確率モデルによる多声音楽演奏のMIDI信号のリズム認識

武田晴登[†] 西本卓也[†] 嵯峨山茂樹[†]

本稿では，人間による器楽演奏の情報を保存したMIDI（Musical Instrument Digital Interface）信号からリズム認識を行う手法について述べる．演奏のテンポが未知であり多声部の構造を持つ楽曲のMIDI信号は従来リズム認識の対象として十分取り上げられてこなかったが，我々は演奏曲に対する事前知識を用いずにリズム認識を行う手法を提案する．演奏のテンポについての事前知識を用いずにリズムを推定するために，テンポに依存しない特徴量（リズムベクトル）を用いたHMM（Hidden Markov Model，隠れマルコフモデル）における事後確率最大化問題を解く．HMMのパラメータは既存の楽曲のリズムと演奏から事前学習により定められる．評価実験として5人のピアノ奏者による電子ピアノ演奏から楽譜推定を行い，リズム認識率として41.6〜94.1%，楽譜の音価の復元率として市販ソフトの14.4〜45.4%を上回る36.8〜92.2%を得た．

Rhythm Recognition of Multiphonic MIDI Signals Using Probabilistic Models

Haruto Takeda,[†] Takuya Nishimoto[†] and Shigeki Sagayama[†]

This paper describes a method of rhythm recognition for automatic transcription of human-performed MIDI (Musical Instrument Digital Interface) signals. For rhythm estimation without a priori knowledge about tempo, we solve maximum a posteriori problem in an Hidden Markov Model (HMM) using tempo invariant features. Parameters of HMMs are optimized through stochastic training with existing scores and performances before estimation. In experimental evaluation using MIDI performances by 5 players with an electronic piano, we obtained 41.6–94.1% accuracy for rhythm recognition and 36.8–92.2% accuracy for restoration of note values of the original scores.

1. まえがき

本稿では，人間の演奏（以下，実演奏）を記録したMIDI（Musical Instrument Digital Interface）信号から楽譜を自動的に書き起こすこと（いわゆる自動採譜[1),2)]）を目的としたリズムの認識について論じる．

実演奏のMIDI信号から演奏されたリズムの適切な楽譜表現を得るために「量子化」（quantization）と呼ばれる手法が，商用ソフトなどで広く用いられている．この手法が対象とするのは，メトロノームなどに従って演奏したテンポが既知で一定に保たれたMIDI信号である．実演奏の音長をユーザが指定した音価の最小単位（量子）の整数倍に対応させることにより音価を決定する．しかし，実演奏では演奏者はしばしば意図的に，あるいは無意識に音長を変動させるため，量子化はその音長の変動を直接に反映した音価を出力し，その結果演奏者の意図したリズムの楽譜表現が得られないことが多い．演奏者の意図した音価を得るには，機械的な正確さで音長を演奏し，かつ適切に音価の最小単位を定めなくてはならないが，それは熟練した鍵盤楽器奏者や熟練したコンピュータ利用者でなければ難しい．さらに，量子化することを意図せずに演奏した場合は，**図1**に示すように演奏者の意図とは異なる楽譜を出力してしまう．量子化の手法を改良する試みには，隣接する音長の比が有理数になれば安定するエネルギー関数を用いる手法[3)]や，演奏の発音位置を確率モデルで補正する方法[4)]が報告されている．

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信号に対しても，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[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，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[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．

† 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

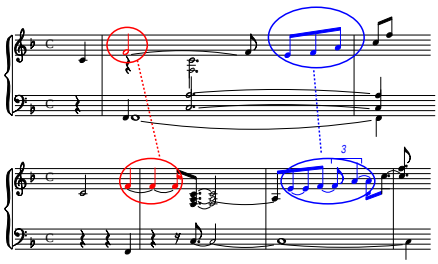

図 1 Träumerei の冒頭部の原譜（上）とその実演奏を市販楽譜作成ソフトにより量子化した結果例（下）
Fig. 1 The result of quantization by a commercial software (lower) compared with the original score (upper) of "Träumerei" played with an electronic piano.

