

リズムベクトルを用いた MIDI 演奏データからのテンポの変動の推定

武田晴登 西本卓也 嵯峨山茂樹

東京大学大学院情報理工学系研究科

{takeda,nishi,sagayama}@hil.t.u-tokyo.ac.jp

あらまし 音楽演奏の MIDI データからテンポを推定する方法について述べる。一般に、演奏の音長は、テンポとリズムから構成されると考えられる。我々は今まで確率モデルを用いて演奏からリズムを推定する手法を提案してきたが、本稿ではリズムだけでなくテンポについても確率を導入し、リズムとテンポの両方を扱う総合的なリズム認識手法を提案する。テンポについての確率モデルを導入する利点は、テンポ誤りを確率に含むことでリズム認識の精度を向上すること、演奏者の意図したテンポの変化を推定できることの 2 点にある。テンポ違いの確率を扱うために有理数倍のテンポ違いを考慮して混合分布を用いて EM アルゴリズムを適用する。さらに、途中から演奏者の意図したテンポが変化する時点を推定するために DP マッチングを行う。実際に人間の演奏に対して実験を行い、提案手法の有効性を示す。

Estimation of Tempo Variations in Performed MIDI Data Signals using Rhythm Vectors

Haruto TAKEDA Takuya NISHIMOTO Shigeki SAGAYAMA

Graduate School of Information Science and Technology, The University of Tokyo

Abstract A method to estimate the tempo in performed MIDI data is described in this paper. In general, note length in musical performance consists of rhythm and tempo. We have proposed methods to estimate rhythm from musical performance data using probabilistic models. The advantage of introducing probabilistic tempo model is the improvement of rhythm recognition accuracy with the probability of misrecognizing rhythm, and making it possible to estimate varying tempo intended by performers. We introduce probabilistic model for tempo as well as rhythm in this paper. To estimate tempo intended by performers, EM (Expectation / Maximization) algorithm are used for mixture probability for different tempos. Moreover, DP (Dynamic Programming) are used for estimation of time point of changing tempo. Experimental results are also discussed in this paper.

1 はじめに

我々は、人間の演奏（以下、実演奏）を記録した MIDI (Musical Instrument Digital Interface) 信号からの自動採譜について研究を行っている [1]。本稿では、自動採譜の要素技術である実演奏からテンポとリズムを推定する手法について論じる。

実演奏の MIDI 信号からリズムを得るためには「量子化」(quantization) と呼ばれる手法が広く用いられているが、この手法は基本的にテンポが既知で一定に保たれた MIDI 信号を対象としている。テンポが既知でなく、また、一定に保たれていない演奏の MIDI 信号に対しては、演奏曲の拍子と発音位置の候補を事前に与えてリズム認識を行う方法が提案されている。Cemgil ら [2] は、テンポを隠れ変数としたカルマンフィルタを用いて、ポピュラー音楽のピアノ演奏の MIDI 信号に対してテンポ推定が可能であることを示している。また、Raphael [3] も確率モデルを用いてリズムとテンポを推定する手法を提案している。これに対して我々は、演奏曲のテンポや拍子についての知識を用いずにリズム認識を行う手法を提案した [4, 5, 6, 7]。この研究では、リズム認識と音声

認識を同型の推定問題として捉え、連続音声認識で現在一般的に用いられている HMM (Hidden Markov Model, 隠れマルコフモデル) [8] を用いてモデル化を行いリズム認識を事後確率最大化問題として解いた。

さて、これらの従来のリズム認識研究では、テンポに関しては局所的に変動するような振舞しか扱われていない。本稿では、さらに演奏を通してのテンポの変動全体を扱う。特に、テンポが演奏の途中で切り替わる場合には、従来の様に局所的には変動を伴いつつも一定であるとしたテンポの扱いだけでは対処できない。複数のテンポを持つ演奏では、局所的なテンポの値だけでなく演奏を通してのテンポの変動過程を扱う必要がある。

本稿では、リズムに加えてテンポについても確率モデルを用い、リズムとテンポを対等な関係で扱うリズム認識手法を提案する。本稿で提案するテンポについてのモデル化を行う利点は、(1) 推定したテンポの情報を用いてリズム再推定によりリズム認識が精度向上できること、(2) 演奏中に変動する複数のテンポが推定できること、にある。以下、2 章でリズム認識の定式化を行い、3 章でリズムの推定方法を述べる。そして、4 章、5 章でテンポ推定方法を提案する。

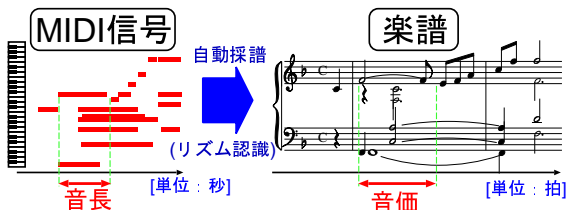


図 1: リズム推定: 実演奏の音長から楽譜の「音価」を推定

2 音長とテンポとリズムの関係

2.1 音価、音長、テンポの関係

本稿では、楽譜上の音符の正規の長さを「音価」(time value; 時価ともいう)と呼ぶ。音価は、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(単位は「拍」)として扱うことができる。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。

