リズム語彙を用いたHMMによるMIDI演奏のリズムとテンポ推定

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　本報告では、人間の演奏を記録したMIDIデータから演奏曲のリズムとテンポを推定する方法について述べる。我々は確率モデルを用いたリズム認識の手法として、単旋律MIDI演奏を対象にしたリズム語彙を用いたHMM(hidden Markov model, 隠れマルコフモデル)を提案した。今回はこれを拡張して、多声音楽のリズム認識を行う。以前に提案したリズムベクトルによる多声音楽のリズム認識法では音価のみを推定したが、提案手法は拍節の情報を含んだリズム情報を推定する。リズム語彙の文法は、既存の楽曲のリズムから学習により求めることができる。また、演奏者の意図したテンポを推定するために実演奏の発音時刻と楽譜での発音位置の関係からテンポを連続関数として定義し、テンポを最尤推定することを提案する。新たに導入されるテンポの定義は、音長と音価の関係から定義されるテンポを拡張したものになっている。装飾音を除去したクラシック音楽3曲のMIDI演奏に対してclosedデータによる学習したモデルで92.3%, openデータで学習したモデルで77.5%の音価推定率を得た。

Estimation of Tempo and Rhythm from MIDI Performance Data based on Rhythm Vocabulary HMMs

Haruto TAKEDA　Takuya NISHIMOTO　Shigeki SAGAYAA
Graduate School of Information Science and Technology, The University of Tokyo

Abstract　This paper describes rhythm recognition technique from performance data recorded in MIDI format. We have already proposed rhythm recognition method for monophonic MIDI performace based on probabilistic models using rhythm vocabulary Hidden Markov Models (HMMs). In this paper, we extend this model to deal with polyphonic MIDI performances. The proposed method can estimate rhythm information including not only note values but beats, which our previous method based on rhythm vectors can not deal with. Grammer of rhythm words in rhythm vocabulary are trained through stochastic training using existing music scores. We also show formulation of tempo as continuous change and method of estimating tempo line using information of observed onset times and onset position in a score. The process of tempo change are model as a stochastic process and tempo lines are estimated based on maximum likelihood. Experimental results are also discussed in this paper.

1　はじめに

本稿では、人間による演奏(以下、実演奏と呼ぶ)を記録したMIDI(Musical Instruments Degital Interface)データからテンポとリズムを推定する手法(以下、リズム認識と呼ぶ)について論じる。我々が以前に提案した多声音楽のリズム認識手法は音価のみを推定するものであったが、今回は、音価だけでなく拍や小節線位置も推定対象に含んだリズム推定の手法を提案する。さらに、演奏者の意図する連続的に変化するテンポをモデル化し、最尤推定によるテンポ推定法を提案する。

本稿で対象とするリズム認識とは、音長に揺らぎが含まれる演奏から、リズム、拍、テンポを推定する技術を指す。人間は音楽演奏においてテンポや音の長さを楽譜によって定められたとおりには演奏せず、意図的な(ときには無意識な)変動をリズムやテンポに施すので、実演奏で観測される音長には揺らぎが含まれている。

我々は、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)[1]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くことを提案してきた[2, 3]。単旋律の演奏のリズム推定では、「リズム語彙」を用いたHMM、音符n-gramモデル、音長比(リズムベクトル)を特徴量としたHMMを提案した。「リズム語彙」を用いたHMMは、リズムだけでなく、小節線の位置や、拍子なども推定できた。リズムベクトルはテンポに依存しないので、

表 1: 通常の音声認識の確率モデルと提案するリズム認識の確率モデルの対応関係

	音声認識	リズム認識
観測信号	スペクトル時系列	音長の時系列
HMM	音素 / 単語	リズム単語
文法	単語	リズム単語
認識対象	文	楽曲

テンポが未知である演奏からリズムを推定するのに有効であったが、リズムベクトルのみを特徴量とする HMM では、拍や小節線の位置などは推定対象に含まれていなかった。その後、多声音楽の MIDI 演奏のリズム認識を行うために、リズムベクトルを特徴量とした HMM を用い音価を推定した [4]。本稿ではさらに「リズム語彙」を用いた HMM を拡張して多声音楽の演奏に適用し、音価のみならず強拍の位置も推定対象に含んだリズム認識について述べる。