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[8)~11)]．リズム認識と音声認識を同型の推定問題としてとらえ，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）[7)] を用いてモデル化を行った．しかし，これらの研究の対象は単旋律に限られていた．

本稿で提案するリズム認識の手法は，演奏曲のテンポや拍子の事前知識をまったく用いず，多声楽曲も対象に含む点が特徴である．また，自動採譜を目的としているので演奏された音のすべてに対して対応する音価を求める点も，先行研究と異なる．以下に，確率モデルを用いたリズム認識の方法について述べ，5 人のピアノ演奏者による MIDI ピアノ演奏の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. リズムベクトルを用いた HMM によるリズム認識

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．ここでは音価 q の値を**表 1** のように扱う．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．

表 1 楽譜から得られる音長情報：音価
Table 1 Information about note length from scores: note values.

音符						...
音価	2	$\frac{3}{2}$	1	$\frac{1}{2}$	$\frac{1}{3}$	...

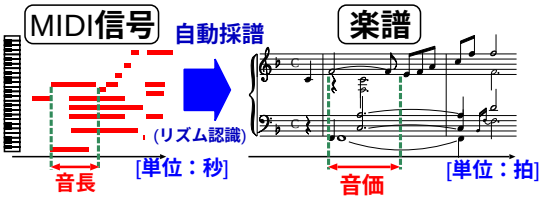

図 2 リズム推定：実演奏の音長から楽譜の「音価」を推定
Fig. 2 Two kinds of information about duration of sound: note values and note lengths.

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

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

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 2** 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 の積への分解ととらえられる．一般にこの分解は一意でない．たとえば，等時間間隔で手拍子を打ったときのリズムは，すべての音を 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．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定性があることを意味する．

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．これを音価の変動 ϵ ととらえ，式 (1) に反映させて，この関係を

$$x\,[\text{秒}] = \tau[\text{秒/拍}] \times (\,q + \epsilon\,)\,[\text{拍}] \qquad (2)$$

のように表せる．さらに，テンポ τ も変動するが，音符ごとにテンポが自由に変化しうるとすると音価の意味がないので，テンポ τ は少なくとも複数の音符にまたがって変動する項，あるいは局所的には一定と見なせる項と考えることにする．また，本稿では急激に

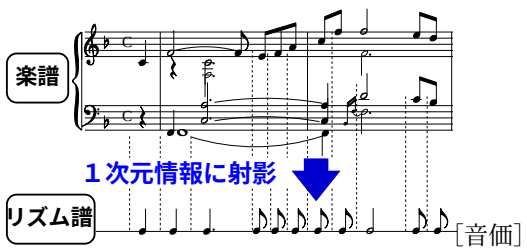

図 3 多声音楽のリズムを単旋律時系列に射影：リズム譜
Fig. 3 Rhythm score: projection of multiphonic rhythm sequence on to monophonic rhythm sequence.

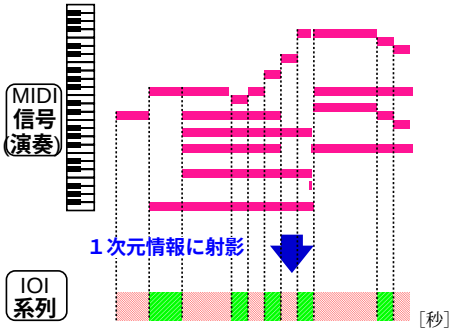

図 4 多声楽曲の演奏の声部間 IOI 系列
Fig. 4 An IOI sequence of performed music having polyphonic structure.

