

論文 / 著書情報
Article / Book Information

論題(和文)	確率モデルによる多声楽曲MIDI演奏からの楽譜推定
Title(English)	
著者(和文)	篠田浩一
Authors(English)	Koichi Shinoda
出典(和文)	情報処理学会研究報告, Vol. 2003-MUS-50, No. , pp. 27-32
Citation(English)	, Vol. 2003-MUS-50, No. , pp. 27-32
発行日 / Pub. date	2003, 5
権利情報 / Copyright	<p>ここに掲載した著作物の利用に関する注意: 本著作物の著作権は(社)情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。</p> <p>The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). This material is published on this web site with the agreement of the author (s) and the IPSJ. Please be complied with Copyright Law of Japan and the Code of Ethics of the IPSJ if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof.</p>

確率モデルによる多声楽曲 MIDI 演奏からの楽譜推定

武田晴登¹ 西本卓也¹ 篠田浩一² 嵯峨山茂樹¹

¹ 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)

² 東京工業大学 情報理工学系研究科 (〒 152-8552 東京都目黒区大岡山 2-12-1)

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

あらまし 多重音を含む楽曲の演奏の MIDI(Musical Instruments Digital Interface) 信号からの自動採譜に確率モデルを用いる。MIDI 信号を対象とした場合は演奏の音高情報は既知であるので、リズム推定が必要である。演奏の速度が一定である場合は量子化によるリズム推定が容易であるが、演奏中にテンポが変動する曲では難しい。我々は、多声部に跨る IOI(発音時刻の間隔) から得られる相対的な音長情報であるリズムベクトルを確率モデルの特徴量とし、モデルのパラメータ値を楽譜や演奏から学習する。実際の採譜手順は、同時発音の検出の後に IOI の時系列を求め、HMM(隠れマルコフモデル) を用いて IOI の時系列から最適なリズム譜を推定し、そしてリズム譜をもとに各音価の推定する。5 人の奏者による電子ピアノの演奏に対して性能評価実験を行い、音価復元率として「フーガ」に対して 92.2%、「トロイメライ」に対して 52.1%を得た。

キーワード 自動採譜、リズム認識、電子ピアノ演奏、HMM(隠れマルコフモデル)、多声楽曲

Score Estimation from Polyphonic MIDI Performance with Probabilistic Models

Haruto TAKEDA¹, Takuya NISHIMOTO¹, Koichi SHINODA², and Shigeki SAGAYAMA¹

¹Graduate School of Information Science and Technology, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan

²Department of Computer Science, Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8852, Japan

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

Abstract This paper proposes an automatic transcription method for polyphonic musical performances in MIDI signals. Pitches and rhythms are basic information which is necessary to write scores. From the MIDI signals of human performances, we just need to recognize rhythms from time information in MIDI signals because pitches are already known in MIDI signals. We propose a method on rhythm recognition, especially targeting at polyphonic music performances. In the proposed probabilistic models for rhythm recognition, we use rhythm vectors, which are obtained from IOI(Inter-Onset Interval) sequence across the multi voices, as a feature of probabilistic models. The values of parameters in our model can be optimized by the learning from scores and human performances. In experiments on performances by 5 piano players with an electronic piano, we obtained score restoration rates of 92.2% for “Fuga” and 52.1% for “Träumerei”.

Key words Automatic transcription, HMM, Rhythm Recognition, electronic pianos, polyphonic music

1. はじめに

我々は、自動採譜の実現に向けて研究を行っている。自動採譜で行う処理は、(1) 音響信号からピッチ (音階名) と発音時刻などの演奏情報を推定する、(2) 演奏情報から人間の常識にあった楽譜を推定する、の 2 つの過程に分けて行うのが一般的

であり [1], [2]、我々もこの枠組にしたがって研究を行っている。ここで、演奏される各音のピッチと発音時刻と継続時間を含む情報を演奏情報と呼び、MIDI(Musical Instrument Digital Interface) 信号の形式で表現され得るものとする、これを楽譜に変換するには、さらに各音符の音価を決定する必要がある。本稿では、この目的で、リズム推定の手法について述べる。