尚、同様の内容を扱う研究はいくつか存在する。Cemgil ら [5] は、テンポを隠れ変数としたカルマンフィルタを用いて、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[6] 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。これらの研究と我々の手法の大きな相異点は、これらの研究で用いられている確率モデルでは、2 つの発音時刻の間の関係がモデルの単位であるのに対し、我々のモデルの単位は一小節内での発音位置でありモデルで扱う情報量が多い点が異なる。

2 リズム語彙を用いたリズム認識

2.1 リズム語彙

提案するリズム認識の基本概念である「リズム語彙」について述べる。音楽経験者は、音長に揺らぎのある演奏を聞いて正しいリズムを楽譜に書き起こすことができる。その理由として、その人がそれまでの音楽についての経験を通して、それぞれのリズムパターンの音長がどのように変動するか、また、リズムパターンが表われやすいかを知っていることが挙げられる。そこで、我々は譜面に表われるであろうリズムパターンを「リズム単語」とし、このリズム単語がどのような音長で演奏されるか、また、このリズム単語がどのような文脈 (単語の繋がり) で現れやすいかを HMM でモデル化し、リズム認識を HMM における最尤状態系列の探索問題として解くことを提案した。表 1 に示すように、スペクトルの時系列から対応する単語を推定し、さらに単語の連鎖から意味のある文章を推定する過程は、リズム単語を用いてリズムを推定する過程と対応が取れるため、リズム認識と音声認識は同型の問題であると言える。音声認識における単語の集合を指す語彙に対応する用語として、リズム認識におけるリズム単語の集合を「リズム語彙」と呼ぶことにする。

以前我々が提案したリズム語彙によるリズム認識法は、単旋律を対象としたものであった [3]。今回は、これを多声音楽を扱えるように拡張する。

2.2 音長と音価

実演奏で観測される音長と、楽譜に記される音の長さの情報である音価との間にある数値としての関係を定式化する。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

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

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 M (bpm, beat per minute, 毎分の拍数) とは

$$M[\text{拍/分}] = \frac{60[\text{秒/分}]}{\tau[\text{秒/拍}]} \tag{2}$$

の反比例の関係がある。我々の目的は、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し、さらにテンポを推定することである。尚、ここでの τ は、IOI の区間で一定値をとるものになるが、4 節でこれを連続関数に拡張する。

2.3 リズムとテンポの推定

確率モデルを用いたリズム・テンポの推定の原理を述べる。音長の時系列をどのような音価に変換するかは、解釈によっては複数の可能性がある。例えば、市販ソフトを用いて量子化して得られるタイや 32 音符をたくさん含んだ楽譜も、ひとつの演奏のリズムのひとつの解釈と言える。そこで、我々は演奏とリズムの対応に確率を与え、最も適切である確率の高いリズムを得るという、確率を用いたアプローチでリズム認識を扱う。

まず、リズム認識を確率的な問題として定式化する。演奏者がテンポ $T{=}\tau$ でリズム $Q{=}\{q_t\}_{t=1}^N$ を演奏しようと意図して音長 $X{=}\{x_t\}_{t=1}^N$ 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

$$\{\hat{Q}, \hat{T}\} = \operatorname*{argmax}_{Q,T} P(Q, T|X) \tag{3}$$

によって推定する問題として扱える。Q がリズム単語の時系列 $W{=}\{w_m\}_{m=1}^M$ として表すことができるならば、

$$\{\hat{W}, \hat{T}\} = \operatorname*{argmax}_{W,T} P(W, T|X) \tag{4}$$

となる。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、リズムの推定とテンポの推定を別々に行う。

リズム推定では、演奏者の意図したテンポを直接扱わずに $\hat{W} = \operatorname{argmax}_W P(W|X)$ として求める。3 節で述べるように、ここで使用する確率モデルにおいて、テンポは確率モデルの中には含まれリズム推定に補助的な役割を果たすが、演奏者の意図したテンポとしては求めない。Bayes の定理を用いると