テンポが切り替わる楽曲は扱わないことにする．

2.2 リズム譜

多声音楽のリズムを扱うために，リズム譜を導入する．**図 3** に示すように，楽譜に記されているすべての音の発音位置に注目したとき，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．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考えてよい．我々の目的は多声音楽の複数の声部を構成するすべての音の音価を推定することであるが，問題を容易にするために，すべての音の音価を同時に推定するのではなく，まずリズム譜の推定を行って，その後にすべての音の音価を決定するアプローチをとる．

2.2.1 リズム譜の確率モデル

楽曲のリズム譜中のリズムパターンには，頻繁に現れるありふれたものや，ほとんどありえないものなどがあり，リズムパターンの出現には統計的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，フレーズの繰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が，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き確率 $P(q_t|q_{t-1},\cdots,q_{t-n+1})$ で近似できるとすると，リズム譜 $Q=\{q_1,\cdots,q_T\}$ の出現確率は

$$\begin{aligned} P(Q) &= P(q_1,\cdots,q_T) \\ &\approx P(q_1,\cdots,q_{n-1})\prod_{t=n}^{T} P(q_t|q_{t-1},\cdots,q_{t-n+1}) \end{aligned} \tag{3}$$

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．n-gram モデルは局所的なリズムパターンの特徴を表現するが，実際の楽曲にしばしば見られるパターンの繰返し構造などの高次の構造は反映していない．

2.2.2 音符の n-gram の学習

現実には，上述の確率を推定するために限られた学習データ量しか得られないことが多い．n-gram のパラメータの推定法として，統計的な信頼性がより高い低次の n-gram 確率を用いて高次の n-gram の確率を推定する線形補間が知られている．本報告では $n=4$ として

$$\begin{aligned} \hat{P}(q_t|q_{t-1},\cdots,q_{t-3}) &= a_0P(q_t)+a_1P(q_t|q_{t-1}) \\ &+a_2P(q_t|q_{t-1},q_{t-2})+a_3P(q_t|q_{t-1},\cdots,q_{t-3}) \end{aligned}$$

を用いた．ここで，補間係数は $\sum_{i=0}^{3} a_i=1$ を満たすものとする．

2.3 多声部間の IOI

演奏の情報の中でリズム譜に対応するものは，**図 4** に示すような多声部間の IOI 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，すべ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 は 0 になる．

多声部間 IOI を用いる利点は，楽曲に対する解釈に関係する声部の分離を必要としない点にある．各声部を分離してから IOI を得，複数の声部間で同期をとりながらリズム推定を行う方法は複雑な処理となるために容易ではない．そのため，我々は，多声部間の IOI からリズム譜を推定し，各音の音価の推定は別処理で行う．以後，便宜的にこの多声部間 IOI 系列を音長系列と呼ぶ．

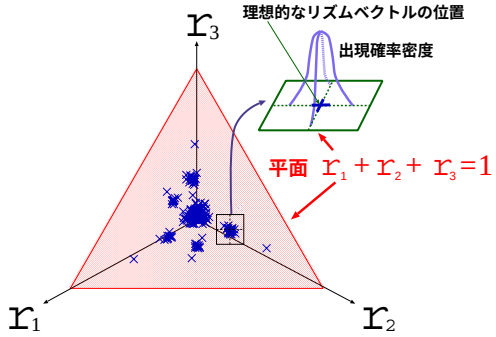

図 5 リズムベクトルの分布の例（$n=3$ の場合，リズムベクトルは 3 次元空間内の平面 $r_1+r_2+r_3=1$ 上の点として観測される）

Fig. 5 Distribution of rhythm vectors (In the case of $n=3$, rhythm vectors are observed as points on the plane $r_1+r_2+r_3=1$ in the 3D space).

2.4 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ために，テンポに依存しない特徴量（リズムベクトル）を用いる．演奏のテンポの変動は小さいので，連続する n 個の音長のテンポ τ はその区間内では一定 $\bar{\tau}$ と見なせる．式 (2) よりテンポ τ が一定と見なせれば，音長 x の比は音価 q の比を表しテンポ τ の値に依存しない．そこで，n 個の連続する音長 $x_t,\cdots,x_{t+n-1}$ 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 $\boldsymbol{r}_t=(r_t^1,\cdots,r_t^n)$ の第 i 成分を