本稿では、多声楽曲の音楽演奏からの採譜を行うため、確率モデルとその学習法、手順を述べ、5人のMIDIピアノ演奏者による実演奏を対象に行った性能評価実験の結果を報告する。

2. 従来手法

音楽の時間情報はリズム(各音符の音価情報)とテンポ(演奏の速度)で構成される。採譜に必要なリズム推定では、この両方を正しく推定しなければならない。

2.1 リズムの推定

演奏データのテンポ、あるいは拍情報が既知であるときに、各々の音について音長から音符に変換する方法として、音長の「量子化」(quantization)が現在広く用いられている。量子化とは、演奏者が楽譜入力を目的としてメトロノームを用いてテンポを一定に保ちながら演奏し、演奏された音長を決められた単位(量子)の整数倍に対応させる処理で、通常は閾値処理で行われる。この手法により音長から音価へ変換できるが、しばしば、演奏者の意図したリズムとは異なる、また、人間が楽譜に対して持っている常識に反する楽譜が出力される。これは、演奏者によって意図的あるいは無意識になされる音長の変動を量子化した結果であり、観測の精度の問題ではない。また、特に楽譜入力を目的としていない通常の演奏からの適切な楽譜化は、上記の方法では容易でない。

これに対して、我々は、演奏される音長のゆらぎを確率的に扱うと同時に、ありふれたリズムや稀なリズムなどのリズムの出現しやすさを学習した確率モデルを用いる、モデルベースのアプローチで研究を進めてきた[3]~[5]。リズムの推定と音声認識[6]を同一の確率モデルにおける推定問題として捉え、現在の連続音声認識で一般的に用いられているHMM(Hidden Markov Model, 隠れマルコフモデル)を用いてモデル化を行い、市販ソフトの量子化機能を上回るリズムの推定性能を得た。この手法は、多声部の構造を持つ楽曲にも応用できることが報告されている[7]。

2.2 テンポ推定

量子化の処理を行うためにはテンポあるいは拍情報が必要であるが、通常の音楽演奏ではメトロノームなどを用いず拍を指定を行わずに演奏を行うので、一般に音楽演奏におけるテンポや拍の情報は未知である。テンポの推定を行わず(あるいはテンポを誤推定して)量子化を行うと、図1に見られるように、全く演奏者の意図とは異なる楽譜が出力され、量子化のみでは採譜の道具として使用することができないことが多い。演奏か

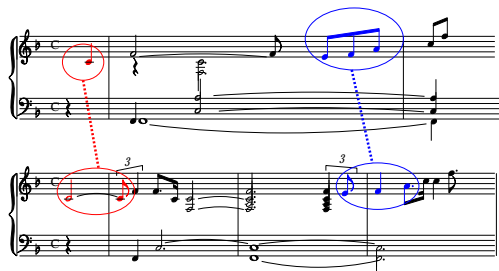


図1 市販の楽譜作成ソフトで量子化(トロイメライを演奏)

らテンポを推定する研究は、主にテンポトラッキングと呼ばれ

る研究テーマとして研究されている。Cemgilら[8]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。

2.3 テンポとリズムの同時推定

リズムとテンポは互いに関連している。リズムとテンポを同時に推定する方法として、Raphael[9]は確率モデルを用いてリズムとテンポを同時に推定する手法を提案し、クラシック音楽のピアノ演奏のMIDI信号に適用可能であることを報告しているが、ここでは各音の音価の推定は扱っていない。また、この方法では、対象曲に応じて小節単位での可能な発音位置を適切に定めておく必要があった。これに対して、我々はテンポに依存しない特徴量を用いたHMMを用い、対象曲に関する事前知識を用いることなく採譜を目的に各音の音価を推定する手法を提案した。単旋律を対象に実験を行い、その有効性を示した[10]。また、この手法は、多声楽曲の演奏に対しても適用可能であることを予備実験で示した[11]。