$$\hat{W} = \operatorname*{argmax}_{W} P(X|W)P(W) \tag{5}$$

となる。

意図されたテンポは、実演奏 X とリズム Q が与えられたときに、用意したテンポモデルにフィッテイングす

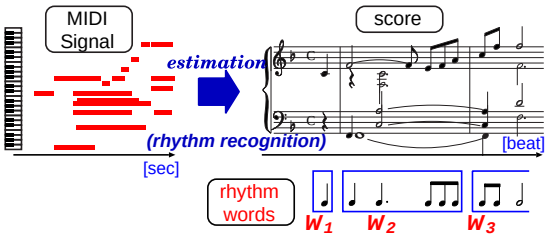

図 1: リズム推定: 実演奏の IOI の時系列 $\{x_t\}_{t=1}^{N}$ からリズム単語の時系列 $\{w_m\}_{m=1}^{M}$ を推定

ることで求める。これは、事前にリズム推定を行うか、あるいは、演奏に対応する楽譜を用意しておく場合を想定している。演奏される音長は、意図したテンポとリズムに対して確率分布に従う揺らぎを持つとし、4 節で述べるテンポモデルのパラメータを

$$\hat{T} = \underset{T}{\operatorname{argmax}}\, P(T|X, \hat{Q}) \tag{6}$$

によって推定する。

3 多声音楽のリズム推定のための HMM

3.1 多声音楽のリズム単語

多声音楽のリズムを扱うために、リズム単語を導入する。楽譜に記されている全ての音の発音位置に注目し、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。強拍によって区切られるこの音価のパターンをリズム単語とする。この区切りは、例えば図 1 に示すように小節を単位にすることができる。今回は、強拍部には必ずひとつの発音があるものとし、全ての声部がシンコペーションとなっているものは扱わない。

3.2 リズム単語の HMM

単旋律のリズム単語 HMM を多声音楽にも対応させるために、同時に複数の音を発音する演奏に対応した HMM について述べる。

リズム単語 $w_i = \{q_1, \cdots, q_{K(i)}\}$ に対応する実演奏の IOI が $x_t, \cdots, x_{t+n-1}$ であったとする。ここで、$k(i)$ は、リズム単語に含まれる音価の個数であり、n は対応する実演奏の IOI の個数である。2 つの音が同時に発音されたとき、この 2 つの音の IOI は理想的には 0 であるが、実際には厳密に同時に発音されることは稀で僅かな時間差で発音されるので、図 2 の左図の x_1 のように短い IOI として観測される。このような IOI を以後「同時発音の IOI」と呼び、それ以外の IOI を「リズムを構成する IOI」と呼ぶ。

同時発音の IOI はリズム単語中のひとつの音価 q_j に対応する状況を、この音価 q_j をひとつの状態 s_j とし、同時発音の際の短い IOI は状態 q に自己遷移するときに出力されるものとモデル化する。また、同時発音でない通常我々がリズムの構成要素として認識される長さの IOI は、リズム単語内の音価が次に遷移するときの出力信号であるとする。リズム単語の音価は、観測信号である IOI の時系列に対して未知であるシンボル時系列であるため、HMM となる。

リズム単語の音価と実演奏の IOI は、この HMM を用いて確率的に対応付けることができる。同時発音の IOI は、$x > 0$ で定義される次の確率分布にしたがって出力

