\tilde{r}(t)$  で音価が  $q$  であるときの音長の変動  $x$  の変動確率は、瞬時テンポ  $x/q$  がテンポ曲線のテンポ  $\tilde{r}(t)$  からの確率変動とも解釈可能である。従って、最大化する目的関数は、同時発音を除いて数えた IOI の番号を  $j$  として、

$$\log P(X|Q, R) = -\frac{1}{2} \sum_{j=1}^J \frac{(\log \frac{x_j}{q_j} - \log \tilde{r}(t_j))^2}{\sigma_j^2} + \text{定数}$$

で与えられる。従って、テンポ曲線の推定は分散に依存した確率的な距離尺度を用いた最小 2 乗誤差推定問題に帰着される。

#### 3.3.3 テンポ曲線のモデル化と推定

テンポ曲線には目的関数  $P(X|Q, R)$  の単調増加が可能である任意の関数を使用することができ、その一例として対数軸上での多項式  $\log \tilde{r}(t) = \sum_{d=0}^D a_d t^d$  を用いた場合は、テンポ曲線は正規方程式を解くことで最適解が得られる。また、区分的に切り替わるテンポのモデルとして、各区分で異なる多項式を定義するテンポ曲線でモデル化した場合、テンポ曲線は、segmental  $k$ -means 法を用いて初期値に依存する最適化が可能であり [2]、Fig. 3 に示すように、滑らかな時間変化を表す曲線が推定できる。以上より、リズム  $Q$  が与えられたときに  $P(X|Q, R)$  を増加させるテンポ  $R$  が推定できる。

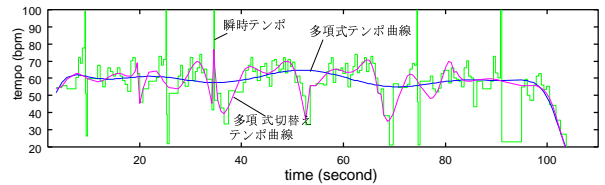


Fig. 3 テンポ曲線推定例 (演奏曲: Träumerei)

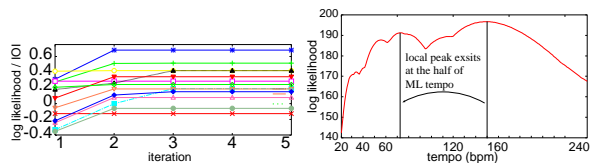


Fig. 4 反復推定の収束 (右) と局所最適解の存在

Table 1 リズム正解率 [単位: %]

	$T_1$	$T_2$	$T_3$
A	83.6	83.1	82.0
B	89.0	87.6	88.2

## 4 評価実験

### 4.1 実験条件

電子ピアノ演奏を記録した MIDI を用いて提案手法の性能評価を行った。評価データの演奏曲は、Burgmüller 作曲の練習曲 Op. 100 の 25 曲と Schumann の Op. 15 「子供の情景」から 7 曲とした。これらの楽曲は装飾音やテンポの変動を多く含む曲である。リズム推定では、A: 演奏曲を含まない 100 曲から学習して構築したモデル (601 個の HMM)、B: 演奏曲を含む 150 曲の統計から構築したモデル B (888 個の HMM) の 2 種類を用い、またテンポ曲線は、 $T_1$ : 定数関数、 $T_2$ : 6 次の多項式、 $T_3$ : 4 次の多項式の切替え 3 種類のテンポ曲線を用いて比較評価を行った。反復推定の初期条件として一定テンポ 120 bpm を与えた。

### 4.2 評価結果

リズム推定正解率は、楽譜と演奏 MIDI から推定した楽譜を音高に基づいてマッチングを取り、音価の置換と同時発音について削除・挿入誤りを数えて、音価もしくは同時発音について楽譜と一致した音の数の MIDI 演奏に含まれる全 IOI の数に対する割合で評価した。全ての音が倍テンポ、半分テンポの違いである場合も正解に数えた。

Table 1 に示す評価結果から、テンポが変動するピアノ演奏に対してリズムの推定が可能であることが示された。局所解への収束による誤推定と考えられるものとして、局所的に音価を正解の有理数倍と推定するテンポの誤推定に対応する誤りが見られた。また、区分的な多項式のテンポ曲線の区分の境界は *rit.* から *a tempo* のようなテンポの切り替えの位置とは対応せず、推定したテンポ曲線は必ずしも演奏の意図と一致したものにはならなかった。

尚、Fig. 4 左に示すようにテンポとリズムの反復推定は数回でほぼ収束した。また、一定テンポでの事後確率を計算すると Fig. 4 右に示すように複数の極大値が存在し、テンポについて複数の局所最適解が存在することが分る。

## 5 おわりに

本稿では人間の演奏者による MIDI 演奏のリズム認識について、テンポ推定とリズム推定の反復によりリズムとテンポの事後確率最大化問題の局所最適解を求める手法を議論した。今後は音楽認知の知見を利用しつつ、発音時刻以外の情報も含めた音楽演奏の数値モデル化の検討を行いたい。

## 参考文献

- [1] 齋藤ら, 情処研報, 99-MUS-33, pp.27-32, 1999.
- [2] Takeda et al., Proc. ISMIR, pp.357-364, 2004.