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。これは、「秒」を単位とする連続的な量である。音長 x 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 (IOI, inter-onset interval) により定義する。たとえば同一音符のスタッカート演奏とレガート演奏では、音符の発音時間自体は異なるが、次の音符までの時間間隔は同一の音価を反映した長さになる。

音長 x [秒] は音価 q [拍] と演奏の音価あたりの時間 τ [秒/拍] に依存し、それらの関係は

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

である。以後、本稿の用語として τ をテンポと呼ぶことにするが、メトロノーム表記のテンポ(毎分の拍数)とは反比例の関係がある。我々の目的は、図 1 に示すように実演奏で観測されたそれぞれの音の音長 x の系列から、音価 q の系列、すなわちリズムに適切に変換することである。これをここでは「リズム認識」と呼ぶことにする。

リズム認識は、式 (1) において与えられた音長 x の 2 変数 τ, q の積への分解と捉えられる。一般にこの分解は一意でない。例えば、等時間間隔で手拍子を打ったときのリズムは、全ての音を 4 分音符としてもよいし、8 分音符としてもよい。すなわち、音価 q を半分にしてテンポ τ を倍に(遅く)すれば、同一の演奏を異なる音価列で表現できる。しかし、複付点 8 分音符の列と解釈するのは常識的ではない。このような音価とテンポの関係は、原理的に楽譜の音価列を決定できない不確実性があるが、楽譜の常識に相当する音価列(リズム)の確率を導入することにより解決できる部分もある。

実際の演奏では、音楽的な演奏意図やランダムな変動やその他の要因により、個々の音符は、その音価に対応する長さから変動した音長で演奏される。音長の変動は、式 (1) におけるリズム q の変動とテンポの変動とテンポ τ の変動に対応する。音符ごとにテンポ τ が自由に変化し得るとすると音価 q の意味がないので、テンポ τ は少なくとも複数の音符に跨って変動する項、あるいは局所的には一定とみなせる項と考えることにする。

2.2 リズムとテンポの推定

ここで、確率モデルを用いたリズム・テンポの推定の原理について説明する。演奏者がテンポ T でリズム Q を演奏しようとして意図して音長 X を演奏したとする。リズム認識は、観測される X から演奏者の意図した T, Q を推定する問題である。これは、観測された X に対して最も尤もらしい T, Q を

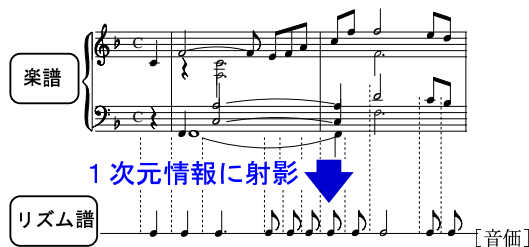


図 2: 楽譜の 1 次元情報への射影: リズム譜

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

によって推定する問題として扱えるが、しかし、最適な \hat{Q}, \hat{T} を同時に推定するのは難しい。そこで、本稿ではテンポとリズムをそれぞれ別の確率モデルで扱い、テンポのついでに事前知識 $P(T)$ は持たないことと、テンポ T とリズム Q が独立であることを仮定すると、式 2 は、それぞれの場合に以下のように 2 段階の(局所)最適化により達成できる。

1. テンポ T が未知である場合: リズム Q を推定

$$\hat{Q} = \underset{Q}{\operatorname{argmax}} P(X|Q)P(Q) \quad (3)$$

2. リズム \hat{Q} が仮定されている場合: テンポ T を推定

$$\hat{T} = \underset{T}{\operatorname{argmax}} P(X|\hat{Q}, T)P(\hat{T}) \quad (4)$$

3. テンポ \hat{T} が仮定されている場合: Q を再推定

$$\hat{Q} = \underset{Q}{\operatorname{argmax}} P(X|Q, \hat{T})P(Q) \quad (5)$$

式 (3)、式 (5) で示されるリズム推定で使用する確率モデルを 3 章で述べ、式 (4) によるテンポ推定で使用する確率モデルを 4 章、5 章で述べる。

3 リズムの推定

3.1 リズム譜と多声部間の IOI

多声音楽のリズムを扱うために、リズム譜を導入する。楽譜に記されている全ての音の発音位置に注目し、図 2 に示すようにそれらの隣り合う発音位置の間隔に対応する音価を考える。この音価の時系列をその曲のリズム譜と呼ぶことにする。