3. 多声楽曲のためのリズムベクトルを用いたHMM

ここではリズム認識の問題を定式化し、リズムベクトルを観測量とするHMMについて述べる。

3.1 音価と音長

音楽の音長を表す情報は、楽譜に書かれる音符から得られるものと、人間の実際の演奏から得られるものの2種類がある。本稿では、音符が持つその音の継続時間の情報「音価」を、4分音符の長さを1とした基準定める。(8分音符の音価は $\frac{1}{2}$ になり、2分音符の音価は2となる。)また、実演奏の音長としては、その音の継続時間ではなく、その音が含まれている旋律でIOI(inter-onset interval, 発音時刻の間隔)を用いることが一般的である。本稿でも、IOIを用い(3.3参照)、単位を秒として行う。観測される観測される演奏の音長情報から全ての音について適切な音価を推定することを、ここではリズム認識と呼ぶことにする(図2)。

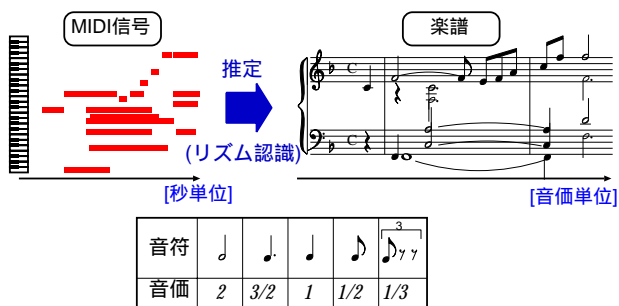


図2 2種類の音長の情報: 楽譜における「音価」と演奏におけるIOI

これら2種類の音長情報である楽譜の音価 q と演奏での音長(実際にはIOI) x は、以下のようにテンポ τ を導入することで両者の関係を定式化できる。

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

テンポは音価あたりの時間という次元を持ち、変動 ϵ は音価 q

と同じ次元を持つとする。我々は、 x についての時系列を観測し、これに対応する q の時系列を推定する。 x の変動は、未知である τ と ϵ の 2 変数の変動から引き起こされる。

テンポ τ が未知であるので、テンポの推定において回避できない誤認識が存在する。例えば、全て等間隔に手拍子を打ったとき、その間隔は 4 分音符としても 8 分音符としても表現可能である。

3.2 リズム譜

楽譜での各音の発音位置において、それらの間隔に対応する音価の音符を配置したものを、リズム譜と呼ぶことにする。リズム譜は、多声楽曲の音価情報を 1 次元の時系列に射影したものと捉えることができ (図 3)、もとの楽譜が 1 次元の音価系列である単旋律の場合には、(休符を除いた) 楽譜の音価の時系列と一致する。楽曲のリズム譜中のリズムパターンには、頻繁に

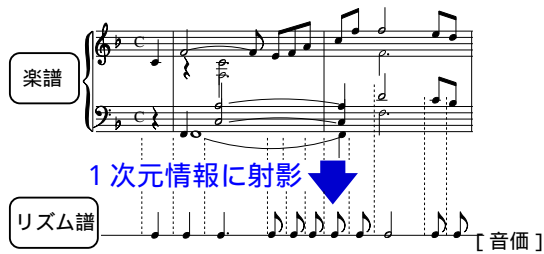


図 3 楽譜の 1 次元情報への射影：リズム譜

現れるありふれたパターンや、ほとんどあり得ないパターンなどが考えられることから、リズムパターンの出現には統計的な偏り・特性があると考えられる。我々は、この出現の統計的な性質を利用するため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、リズム譜全体の音価系列 $Q = \{q_1, \dots, q_T\}$ の出現確率は

$$P(Q) = P(q_1, \dots, q_T) \approx \prod_{t=n}^T P(q_t|q_{t-1}, \dots, q_{t-n+1}) \quad (2)$$

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。実際の楽曲では、パターンの繰り返し構造などの高次の構造を持つことが多いが、上記の式はそれらを含んでいないという点で近似表現となっている。このことは、音声認識でよく用いられる n -gram 言語モデル [12] の場合と同様である。

3.3 多声部にまたがる IOI

複数の旋律 (声部) が同時に演奏される場合、声部ごとの音長 (IOI) 系列を求めて楽譜を推定することは容易でない。各声部を分離してから IOI を得、複数の声部間で同期を取りながらリズム推定を行なうのは、複雑な処理となるためである。そこで、我々は、多声楽曲の演奏から明確に得られる観測量として、多声部にまたがる単一の時系列の IOI を用い (図 4)、各音の音