$$r_t^i=\frac{x_{t+i-1}}{x_t+\cdots+x_{t+n-1}}$$

と定義する．

2.4.1 観測量としてのリズムベクトル

すでに式 (2) で述べたように，一般に実演奏の音長は音価に忠実とは限らず，変動成分 ϵ を含む．このため，実演奏のリズムベクトル $\boldsymbol{r}$ 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変動し分布する．実演奏の音長の時系列 $X=\{x_1,\cdots,x_T\}$ から求めた $n=3$ の場合のリズムベクトルの時系列 $\{\boldsymbol{r}_1,\cdots,\boldsymbol{r}_{T-n+1}\}$ は，**図 5** に示すように $\boldsymbol{r}_t$ が「理想的な」リズムベクトルの点に一致せず，その周辺に分布している様子が観測される．そこで，$\boldsymbol{r}$ の変動は確率分布に従うと仮定する．音価の n 個の組 $s_t=\{q_t,\cdots,q_{t+n-1}\}$ で表されるリズムをリズムベクトル $\boldsymbol{r}_t$ で演奏する確率を，$b_{s_t}(\boldsymbol{r})$ で表すことにする．b_{s_t} としては，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．

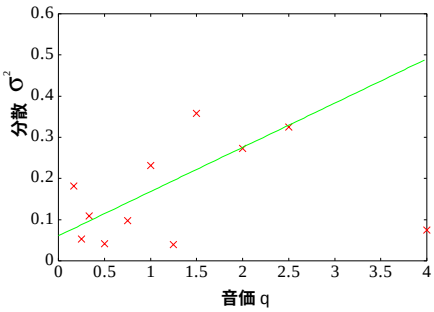

図 6 人間の演奏の音価 q と分散 ϵ の関係

Fig. 6 Relation of time value q and variance ϵ of human performances.

2.4.2 リズムベクトルの確率分布の学習

音符 n 個組の組合せの数は大きいので，すべての組合せに対応する実演奏の統計を得ることは容易でない．このため，それぞれの音符 n 個組 s に対応するリズムベクトルの変動確率の分散 Σ を以下に述べるような近似により求める．

実演奏の多声部間 IOI の各 x_t に演奏曲のリズム譜の音価 q_t が対応付けられているとする．リズムベクトルの各成分の間の相関は考慮せず，Σ は対角共分散行列として求める．まず，音長 n 個の組の平均のテンポ（以後，局所テンポと呼ぶ）

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

を求める．この局所テンポ $\bar{\tau}$ を演奏された IOI x のテンポとし，式 (2) から $q+\epsilon$ を計算し，各 q に対するリズムの変動 ϵ の統計を得る．この ϵ の統計は，平均 0，分散 σ^2 の正規分布に従うと仮定する．さらに，音価が大きくなるに従いその分散も大きくなると考えられるので，分散と音価の関係を次の 1 次式で近似する．

$$\sigma_\epsilon^2(q)=\alpha\cdot q+\beta \tag{4}$$

サンプルから得られた ϵ から最小二乗法により定数 α，β を求め (**図 6**)，式 (4) により音価 q に対応する分散を求める．最後に，対角共分散行列 $\Sigma=\mathrm{diag}(\sigma_1^2,\cdots,\sigma_n^2)$ を

$$\sigma_i^2\approx\frac{\sigma_\epsilon^2(q_i)}{\sigma_\epsilon^2(q_1)+\cdots+\sigma_\epsilon^2(q_n)}$$

による近似で求める．

なお，式 (4) による近似では音価個有の音長変動の情報が損なわれるので，本来ならこの近似を用いずに ϵ の統計から直接に分散値を推定することが望ましい．この近似を用いる理由は，学習データにほとんど出現しない音価の分散値を出現頻度の高い音価の統計を用

いて推定値の統計的信頼性を補うためである．しかし，たとえばバロック音楽や古典派，ロマン派などのクラシック音楽作品でしばしば表れる符点 8 分音符と 16 分音符から成るリズムは，16 分音符のタイミングを速めたり遅くしたり☆して表情を変化させて演奏されることがしばしばあるので，演奏曲や演奏スタイルによっては符点 4 分音符と 16 分音符の音長変動は他の音価の音長変動より大きくなる．このような演奏には式 (4) の近似は不適切である．本稿で扱う演奏は，このような奏法を行わないものとする．