楽曲のリズム譜に見られるリズムパターンの出現の統計的な性質を利用するためにリズム譜に現れるリズムに確率モデルを導入する。楽曲のリズムパターンの統計には、フレーズのくり返しなどの大域的な特徴と、フレーズ中のリズムを構成する局所的な特徴があると考えられる。今回は、局所的な特徴を扱うために、連続する n 個の間の確率的依存性を与えるモデルとして音価の n -gram を考える。音価の出現確率は、直前の $n-1$ 個の音価の履歴に依存する条件付確率 $P(q_t|q_{t-1}, \dots, q_{t-n+1})$ で近似できるとすると、リズム譜 $Q = \{q_1, \dots, q_N\}$ の出現確率は

$$\begin{aligned} P(Q) &= P(q_1, \dots, q_N) \\ &\approx P(q_1, \dots, q_{n-1}) \prod_{t=n}^N P(q_t|q_{t-1}, \dots, q_{t-n+1}) \quad (6) \end{aligned}$$

により近似できる。履歴に依存する各音価の出現確率値は、既存の楽曲のリズム譜から統計的な学習を行うことで適切な値を定められる。

一方、演奏の情報の中でリズム譜に対応するものが、多声部間 IOI (inter-onset interval, 発音時刻の間隔) である。多声部間 IOI は、複数の旋律(声部)が同時に演奏される場合でも演奏楽曲の多声部の構造を考慮せずに、全ての音の発音時刻の間隔をとったものである。2 つの音が同時に発音される場合は、IOI が 0 になる。

3.2 リズムベクトル

テンポが未知である実演奏のリズム推定を行う場合は、テンポに依存しない特徴量としてリズムベクトルを用いる。式 (1) よりテンポ τ が局所的に一定と見なせれば、 τ の値に無関係に音長 x の比は音価 q の比を反映する。そこで、 n 個の連続する音長 x_t, \dots, x_{t+n-1} の比を成分とするベクトルをリズムベクトルと呼ぶ。成分の和が 1 になるように規格化するため、リズムベクトル $r_t = (r_t^1, \dots, r_t^n)$ の第 i 成分を

$$r_t^i = \frac{x_{t+i}}{x_t + \dots + x_{t+n-1}} \quad (7)$$

と定義する。

実演奏のリズムベクトル r は、 τ が一定で $\epsilon = 0$ である「理想的な」リズムベクトルに対して変動する。 r の変動は確率分布に従うと仮定する。音価の n 個の組 $s_t = \{q_t, \dots, q_{t+n-1}\}$ で表されるリズムをリズムベクトル r_t で演奏する確率を、 $b_{s_t}(r)$ で表すことにする。 b_{s_t} はリズム s_t に対応する理想的なリズムベクトルを平均とし、分散が Σ である正規確率分布を用いる。 Σ は、実演奏データを用いた事前学習により求める。

3.3 局所的テンポの変動

さらに、テンポについても確率を導入する。音長 n 個の組の平均のテンポ (以後、局所テンポと呼ぶ)

$$\bar{\tau}_t = \frac{x_t + \dots + x_{t+n-1}}{q_t + \dots + q_{t+n-1}} \quad (8)$$

を求める。

テンポの変動が小さいことを仮定して、局所テンポ $\bar{\tau}_t$ の変動 $\Delta\bar{\tau}_t = \bar{\tau}_t - \bar{\tau}_{t-1}$ は、0 を平均としたが確率分布 $\phi(\tau_t)$ に従うとし、本稿では正規分布を用いる。

また、演奏者の意図したテンポ τ'_t が与えられているときは、実演奏と意図とのテンポについての差 $\bar{\tau}_t - \tau'_t$ が、0 を平均としたが確率分布 $\phi(\tau_t)$ に従うとし、本稿では正規分布を用いる。

以上より、演奏者がリズム譜が Q であるような楽曲を演奏したとき、演奏の音長系列が X である確率は次のように書ける。 X から得られるリズムベクトルの時系列を $\{r_t\}$ 、 X と Q から得られる局所テンポの変動の時系列を $\Delta\bar{\tau}_t$ として、

$$P(X|Q) = b_{s_1}(r_1) \cdot \prod_{t=2}^{N-n+1} b_{s_t}(r_t) \cdot \phi(\Delta\bar{\tau}_t) \quad (9)$$

である。更に、演奏者がテンポ $T = \{\tau'_1, \dots, \tau'_T\}$ でリズム譜 Q を演奏する確率は

$$P(X|Q, T) = b_{s_1}(r_1) \cdot \prod_{t=2}^{N-n+1} b_{s_t}(r_t) \cdot \phi(\bar{\tau}_t - \tau'_t) \quad (10)$$

である。

3.4 リズムベクトルを用いた HMM

n 次元のリズムベクトル r には音価 n 個の組 s と 1 対 1 に対応するので、音価 n 個の組を状態 s とし r を状態 s からの出力とするモデルを考える。この状態の遷移のマルコフ過程であるとする、状態遷移確率は 3.1 で述べた音符の $(n+1)$ -gram との遷移確率に対応する。リズムベクトルの変動の確率 $P(X|Q)$ と音符 n -gram による確率 $P(Q)$ を組み合わせた確率モデルは、図 3 に示されるようにリズムベクトルを出力とする HMM と仮えられる。この HMM により、ある音価系列 Q の仮説に対して音長の時系列 X が観測される確率 $P(X|Q)P(Q)$ または $P(X|Q, T)P(Q)$ を計算できる。