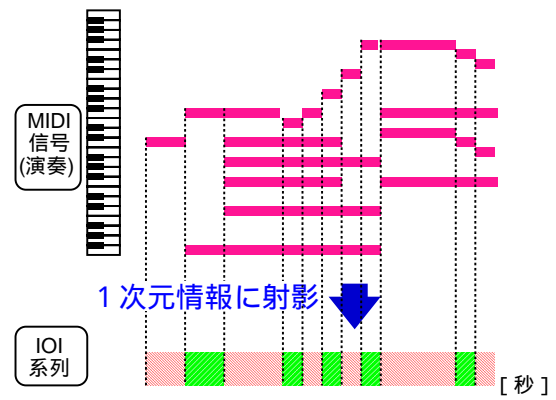


図 4 多声楽曲の IOI 系列

価の推定は別処理で行うことにした。

多声部にまたがる IOI は、リズム譜における音長に対応する。以後、便宜的にこのような多声部間 IOI を単に「音長」と呼ぶ。我々は、演奏の時間情報から楽譜への時間情報への対応付けを、音長系列とリズム譜との間で行う。

3.4 リズムベクトル

テンポが未知であり、また演奏の途中でテンポが変動しうる演奏を、本研究で対象としている。テンポが未知である演奏のリズム推定を行うため、テンポに依存しないリズムを表す特徴量を導入する。

テンポが曲の途中で変動しても、演奏のある一部分における局所的にはテンポはほとんど一定であると仮定する。連続する n 個の音長 (実際には IOI) の中でテンポ τ が一定 $\bar{\tau}$ の場合、音長 x の比は式 (1) より $\bar{\tau}$ に依存しない。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び、観測量として用いる。具体的には、リズムベクトル $r_t = (r_t^1, \dots, r_t^n)$ の各成分を

$$r_t^i = \frac{x_{t+i}}{x_t + \dots + x_{t+n-1}} \quad (3)$$

と定義する。これから、音長の時系列 $X = \{x_1, \dots, x_T\}$ に対するリズムベクトルの時系列 $R = \{r_1, \dots, r_{T-n+1}\}$ が計算される。

実際に観測される音長 (IOI) x およびテンポ τ は局所的にも厳密には一定でなく、リズムに変動 ϵ が含まれる。このため、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」リズムベクトルの周りに分布して観測される。ここで r の変動は確率分布に従うとし、今回は正規分布 $N(\mu, \Sigma)$ を用いる。正規分布のパラメータである平均 μ 、分散 Σ は、実演奏のリズムベクトルから統計的に適切な値を定める。音価 $s_t = \{q_t, \dots, q_{t+n-1}\}$ をリズムベクトル r_t として演奏する確率を、 $b_{s_t}(r)$ で表すことにする。

3.5 局所的テンポの変動

テンポは、局所的には一定と見なせる場合でも、大局的には変動しうる。ここでは、局所的なテンポを

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

とする。局所テンポの時系列 $\tau = \{\tau_1, \dots, \tau_{T-n+1}\}$ に対して、この変動 $\Delta\tau_t = \tau_t - \tau_{t-1}$ が確率分布 $\phi(\tau_t)$ に従うとする。

以上より、演奏者がリズム譜が Q である楽曲を演奏したとき、演奏の音長系列が X である確率が定式化される。 X から得られるリズムベクトルの時系列 R と、 X と Q から得られる局所テンポの時系列 τ を用いて

$$P(X|Q) = b_{s_1}(r_1) \cdot \prod_{t=2}^{T-n+1} b_{s_t}(r_t) \cdot \phi(\Delta\tau_t) \quad (4)$$

と表される。

3.6 リズムベクトルを用いた HMM

n 次元からなるリズムベクトルには n 個の音符に対応するので、この音符の組を状態としたモデルを考える。このとき、状態遷移をマルコフ過程とすることは、3.2 で述べた音符の $(n+1)$ -gram と等価である。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、リズムベクトルを出力とする HMM である (図 5)。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 $P(X|Q)P(Q)$ を計算できる。