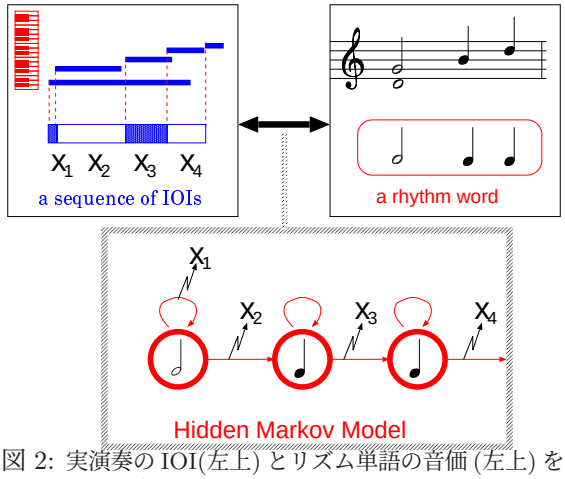

図 2: 実演奏の IOI(左上) とリズム単語の音価 (左上) を確率的に関連付ける HMM(下)。リズム単語をひとつの HMM とモデル化し、一つの音価を一つの状態にすると同時発音は状態の自己遷移に対応する。

されるとする。

$$b_{ss}(x) = \frac{2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{7}$$

リズムを構成する IOI に対する確率値は、3.3 節で述べるリズムベクトルで与える。また、状態遷移確率 $a_{s(t)s(t+1)}$ はその状態に留まる回数の確率値を与えるので、同時発音数についての確率を与えている。

3.3 リズムベクトル

リズムを構成する IOI に対する確率値は、テンポに依存しない特徴量である IOI の比に対して与える。

式 (1) より、テンポ τ が一定と見なせれば、IOI の比と音価の比はおよそ等しいことが分る。そこで、IOI の組をベクトルと見なし、このベクトルを成分の和が 1 になるように正規化したものをリズムベクトル $\boldsymbol{r}$ と呼び、我々は以前にリズム認識の特徴量として導入した []。実演奏のリズムベクトルは、音価の比から計算されるリズムベクトルに対して変動するので、ここではその変動が多次元正規分布 $c(\boldsymbol{r})$ に従うものとモデル化する。

以上よりリズム単語 w_i が IOI の時系列 $\{x'_t\}_{t'=t}^{t+n-1}$ として演奏される確率は、

$$\begin{aligned} &P(x_t, \cdots, x_{t+n-1}|w_i) \\ &= \prod_{t'=t}^{t+n-1} a_{s(t')s(t'+1)} b_{s(t')s(t'+1)}(x'_t) c(\boldsymbol{r}_{t'}|w_i) \end{aligned} \tag{8}$$

と表される。

3.4 テンポの揺らぎの確率

演奏中のテンポの揺らぎについても確率を与える。Adagio から Allegro のようにテンポが急激に変化しないならば、実演奏のリズム単語単位でのテンポはほとんど一定であると考えられる。これは、テンポの変動 (差分) の統計は 0 を中心に分布することが期待される。そこで、リズム単語の平均テンポを

$$\bar{\tau}_t = \frac{x_t + \cdots + x_{t+n-1}}{q_t + \cdots + q_{t+n-1}} \tag{9}$$

とし、この時系列 $\{\bar{\tau}_m\}_{m=1}^{M}$ の差分は確率を

$$\bar{\tau}_{m+1} - \bar{\tau}_m \sim N(0, \sigma) \tag{10}$$

とする。

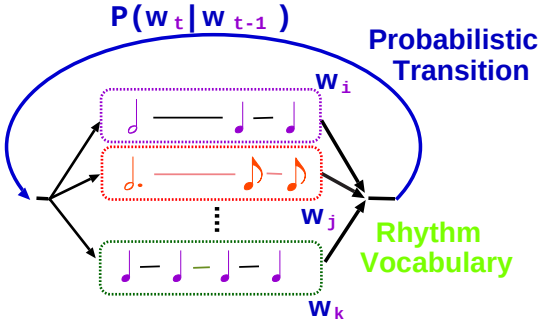

図 3: リズム単語の bigram モデル

表 2: リズム語彙の HMM によって与えられる尤度

尤度を与える対象	HMM で尤度を与える部分
同時発音の IOI	状態自己遷移の出力確率
同時発音数	状態遷移確率
リズムを構成する IOI	リズムベクトルの出力確率
リズムの接続	リズム単語 n-gram
テンポの揺らぎ	平均テンポの揺らぎの確率