式 (3), (5) におけるリズム Q の推定問題は、HMM における最尤状態系列の探索問題となる。これは、効率的

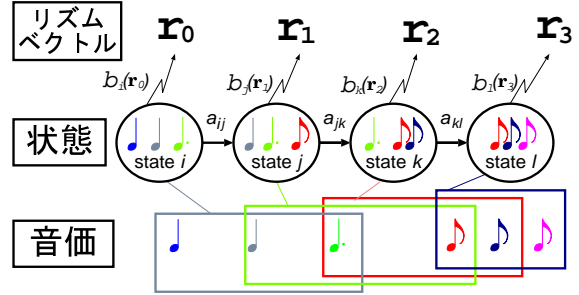


図 3: リズムベクトルを出力とし音符の組を状態とする HMM

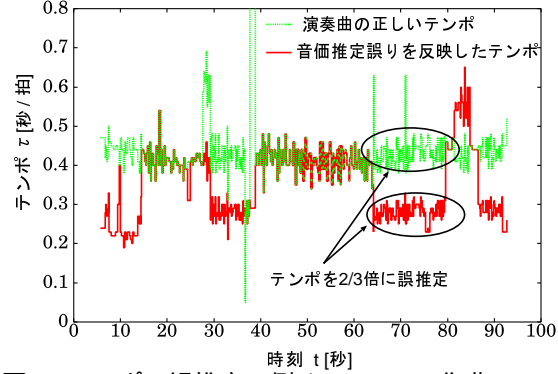


図 4: テンポの誤推定の例 (Beethoven 作曲 Sonata No.20 の前半の演奏)

な探索アルゴリズムである VDA[9] (Viterbi Decoding Algorithm: ビタビ復号化アルゴリズム) を利用して解くことができる。

4 一定テンポの推定

4.1 リズム譜から得られるテンポ

推定された音価の時系列 $Q = \{q_1, \dots, q_t\}$ と観測された音長の時系列 $X = \{x_1, \dots, x_N\}$ から、音価の変動がなかったと仮定して式 (1) から $\tau_t = x_t/q_t$ としてテンポの時系列 $\{\tau_1, \dots, \tau_N\}$ を計算できる。演奏者がテンポの変動を意図していなかった場合、テンポは小さな変動はあるが曲全体では一定と見なせる。ここでは、演奏者の意図した曲全体のテンポは一定であると仮定し、このテンポを推定する。

リズムベクトル (音長の比) を用いて音価推定を行っているため、一度音価の推定誤りを犯すと後続する音の音価は、音価の比を正しく推定しても誤りになる。これは、本来ならテンポがほとんど一定であるはずなので、音価の推定誤りを続けることは図 4 のように推定した音価から計算されるテンポの変動として観測できる。

4.2 確率モデルを用いたテンポの推定

演奏者の意図したテンポとは異なる倍テンポ違いに推定された可能性を含めるために、確率的距離尺度を導入する。観測されたテンポが τ であり、演奏者の意図したテンポが τ' であると確率を倍テンポ誤りの可能性を考慮するため、混合正規分布

$$P(\tau|\tau') = \sum_{c=0}^5 w_c p_c(\tau) \quad (11)$$

で与える。ただし、 $p_c(\tau)$ は平均 μ_c 、分散 σ_c の正規分布の確率密度関数

$$p_c(\tau) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left\{-\frac{1}{2} \frac{(\log \tau - \log \mu_c)^2}{\sigma_c^2}\right\} \quad (12)$$

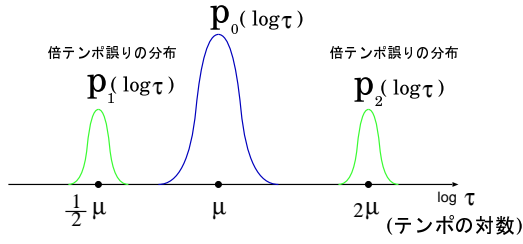


図 5: 対数軸上のテンポの混合正規分布

を表し、平均 $\mu_c = \alpha_c \mu$ は $\alpha_0 = 1, \alpha_1 = 2, \alpha_2 = \frac{1}{2}, \alpha_3 = \frac{3}{2}, \alpha_4 = \frac{2}{3}$ により、がテンポ誤りであるときの確率分布の中心を表す (図 5)。それぞれ時間間隔 $[t_0, t_1], [t_1, t_2], \dots, [t_{N-1}, t_N]$ の間のテンポとして τ_1, \dots, τ_T が観測される確率は、各テンポの観測される継続時間 $\Delta t_k = t_k - t_{k-1}$ を考慮して

$$P(Q, X|T) = \prod_{k=1}^T p(\tau_k | \tau')^{\frac{\Delta t_k}{T}} \quad (13)$$