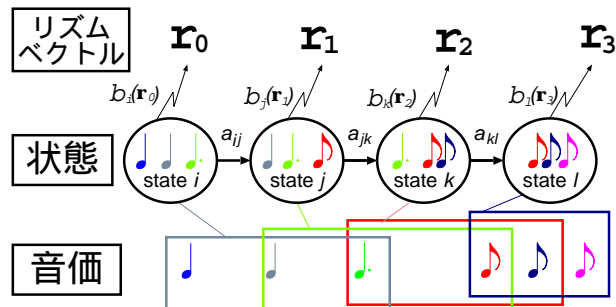


図 5 リズムベクトルを出力とし音符の組を状態とする HMM

4. 多声楽曲のリズム認識

HMM を用いて、声部にわたる音長からリズム譜を推定することができる。実際にリズム認識の確率モデルを用いて採譜を行うには認識の前後に処理が必要である。

4.1 逆問題

音長の系列 X からリズム譜 Q の推定は、可能性のある全ての Q から観測された X に対して HMM の中で最も尤もらしい \hat{Q} を求めることによって行われ、Bayes の定理を用いて以下のように定式化される。

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

ここで、式 (2)(4) を用いると、

$$\hat{Q} = \operatorname{argmax}_Q b_{s_1}(r_1) \cdot \prod_{t=2}^{T-n+1} a_{s_{t-1}, s_t} \cdot \phi_{s_t}(r_t) \quad (5)$$

と表せる。ただし、HMM の遷移確率 a_{s_{t-1}, s_t} は、音符の $(n+1)$ -gram の遷移確率 $P(q_t|s_{t-1})$ と同一のものである。 \hat{Q} を求めるには、HMM における最尤状態系列を求める Viterbi アルゴリズムを用いればよい。

4.2 モデルの学習

リズムベクトルの HMM の状態数は、 N 種類の音符の n 個の組である N^n (本実験では $N = 19, n = 3$ で状態数 6859) と

非常に大きい。このため、各状態についてモデルパラメータを学習するためには、必要な大量のデータが必要である。十分な学習データが得られない場合は、確率モデルのパラメータとして統計的に信頼できる値を推定できないので、限られた学習データからできるだけ多くのサンプルから得られる統計値を用いてパラメータの推定を行うように工夫する必要がある。

リズム: 音符の n -gram の学習

n -gram のパラメータの推定法として、より統計的な信頼性の高い低次の n -gram 確率を用いて高次の n -gram の遷移確率推定を推定する線形補間が知られている [12]。

$$\hat{P}(q_t|q_{t-1}, \dots, q_{t-n+1}) = \sum_{i=1}^n a_i P(q_t|q_{t-1}, \dots, q_{t-i+1})$$

本報告では a_i を実験的に定めて、線形補間を用いた。

リズムの演奏: リズムベクトルの確率分布の学習

リズムの変動の統計は、統計値を得るための統計のサンプル数を増やすため、状態ごとではなく音価ごとに求めた。演奏データと演奏曲のリズム譜を対応させることで、局所的にテンポが一定であるとしたときに計算される音価の変動分 ϵ を各音符ごとの統計を求めた。ここでも学習データで出現頻度の低い音価 q の ϵ の分散 $v(\epsilon; q)$ を推定するために、我々は線形の関係式 $v(\epsilon; q) = aq + b$ の a, b を推定する。これをもとに、リズムベクトルの確率パラメータを近似的に求めた。