3.5 リズム語彙の文法

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単語に文法を導入する。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。ここでは、リズム単語の n-gram を考え、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 $P(q_t|q_{t-1},\cdots,q_{t-n+1})$ で近似できるとする。リズム単語の時系列 $W=\{w_1,\cdots,w_N\}$ の出現確率は、図 3 に示す bigram モデルであれば

$$P(W)=P(w_1)\prod_{m=2}^{M}P(w_m|w_{m-1}) \qquad (11)$$

と近似される。

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。n-gram の初期単語の出現確率からは、惹起で始まる小節を学習できる。また、n-gram モデルにより、2 拍子、3 拍子などのそれぞれの拍子のリズム単語で接続することが学習されるので、リズム単語の文法により拍子についても構造が学習される。

3.6 HMM ネットワークの探索

以上の HMM ネットワークを用いて、与えられた IOI の時系列を出力信号とする HMM の状態遷移系列からの出力とすることで、表 2 に示すように実演奏の各要素に尤度を与えられ、その結果、式 (5) の確率が与えられる。従って、式 (5) によるリズム推定の問題は、リズム単語の HMM によって構成される HMM ネットワークにおいて最も尤度の高い HMM の経路を求める探索問題となる。このため、HMM における最適な状態遷移系列と、HMM ネットワークでの最適な経路を計算する必要がある。HMM における最尤状態系列は、効率的な探索アルゴリズムである VDA[7] (Viterbi Decoding Algorithm: ビタビ復号化アルゴリズム) を利用して求められ、さらに HMM のネットワークでの探索はレベルビルディングを用いて行うことができる。

表 3: 評価データに用いたクラシック音楽のピアノ作品

作曲者 (曲名)	演奏曲
J. S. Bach (Fuga)	平均律クラヴィーア曲集第 1 巻より ハ短調の Fuga BWV847
R. Schumann (Träumerei)	組曲「子供の情景」op.15, より no.7 トロイメライ (Träumerei, 夢)
L.v.Beethoven (Sonata)	Piano Sonata Op.49-2 より 第 1 楽章前半部

表 4: 実演奏のリズム認識評価：音価正解率 (リズム単語正解率) [%]

学習データ	**closed 1**	**closed 2**	**open**
Fuga	100 (100)	100 (97.2)	100 (52.0)
Träumerei	96.0 (75.0)	77.7 (29.1)	87.6 (29.1)
Sonata	100 (100)	100 (78.9)	45.0 (42.0)
平均	98.6 (91.6)	92.3 (68.4)	77.5 (41.0)

3.7 評価実験

提案手法を既存のクラシック音楽の実演奏を MIDI データとして記録したものを用いて評価した。評価データには表 3 に示す 3 曲の演奏を使用した。Träumerei は前打音を含み、楽曲中に rit.(だんだん遅く) やフェルマータの指示があるなど、テンポの変動が大きい曲である。Sonata は拍の刻みが 8 分音符である場合と符点 8 分音符である場合の 2 つがある。bigram 文法は文法の学習に次の 3 種類の楽曲を用いた。

closed 1: 演奏楽曲
closed 2: 演奏楽曲を含むピアノ作品 13 曲 (リズム単語 107 個)
open: 演奏楽曲を含まないピアノ作品 10 曲 (リズム単語 119 個)

正解率は、音価とリズム単語のそれぞれについて、

$$\frac{\text{正解音価数 (単語数)} - \text{挿入誤り} - \text{削除誤り} - \text{置換誤り}}{\text{正解音価数 (単語数)}}$$

を用いて数え、表 4 に示す正解率を得た[1]。リズム単語は 1 拍 (惹起の曲の冒頭)、2 拍、3 拍、4 拍のものがあり、同じ音価列を異なるリズム単語の組合せで表現できる。このため、リズム単語の推定としては不正解でも音価や拍位置 (強拍・弱拍の関係) としては正しく推定されることが多く見られた。Fuga において、小節線は多めに挿入されたが、音価は正しく推定できた。Träumerai ではテンポがゆっくりになるのを音価の変動と見なした結果、符点 4 分音符がより長い音価と推定される誤りが見られた。