で表される。この対数を取ると、

$$F = \sum_{k=1}^N \log P(\tau_k | \tau') \frac{\Delta t_k}{T} \quad (14)$$

である¹。推定したいテンポは、与えられたテンポの時系列に対して最も尤もらしいものであるので、

$$\hat{T} = \operatorname{argmax}_T P(X, Q|T) = \operatorname{argmax}_T F \quad (15)$$

ここでは具体的に

$$\hat{\tau} = \operatorname{argmax}_{\tau} \sum_{k=1}^N \log P(\tau_k | \tau') \frac{\Delta t_k}{T} \quad (16)$$

となる。

4.3 EM アルゴリズムとしての定式化

式 (16) により、未知変数である正規分布のパラメータ $\theta = (\mu, \sigma_c)$ とその重み w_c を直接計算することはできない。何故なら、確率値 $P(\tau_k | \mu, \sigma_c)$ の計算するのに必要である、テンポ τ_k がどの正規分布に属するかという情報が与えられていないからである。そこで、このような不完全データからの目的関数を最大にするパラメータ推定を行うために、EM アルゴリズムを用いる。定式化は、以下に示すようにまず、4.3.1 で目的関数の増分を Q 関数で評価し、4.3.2 で Q 関数を最大にするパラメータ更新の式を導く。

4.3.1 パラメータ更新時の目的関数の増分

はじめに、パラメータ $\theta = (\mu, \sigma_c, w_c)$ を $\theta = (\mu', \sigma'_c, w'_c)$ に更新したときの目的関数の増分

$$\Delta F = \sum_{k=1}^N \log P(\tau_k | \theta') \frac{\Delta t_k}{T} - \sum_{k=1}^N \log P(\tau_k | \theta) \frac{\Delta t_k}{T} \quad (17)$$

を計算する。テンポの正規分布への所属についての ΔF の期待値を計算しても ΔF の値は変わらないので、

$$\begin{aligned} \Delta F &= \sum_{k=1}^N \log \frac{P(\tau_k | \theta')}{P(\tau_k | \theta)} \frac{\Delta t_k}{T} \\ &= \sum_{k=1}^N \sum_{c=0}^4 p(c | \tau_k, \theta) \cdot \log \frac{P(\tau_k | \theta')}{P(\tau_k | \theta)} \end{aligned}$$

¹テンポの観測時間を細かくし連続への極限を取ると $F = \int_0^T \log P(\tau(t) | \tau'(t)) dt$ を得る。以降の全ての $\sum_{k=1}^N \dots \frac{\Delta t_k}{T}$ は、観測時刻 t_k の関数 $g(t_k)$ を用いることにより $\int g(t) dt$ の形で書くことができる。

ここで、テンポ τ_k が正規分布 c に所属する確率 $p(c | \tau_k, \theta)$ は

$$\frac{p(\tau_k, c | \theta)}{p(\tau_k | \theta)} = \frac{w_c p_c(\tau_k)}{\sum_{c=0}^4 w_c p_c(\tau_k)} \quad (18)$$

で与えられることに注意する。さらに、同時確率についての関係式 $p(\tau_k | \theta) = \frac{p(\tau_k, c | \theta)}{p(c | \tau_k, \theta)}$ を用いると

$$\begin{aligned} \Delta F &= \sum_{k=1}^N \sum_{c=0}^4 p(c | \tau_k, \theta) \log \left[\frac{P(\tau_k, c | \theta')}{P(\tau_k, c | \theta)} \right] \cdot \frac{\Delta t_k}{T} \\ &+ \sum_{k=1}^N \sum_{c=0}^4 p(c | \tau_k, \theta) \log \left[\frac{P(c | \tau_k, \theta')}{P(c | \tau_k, \theta)} \right] \cdot \frac{\Delta t_k}{T} \end{aligned}$$

となる。ここで、 $p(\tau_k, c | \theta) = w_c p_c(\tau_k)$ である。ここで、Jensen の不等式より

$$\sum_{c=0}^4 p(c | \tau_k, \theta) \log \left[\frac{P(c | \tau_k, \theta')}{P(c | \tau_k, \theta)} \right] \geq 0 \quad (19)$$

が成立するので、第 2 項が正であることは保証される。ここで、 Q 関数

$$Q(\theta, \theta') = \sum_{k=1}^N \sum_{c=0}^4 p(c | \tau_k, \theta) \log p(\tau_k, c | \theta') \cdot \frac{\Delta t_k}{T} \quad (20)$$

を導入すると、

$$\Delta F \geq Q(\theta, \theta') - Q(\theta, \theta) \quad (21)$$

を得る。従って、パラメータ更新により目的関数 D を増加させるには $Q(\theta, \theta')$ 関数を増加させるように θ' を推定すれば良い。このパラメータ更新を目的関数 F が収束するまで繰り返す。(収束判定には $\Delta F \leq \epsilon_d$ を用いる。)