4.3 楽譜推定の手順

リズムベクトルの HMM による推定を行う前後に処理を必要とする。実際には、次のような 3 段階の処理 (図 6) を行う。

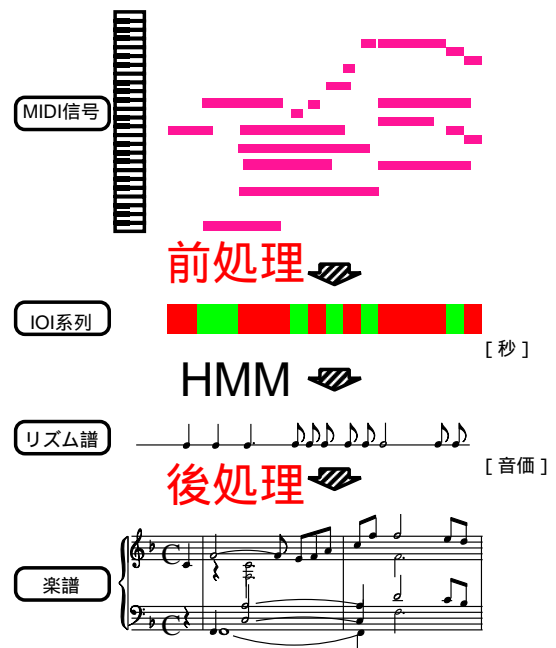


図 6 多声楽曲演奏 (MIDI 信号) の採譜の処理手順

Step 1: 声部間にまたがる音長系列の取得

声部間にまたがる音長 (IOI) 系列を得るため、まず同時発音を検出する。今回は、音長が閾値 δ 以下である場合に同時発音と判定する。これにより、MIDI 信号から音長の時系列 $X = \{x_1, \dots, x_T\}$ を得る。

Step 2: 音長に対応する音価を推定

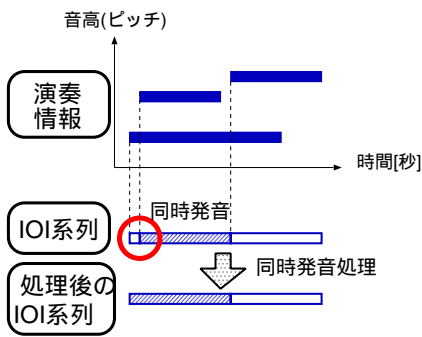


図7 同時発音の検出

リズムベクトルを用いたHMMにおいて式(5)により、音長の時系列 X に対応する音価列を推定する。Step 1 で得られた音長時系列 X から、リズム譜 $Q = \{q_1, \dots, q_T\}$ を得る。

Step 3: 各音の音価を推定 (後処理)

Step 2 で得たリズム譜をもとに、各音の継続時間からその音価を決定する。これにより、Step 1 の入力である MIDI 信号に対応する楽譜が得られる。今回は、音価を決定する手段として、表1で定められた分解能で量子化を用いた。

表1 各音の音価を量子化するときの分解能の設定

音長の音価	量子化単位
3連符	3連符単位
8分音符以上	8分音符単位
8分音符以下の短い	音長の音価

5. 評価実験

5.1 楽譜復元について

提案手法は既存の楽譜を復元することに特化したものではないが、自動採譜の性能評価を行うために客観的な「正解」の楽譜が必要であるため、演奏からの楽譜復元の実験評価を行った。但し、以下に述べる楽譜や演奏の特性のため、楽譜復元には原理的な限界がある。

(1) [楽譜の特性] 同一の音楽リズムに対して、複数の表記があり得る。例えば同一の演奏に対し、全部の音符の音価を倍にし、演奏のテンポも倍とする解釈が可能である。また、拍子の基準の音価の取り方にも任意性がある。例えば、3拍子のリズムは $3/8$ でも $3/4$ でも書け、さらに6拍子との区別も容易でない。

(2) [記譜習慣] たとえば図8の冒頭の最低音Fは、楽譜通りに長く鍵を押し続け得ないが、記譜習慣としては許され、しかもかなりの任意性がある。演奏からこのような楽譜を復元することは、原理的に困難である^(注1)。装飾音の記譜法も多様である。

5.2 実験条件

5人の演奏者が2曲を2回ずつ演奏した電子ピアノでの演奏をMIDI信号として記録し、これから演奏曲の音価の復元を試みた。演奏曲は、「フーガ」(J. S. Bach 作曲、平均律第1巻よ

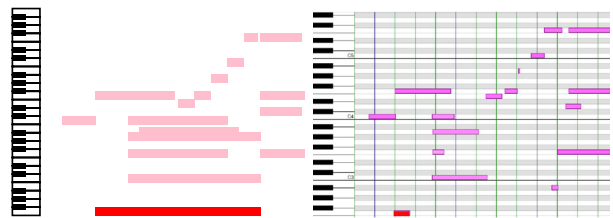


図8 実演奏と原楽譜の音符の継続時間の違い(トロイメライの冒頭部)