4 テンポの推定

4.1 演奏者が意図したテンポ

4 節では、実演奏のデータと演奏された音楽のリズムが与えられたときに、演奏者が意図したであろうテンポを連続関数として求める方法を述べる。2.2 節で定義したテンポは各 IOI に対して計算されるものであった。この実演奏中の時間と楽譜の音価を結びつける関係を保ったまま、連続的に変化するようにテンポの定義を拡張する。また、演奏者の演奏するテンポを、各 IOI ごとでは

[1] 4 拍のリズム単語を 2 拍のリズム単語 2 つで正しい音価を与える推定結果は正解とした。

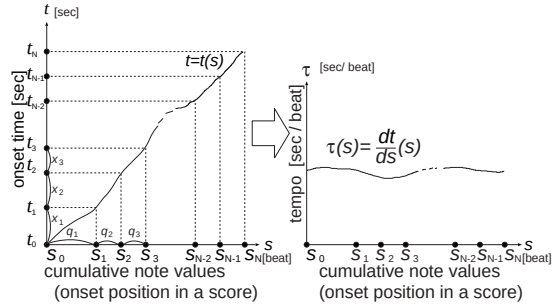

図 4: 累積音価 s と発音時刻 t の関係 (左)、およびその微分として求められるテンポ (右)

表 5: テンポと物体の物体の運動の「はやさ」

	物体の運動	音楽演奏
媒介変数	時刻 t	楽譜上の発音位置 s
変化量	位置 y	実演奏の発音時刻 t
観測量	(t, y) の組	(s, t) の組
平均の速さ	$\bar{v} = \Delta y/\Delta t$	$\tau = \Delta t/\Delta s$ (式 (1))
瞬間速度	$v = dy/dt$	$\tau(s) = dt/ds$ (式 (13))

なく、より大きなフレーズの単位でテンポを意図し、フレーズ単位で滑らかな曲線で図示できると考えられる。このような演奏者の意図したテンポ曲線を、実演奏データから求めるために、実演奏は意図したテンポに変動が加わったものとモデル化し、テンポ曲線を最尤法で推定する。このテンポ推定方法は、リズム認識以外にも、音楽練習や名演奏家の奏法を客観的な数値により解析するなどの演奏解析に有効であり、また、自動演奏のためのモデルやデータベース作成 [8] のための技術にも関連している。

4.2 微分係数としてのテンポ

実演奏の n 番目の発音時刻 t_n は、同時発音をひとつの発音としたときの IOI の時系列 $\{x_k\}_{k=1}^N$ を用いて $t_n = \sum_{k=1}^{n} x_k$ として表される。t_n に対応する楽譜上の累積音価 s_n は、楽譜上の発音位置の間の音価の時系列 $\{q_k\}_{k=1}^N$ を用いて、$s_n = \sum_{k=1}^{n} q_k$ として表される。式 (1) で定義されたテンポは、

$$\tau_n = \frac{x_n}{q_n} = \frac{t_{n+1} - t_n}{s_{n+1} - s_n} = \frac{\Delta t_n}{\Delta s_n} \tag{12}$$

と表された。そこで、実数値 s で指定される楽譜上の位置のテンポは、

$$\frac{\Delta t}{\Delta s} \longrightarrow \frac{dt}{ds} \tag{13}$$

で定義するのが自然である。即ち、図 4 に示すように、発音時刻 t を譜面上での累積音価によって表された発音位置 s の区分的に連続な関数 t=$t(s)$ とし、その導関数としてテンポを定義する。