4.3.2 Q 関数の最大化

$Q(\theta, \theta')$ を最大にする $\theta = (\mu, \sigma_c, w_c)$ ($c = 0, 1, \dots, 5$) を、 $\sum_{c=0}^4 w_c = 1$ という条件付で求める。Lagrange の未定乗数法により、未定乗数 γ を導入して、

$$L = Q(\theta, \theta') - \gamma \left[1 - \sum_{c=0}^4 w_c \right]$$

を最大にする θ' を求めれば良い。各変数で偏微分が 0 になる点として極値を求め、式 (18) を用いることにより次式に示す推定式を得る。

$$\hat{w}_c = \frac{w_c \sum_{k=1}^N p_c(\tau_k) \Delta t_k}{\sum_{c=0}^4 w_c \sum_{k=1}^N p_c(\tau_k) \Delta t_k} \quad (22)$$

$$\hat{\mu} = \frac{\sum_{k=1}^N \sum_{c=0}^4 \left(\frac{\alpha_c w_c p_c(\tau_k)}{(\sigma_c)^2} \right) \cdot \tau_k \cdot \Delta t_k}{\sum_{k=1}^N \sum_{c=0}^4 \left(\frac{\alpha_c^2 w_c p_c(\tau_k)}{(\sigma_c)^2} \right) \cdot \Delta t_k} \quad (23)$$

$$\hat{\sigma}_c^2 = \frac{\sum_{k=1}^N (\tau_k - \hat{\mu}_c)^2 p_c(\tau_k) \Delta t_k}{\sum_{k=1}^N p_c(\tau_k) \Delta t_k} \quad (24)$$

表 1: 電子ピアノによる演奏からの楽譜復元の正解率 [単位 : %]

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	3 曲ピアノ演奏 (5 名 2 回)

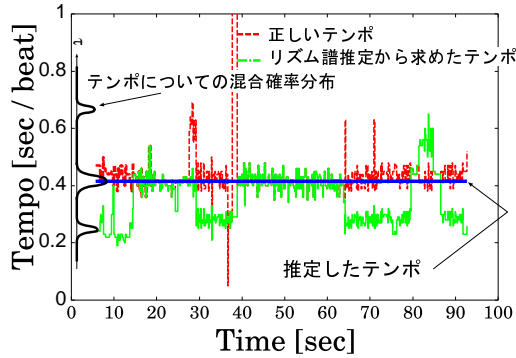


図 6: 混合確率分布を用いた一定テンポの推定例

4.4 一定テンポ推定の手順

EM アルゴリズムを用いてテンポを推定するには、以下のような手順で行う。

1. 推定した音価の時系列 $Q = \{q_1, \dots, q_N\}$ と観測した IOI の時系列 $X = \{x_1, \dots, x_N\}$ からテンポの時系列 $\{\tau_1, \dots, \tau_N\}$ を求める。
2. EM アルゴリズムの初期値 $\theta = (\mu, \sigma_c)$ を定める。テンポの平均 μ には、演奏曲を通しての平均のテンポ $\bar{\tau}$ を用いる。

$$\bar{\tau} = \frac{\sum_{k=1}^N \tau_k \Delta t_k}{\sum_{k=1}^N \Delta t_k} \quad (25)$$

3. 式 (22)、式 (23)、式 (24) によるパラメータ更新を目的関数の値が収束するまで行う。収束判定は、ある正数 ϵ_F を定め、 $\Delta F < \epsilon_F$ により行う。

4.5 評価実験

提案手法を用いてテンポ推定、および推定テンポを用いてリズムを推定を行う評価実験を行った。実験条件を表 1 に示す。実験結果を表 2 に示す。EM アルゴリズムでは、分散の再推定は行わず $\sigma_c = 0.01$ と固定し、また収束の判定基準は $\epsilon_F = 10^{-5}$ を用いた。

テンポ推定前のリズム \hat{Q} の正解率とテンポ推定後のリズム \hat{Q} の正解率、および推定したテンポの値を表 2 に示す。テンポが正しく推定できれば、高いリズム認識率を得られている。テンポの誤推定は、正しい値の $\frac{2}{3}$ の値と推定した場合が多く、これは 3 連符の音価を 8 分音符の音価と誤認識したものに对应する。

5 複数テンポの推定

5.1 複数のテンポを持つ楽曲

本章では、テンポ変化時点の検出方法について述べる。

2.1 で述べたようにテンポは局所的に一定と見なせる項として考えてきた。しかし、音楽演奏においてはある時点からテンポが切替わり、それまでとは別のテンポで演奏を続けることがある。このようなテンポの変動は、

表 2: 推定テンポを用いたときのリズム正解率 [%]