り八短調のフーガ)と「トロイメライ」(R. Schumann 作曲、組曲「子供の情景」より)である。「フーガ」はテンポが中庸で、16分音符の刻みが多くリズム譜は単純である。これに対して、「トロイメライ」はテンポが遅めで、リズム譜も複雑である。一般的にテンポが遅い曲は人間に取ってもリズムを認識しにくく、またテンポが遅い分だけ表情のある演奏をするためにテンポの変動が大きくなる傾向があるが、「トロイメライ」はこの傾向が当てはまる曲である。

Step 1 の同時発音の検出には、実際の演奏データを調べて $\delta = 0.04$ [秒] と定めた。この値は、人間が同時発音を区別する限界と言われている値に近い。HMMでは19種類の音価の3つ組を状態とし、特徴量として3次元のリズムベクトルを用いた。HMMの学習には、遷移確率の学習に学習データを、出力確率の学習に演奏データを用いた。楽譜データには13曲のクラシックのピアノ作品から得られたリズム譜(総音価数4355)を使用した。演奏データには「トロイメライ」の評価データの演奏者とは異なる奏者1名による3回の演奏を用いた。

5.3 結果

提案手法の評価のために、4.3で述べたStep毎に正解率を計算した。5.1で述べた楽譜推定の限界を考慮して、テンポ違いの音価推定は正解に含めた。実際に、フーガのリズム譜の推定結果は、全ての演奏が2倍の音価(倍テンポ)として推定された(図9)。複数の楽譜表現の例としては、スタッカート奏法を



図9 リズム譜の推定におけるテンポの多義性(正解に含める)

考慮する。実際のフーガの推定結果では図10のように音の継続時間が短くなった。これは、4分音符の音をスタッカート気味で演奏した結果、8分音符と8分休符の様に演奏したためであるので、正解と見なすことにした。以上を踏まえて各ステッ

スタッカートで演奏していると解釈できる
(正解と見なす)



図10 音価推定での正解音価判定の補正方法

プでの正解率を次式で計算した。その結果を表2および表??に示す。なお、各評価は全てDP(Dynamic Programming)による対応付けに基づいて行った。

(注1): 記譜法は各時代の楽曲様式や作曲家の習慣に依存する部分もあり、作曲家は意図を持って楽譜を書いているのであり、例えば $3/4$ と $3/8$ は厳密に同一ではないが、このような背景は本稿では扱わない。

- STEP 1 (同時発音の検出)
 (前処理処理の精度) = $\frac{(\text{演奏の総音数}) - (\text{脱落} \cdot \text{挿入} \cdot \text{同時発音の誤り})}{(\text{楽譜の総音符数})}$
- STEP 2 (リズム譜の推定)
 (リズム正解率)
 = $\frac{(\text{演奏の総音数}) - (\text{脱落} \cdot \text{挿入} \cdot \text{同時発音の誤り}) - (\text{リズム譜の音価の置換誤り})}{(\text{楽譜の総音符数})}$
- STEP 3 (音価復元率)
 (音価復元率)
 = $\frac{(\text{演奏音の数}) - (\text{脱落} \cdot \text{挿入誤り}) - (\text{音価の置換誤り})}{(\text{楽譜の総音符数})}$

ここで、各 Step での脱落誤り、挿入誤りは、演奏ミス (ミスタッチ) に相当する。

表 2 電子ピアノによる演奏からの楽譜復元の正解率 [単位: %]

演奏データ	フーガ			トロイメライ		
	同時発音処理	リズム正解率	音価復元率	同時発音処理	リズム正解率	音価復元率
A、1 回目	96.9	93.0	89.7	90.1	70.5	52.2
A、2 回目	97.8	95.0	91.2	92.8	52.6	41.6
B、1 回目	97.6	95.0	93.0	94.3	64.2	55.7
B、2 回目	98.0	95.0	93.4	95.8	78.8	58.4
C、1 回目	98.0	94.5	93.0	91.3	71.1	49.6
C、2 回目	97.8	94.3	93.0	96.2	71.1	54.6
D、1 回目	96.3	93.5	90.8	90.1	76.1	53.2
D、2 回目	96.9	93.5	91.3	92.3	67.3	53.5
E、1 回目	98.0	92.5	93.3	90.4	49.6	51.2
E、2 回目	98.1	95.0	93.6	94.3	59.7	51.0
平均	97.5	94.1	92.2	92.8	69.4	52.1