式 (1) と式 (13) の 2 つのテンポの定義は、表 4.2 に示すように物体の運動を記述するときに用いる「平均速度」と「瞬間の速度」の関係に相当する。

実演奏とリズムの情報は、楽譜での発音位置 s と実演奏の発音時刻 t の組 (s_n, t_n) として与えられている。これらの情報から連続テンポを求めるには、

$$t_n = t(s_n) \quad (n = 0, 1, \cdots, N) \tag{14}$$

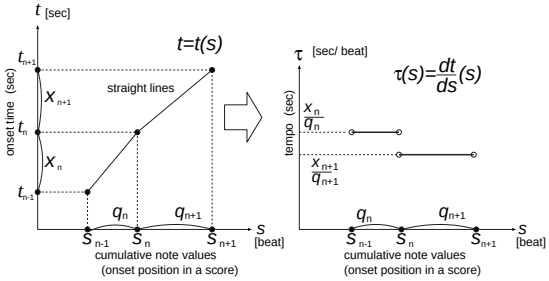

図 5: 観測した (s_n, t_n) を直線で結んで得られる $t = t(s)$(左). および式 (1) による定義のテンポ (右)

を満たす連続関数 (やや詳しく言えば、区分的に微分可能な関数) t=$t(s)$ を求めればよい。式 (1) によるテンポは、図 5 に示すように、st 平面で観測点を直線で結んだ場合に相当する。

また、$t(s)$ の逆関数 s=$s(t)$ を用いることで、メトロノーム表記で用いるテンポ (bpm) を時間の連続関数 $M(t)$ としての定義は、式 (2) を拡張して

$$M(t) = 60 \cdot \frac{ds}{dt}(t)$$

として計算することもできる。

4.3 対数スケールのテンポ変動

ここでは、人間のテンポの変動に対する感覚が対数的であると仮定して、テンポを対数スケールで考える。例えば 2 倍の速さに変化する場合と半分のテンポに変動する場合が同じ距離尺度を与えることができる。また、対数軸上で直線で表されるテンポ変動は、線形時間軸では指数関数的な変化に対応する。

ここで、演奏者はテンポ演奏について意図を持って演奏し、実演奏はこの意図されたテンポ $\tau(s)$ に対して対数スケールで近くされるテンポ変動 $\epsilon(s)$ を伴い、その結果実演奏のテンポは対数スケールで

$$\log \tau(s) + \epsilon(s)$$

として観測されるとする。誤差が全くない ϵ=0 の場合に線形時間軸で観測される発音時刻 t と一致するとして、

$$t = e^{\epsilon(s)} t(s)$$

の関係が成立する。この確率過程で、観測される発音時刻が t_n として観測されるので、

$$t_n = e^{\epsilon_n} \cdot t(s_n)$$

が成り立つ。ϵ_n が正規分布 $N(0, \sigma^2)$ に従うと仮定すると、演奏者が意図したテンポに対して発音時刻の揺らぎは $e_1, \cdots, \epsilon_N$ である確率は、

$$p(\epsilon_1, \cdots, \epsilon_N) = \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log t_n - \log t(s_n))^2}{2\sigma^2} \right) \tag{15}$$

で与えられる。

4.4 テンポの最尤推定

テンポを求めることは、譜面上での発音位置と実演奏での発音時刻の組 $\{(s_n, t_n)\}_{n=1}^N$ から累積音価 s と実演奏での時刻 t の対応を与える関数 $t(s)$ を求めることと等価であった。そこで、与えられた発音位置と累積音価の

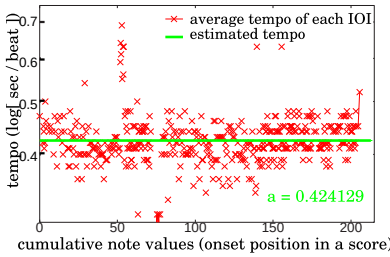

図 6: 一定テンポで意図された演奏のテンポ推定 (点のプロットは対数スケールでテンポの観測点)

組 (s_n, t_n) に対して最も尤もらしい関数 $\hat{t}(s)$ として推定することを考える。

$$\hat{t}(s) = \underset{t(s)}{\operatorname{argmax}} P(t(s)|s_1, \cdots, s_N, t_1, \cdots, t_N)$$