演奏データ	テンポの値		リズム正解率		
	実際	推定	同時発音推定前	テンポの処理	テンポ推定後
A、1 回目	0.453	0.416	99.4	43.8	92.2*
A、2 回目	0.453	0.405	97.8	42.3	86.3*
B、1 回目	0.467	0.277	96.9	38.6	47.3
B、2 回目	0.470	0.440	97.1	42.3	78.0*
C、1 回目	0.486	0.314	97.2	45.1	47.3
C、2 回目	0.468	0.360	97.3	41.6	42.3
D、1 回目	0.453	0.417	97.8	42.5	63.2*
D、2 回目	0.455	0.349	97.8	43.3	46.4
E、1 回目	0.477	0.446	97.2	43.8	73.0*
E、2 回目	0.476	0.304	97.9	32.5	47.9
平均	0.466	0.372	97.5	41.6	60.4(78.5*)

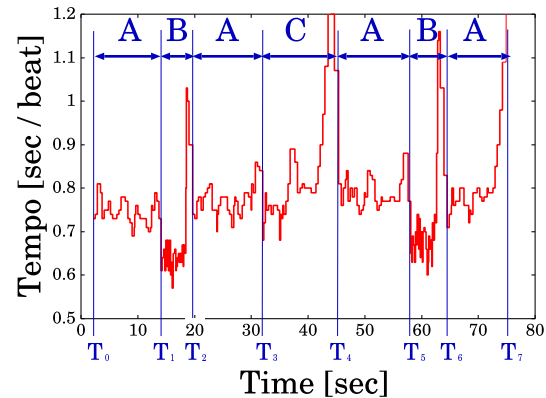


図 7: テンポ変化の指示を含む楽曲の実演奏のテンポの推定例

従来のリズム認識手法でテンポが局所的に一定であるとした単純なモデルでは扱えない問題である。そこで、本稿では曲を通してのテンポ変動全体に注目し、テンポ切替の検出を行う手法を提案する。

例としてピアノ曲 “Fürchtenmachen” (“怖いお話”)² を選んだ。この曲は、A(1-8) - B(9-12) - A(13-20) - C(21-28) - A(29-36) - B(37-40) - A(41-48)³ という 3 部構成になっており、譜面に A は遅いテンポで、B の部分は “Schneller” (早く) 演奏するよう指示されている。この曲の実演奏から抽出されたテンポを図 7 に示す。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ことがわかる。

5.2 テンポ推定のための目的関数

このように区間によってテンポが異なる曲について、テンポの変化時刻を推定する手法について述べる。 $[T_{m-1}, T_m]$ でのテンポが τ_m であるとき、この区間での目的関数の値は式 (14) より、

$$F[T_{m-1}, T_m] = \sum_{t_k \in [T_{m-1}, T_m]} \log P(\tau_k | \tau_m) \frac{\Delta t_k}{T} \quad (26)$$

と書ける。区間の数を M とし、各区間を区分する時刻を T_m で表す。 T_0 は演奏開始時刻であり、 T_M は終了時刻を表す。したがって、曲全体についての目的関数 F は

$$F = F[0, T] = \sum_{m=1}^M F[T_{m-1}, T_m] \quad (27)$$

と表すことができる。

²R. Schumann 作曲, Kinderszenen(「子供の情景」)Op. 15, no. 11

³() は小節番号を表す。

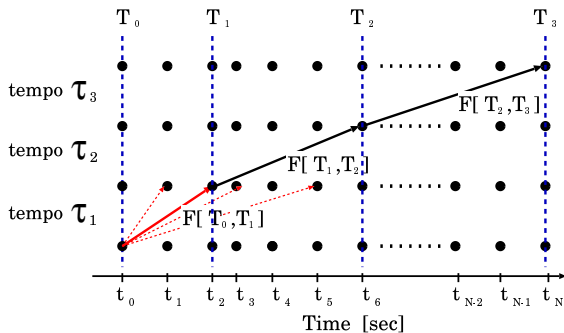


図 8: DP を用いたテンポ変動の推定

5.3 2 段 DP アルゴリズムによるテンポ変化時点の推定

複数区間で異なるテンポを許す式 (27) において、テンポの変化時点 T_1, \dots, T_{M-1} を推定するために、2 段 DP (動的計画法、Dynamic Programming) を用いる。以下に手順を示す。

まず、3 章で述べた手法を用いて各区間の間の累積距離 $F(t_i, t_j)$ を計算し、テーブルに用意する。次に通常の DP と同じ原理で、図 8 に示すように目的関数が最大となる各テンポ区間の各時刻での関数の累積値を逐次計算する。時刻 t_i に l 番目のテンポ区間が終了するときの、目的関数の時刻 t_i までの累積値を $\delta(t_i, l)$ とし、このテンポ区間の開始時刻を $\Psi(t_i, l)$ とする。累積値の計算は、以下の手順で行う。