5.4 考 察

フーガでは、音価正解率が 92.2% であり、テンポが未知でありかつ変動する多声楽曲の演奏から採譜を行う場合に、提案手法が有効であることが示された。トロイメライの場合も、quantize による従来の楽譜推定よりは遥かに適切な楽譜が得られたが、音価推定率自体は 52.1% と低い。この原因としては次の 2 つが挙げられる。

一つの原因は、5.1 で述べたピアノ特有の奏法であるペダルの使用にある。ペダルを使用する場合は、音の継続時間が鍵盤を押さえている時間と異なるため、継続時間が正しく観測できない。図 8 に示したトロイメライの冒頭部の演奏に対応する認識結果を、図 11 に示すように音価の誤推定となる。ペダルへの対処はピアノ演奏に特有の課題であるが、今後、検討をしていきたい。もう一つは、トロイメライでは多用されている装飾



図 11 継続時間が短いためにおこる音価推定誤り

音について、本手法では十分に考慮していない点にある。装飾音も同時発音として処理されることを期待していたが、実際の演奏では装飾音は目立つように閾値 $\delta = 0.04$ 秒よりも長く演奏されることが多かった。装飾音を独立した短い音価として誤認

識し、途中から音価を 2 倍に誤推定する原因となった。但し、この問題はテンポ推定による後処理で解決可能である。

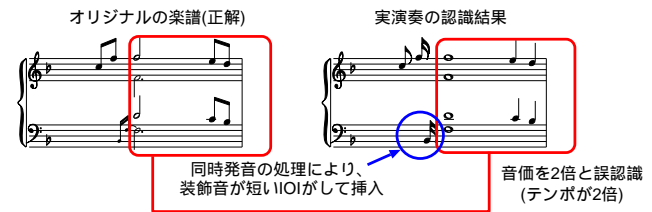


図 12 途中から倍テンポとなる誤推定

6. おわりに

本稿では、HMM を用いた多声部の楽曲構造を持つ演奏データ (MIDI 信号) から採譜を行う方法を提案した。必要な処理はリズム認識であり、声部間にまたがる IOI (音長) の時系列からリズム譜を推定する方法を提案した。推定には、テンポに依存しない時間情報を特徴量とし、リズムパターンの生成確率を含んだ HMM を用いた。実際の採譜手法として、HMM による認識を行う前に同時発音の検出、後に各音の音価推定を行った。電子ピアノの演奏の MIDI データに対して評価実験を行い、「フーガ」に対して 92.2%、「トロイメライ」に対して音価復元率 52.1% を得た。今後は、声部にまたがる IOI 以外の観測量も含めた確率モデルを用いて性能向上の目指したい。

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して下さった方々に感謝する。

文 献

- [1] R. Curtis: The Computer Music Tutorial, MIT Press, Cambridge, 1996.
- [2] 片寄, 井口, “知能採譜システム,” 人工知能学会誌, Vol. 5, No. 1, pp. 59-66, 1990.
- [3] 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.
- [4] 大規, 齊藤, 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.
- [5] H. Takeda, N. SAITO, T. OTSUKI, M. NAKAI, H. SHIMODAIRA, S. Sagayama, “Hidden Markov Model for Automatic Transcription of MIDI Signals,” Proc. MMSP, 2002.
- [6] L. Rabiner, and B.-H. Juang: Fundamentals of Speech Recognition, Prentice-Hall, 1993.
- [7] 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.
- [8] A. Cemgil, B. Kappen, P. Desain, H. Honing, “On tempo tracking: Tempogram Representation and Kalman filtering” Journal of New Music Research, 2000.
- [9] C. Raphael, “Automated Rhythm Transcription,” ISMIR, pp. 99-107, 2001.
- [10] 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.
- [11] 武田, 西本, 篠田, 嵯峨山, “HMM を用いた多声部 MIDI 信号からの楽譜復元,” 音講論集, 春季 3-7-3, pp. 837-838, 2003.
- [12] 北, “確率的言語モデル,” 東京大学出版会, 1999.