ここで、式 (15) と上式の対数を用いると、$t(s)$ の最尤推定は二乗誤差

$$D = \sum_{k=1}^{N} \left(\log \frac{t_n}{t(s_n)} \right)^2 \qquad (16)$$

を最小にする $t(s)$ を求めることと等価である。

4.5 種々のテンポモデルのパラメータ推定式

type. 1 定数 $\log \tau(s) = a$**(const.) のとき**
$t(s)$=as なので、モデルパラメータ a の最尤推定値は

$$\hat{a} = \frac{1}{N} \sum_{n=1}^{N} \log \frac{t_n}{s_n}$$

で得られる。

type. 2 一次関数 $\log \tau(s) = as + b$ **のとき**
この場合は $t(s)$=$e^a(e^{bs} - 1)/b$ であり、式 (16) を直接最小化するのは難しい。おおまかな傾向を掴むために最も簡単な方法は、対数スケールでの各 IOI の平均テンポから傾き b を最小二乗法で求める。線形時間軸との整合性を保つには、観測点の最後の点 (s_N, t_N) が一致するように a を定めればよい。

4.6 テンポの推定例

実演奏のテンポを提案手法で推定する。Beethoven のピアノソナタ (Op.49-2, 1st Mov) の実演奏を **type.1** のテンポモデルを全区間を通した結果を、図 6 に示す。推定した一定テンポの周りに、ほぼ均等に分布しているのが観測される。さらに、R. Schumann の「子供の情景」(Op.15) より第 11 曲「こわいお話 (Fürchtenmachen)」の前半部を複数の区間に分け、**type2** のモデルでテンポを表したものを図 7 に示す。この曲は 4 小節 (8 拍) 単位でフレーズが構成されているので、4 小節毎にテンポ推定を行った。途中でテンポが変わること、また各フレーズはテンポが僅かながら減速傾向にあることが分る。

5 おわりに

本稿では、以前に提案した「リズム語彙」の HMM による単旋律のリズム認識手法を拡張し、多声音楽の MIDI 演奏から小節線を含むリズム情報を推定する方法を提案した。3 曲のクラシックピアノ曲の演奏を記録した MIDI データを用いた性能評価実験で、open データで学習した

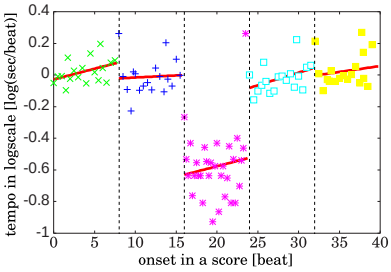

図 7: Schumann の「こわいお話 (Fürchtenmachen)」の実演奏のフレーズ毎のテンポ推定。発音位置 17.0 から 24.0 までは楽譜の指示どおり「早く (Schneller)」演奏している。

モデルで 92.3%, closed データでのモデルでは 77.5%の音価正解率を得た。また、リズムと演奏のデータから演奏者の意図したテンポを推定する方法を提案した。

今後は、和声の動きや音の強さなど、人間がリズムを推定するときに用いる情報をより多く盛りこんだ確率モデルを用いて、リズム認識率の向上を図るとともに、調整、和声なども同時に推定する自動採譜のための確率モデルを検討したい。

参考文献

[1] L. Rabiner, and B.-H. Juang: Fundamentals of Speech Recognition, Prentice-Hall, 1993.
[2] 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.
[3] 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.
[4] 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.
[5] A. Cemgil, B. Kappen, P. Desain, H. Honing, “On tempo tracking: Tempogram Representation and Kalman filtering” Jounal of New Music Research, 2000.
[6] C. Raphael, “Automated Rhythm Transcription,” In Proc. of ISMIR, pp. 99–107, 2001.
[7] A. J. Viterbi,“Error bounds for convolutional codes and an asymptotically optimum decodeing alogorithm,” IEEE Trans. Inform. Theory, vol. IT-13, pp260-129, 1967.
[8] 豊田, 片寄, 野池, “音楽解釈研究のための演奏 deviation データベースの作成,” 情処研報, 2003-MUS-51, pp.65–70, 2003.