1. $l = 1$ のとき、
 $\delta(t_i, l) = F[0, t_i]$
 $t_0 = \Psi(t_i, 1)$
2. $l = 2 \sim M - 1$ について、
 $\delta(t_i, l) = \max_j [\delta(t_j, l - 1) + F[t_j, t_i]]$
 $\Psi(t_i, l) = \operatorname{argmax}_{t_j} [\delta(t_j, l - 1) + F[t_j, t_i]]$
3. $l = M$ のとき、
 $F = \delta(t_N, M)$
 $= \max_j [\delta(t_j, M - 1) + F[t_j, t_N]]$
 $T_{N-1} = \Psi(t_N, M)$
 $= \operatorname{argmax}_{t_j} [\delta(t_j, M - 1) + F[t_j, t_N]]$

最後に、 $l = N - 1$ から $l = 2$ まで

$$T_{l-1} = \Psi(T_l, l)$$

によりトレースバックを行い、テンポ変化時点を求める。

なお、2 段 DP 法でなく one-pass DP アルゴリズム (あるいは Viterbi アルゴリズム) を用いることも可能だが、 $F[\cdot, \cdot]$ の計算が大きな部分を占めるので、計算効率について両者の間に大差はないと考えられる。

5.3.1 実験結果

ピアノ奏者 2 名 (A,B) による “Fürchtenmachen” (恐いお話) の演奏に対して、提案したテンポ変化時点の検出方法の評価実験を行った。異なるテンポの区間の数として $M = 7$ を用い、テンポが早くなる区間 B の検出を試みた。また、混合正規分布 $p(\tau)$ は、分布の中心から大きく外れたテンポに対して等しい尤度を与えるように補正を行った。⁴ 評価は次式で定める検出率で評価した。

$$\text{検出率 1} = \frac{\text{正しく検出した区間 } B \text{ の長さの合計}}{\text{正解の区間 } B \text{ の長さ}}$$

$$\text{検出率 2} = \frac{\text{正しく検出した区間 } B \text{ の長さの合計}}{\text{検出した区間 } B \text{ の長さ}}$$

⁴ $|\tau - \mu| \gg 1$ のとき $p(\tau) = 10^{-7}$ とした。

表 3: 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	95.2	100	67.8	51.4
検出率 2	42.6	40.2	25.8	49.1

検出結果を表 3 に示す。テンポの値が急に変化しているのを音価の誤推定に起因するテンポ誤りとして捉えてしまうため、正しく検出できないことが多い。このため、確率 $p(\tau)$ に単一の正規分布などを用いれば、精度を上げられると思われる。

6 おわりに

本稿では、確率モデルを用いてテンポを推定する方法について述べた。テンポの定義には任意性があり、テンポの値は必ずしも一意に決定できないが、テンポとリズムの関係を考慮して適切な演奏を表すリズムとテンポを推定することは本稿で示したとおり可能である。最初に、テンポに依存しないリズム特徴量を用いてリズムを推定し、その結果からテンポの変動を推定し、曲全体で一定のテンポを仮定する場合のテンポ推定、および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。

今後は、テンポが連続的に次第に遅くなる場合などに、テンポの連続変化も確率モデルで扱いたい。図 7 に見られるようにフレーズの終わりでは急激にテンポが遅くなる場合は多く、指数関数などを用いてモデル化が可能であると思われる。

また、今回はテンポの変化区間の数 M を与えて使用したが、これも自動で与えるように改善したい。今後は M の推定のためにパラメータの数を推定するために、MDL (minimum description length)、AIC (Akaike information criterion) など情報論的尺度の導入も試みたい。

参考文献

- [1] H. Takeda, N. Saito, T. Otsuki, M. Nakai, H. Shimodaira, S. Sagayama, “Hidden Markov Model for Automatic Transcription of MIDI Signals,” Proc. MMSP, 2002.
- [2] A. Cemgil, B. Kappen, P. Desain, H. Honing, “On tempo tracking: Tempogram Representation and Kalman filtering” Journal of New Music Research, 2000.
- [3] C. Raphael, “Automated Rhythm Transcription,” In Proc. of ISMIR, pp. 99–107, 2001.
- [4] 齋藤, 中井, 下平, 嵯峨山, “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,” 情処研報, 99-MUS-33, pp.27-32, Dec 1999.
- [5] 大規, 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,” 情報処理学会論文誌, Vol. 43, No. 2, pp. 245–255, 2002.
- [6] 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, 2002-MUS-46, pp. 23–28, 2002.
- [7] 武田, 篠田, 嵯峨山, “確率モデルによる多声楽曲 MIDI 演奏からの楽譜推定,” 情処研報, 2003-MUS-50, pp. 21–26, 2002.
- [8] L. Rabiner, and B.-H. Juang: Fundamentals of Speech Recognition, Prentice-Hall, 1993.
- [9] A. J. Viterbi, “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm,” IEEE Trans. Inform. Theory, vol. IT-13, pp.260-129, 1967.