2.5 局所的テンポの変動

さらに，テンポの変動についても確率モデルを導入する．演奏曲のテンポ自体は未知であるが，テンポ変動は小さいことを仮定して，局所テンポ τ_t の変動 $\Delta\tau_t = \tau_t - \tau_{t-1}$ は，0 を平均とした正規分布 $\phi(\tau_t)$ に従うものと仮定する．

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 $\{\boldsymbol{r}_t\}$，X と Q から得られる局所テンポの変動の時系列を $\Delta\tau_t$ として，

$$P(X|Q) = b_{s_1}(\boldsymbol{r}_1) \cdot \prod_{t=2}^{T-n+1} b_{s_t}(\boldsymbol{r}_t) \cdot \phi(\Delta\tau_t) \quad (5)$$

である．

2.6 リズムベクトルを用いた HMM

リズムベクトル $\boldsymbol{r}$ 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 $\boldsymbol{r}$ を状態 s からの出力とするモデルを考える．この状態の遷移をマルコフ過程とすると，2.2 節で述べた音符の $(n+1)$-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 7**）．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．

3. 多声楽曲のリズム認識

3.1 逆 問 題

音長の系列 X からリズム譜 Q の推定は，可能性のあるすべての 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．

☆ 音価どおりの音長比 3/4 : 1/4 ではなく，3 連符で表される 2/3 : 1/3 や複付点音符で表される 7/8 : 1/8 などの音長比で演奏される場合がある．

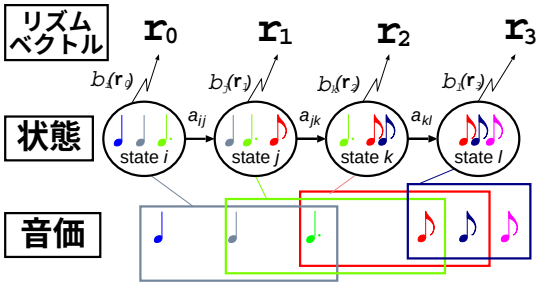

図 7 リズムベクトルを出力とし音符の組を状態とする HMM
Fig. 7 Rhythm vectors as outputs of an HMM.

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

ここで，式 (3)，(5) を用いると，

$$\hat{Q} = \operatorname*{argmax}_Q \pi_{s_0} b_{s_1}(\boldsymbol{r}_1) \prod_{t=2}^{T-n+1} a_{s_{t-1}s_t} b_{s_t}(\boldsymbol{r}_t) \cdot \phi_{s_t}(\Delta t_k) \quad (6)$$

と表せる．ただし，$a_{s_{t-1}s_t}$ は HMM の遷移確率，π_{s_0} は初期状態の出現確率を表し，音符の $(n+1)$-gram のパラメータとは

$$\pi_{s_0} = P(q_1, \cdots, q_n)$$
$$a_{s_{t-1}s_t} = P(q_t|s_{t-1})$$

のように対応する．HMM における最尤状態系列 $\hat{Q}$ を求めるには，効率的な探索を可能にする VDA（Viterbi Decoding Algorithm：ビタビ復号化アルゴリズム）[12] を用いる．

3.2 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行う．実際には，次のような 3 段階の処理（**図 8**）を行う．

Step 1：**音長系列（声部間 IOI）の取得**

音長系列（声部間 IOI 系列）を得るため，まず同時発音を検出する．同時打鍵を意図しても打鍵時刻がずれることが多いので，今回は，音長（声部間 IOI）が閾値 δ_s 以下である場合に同時発音と判定する．また，トリル，ターンは音高差（音程）δ_n と発音時刻間隔 δ_t を閾値により検出し☆☆，検出した装飾音を取り除いた音の発音時刻から IOI を求める．複数の発音時刻が同時発音であると検出された場合，最初の発音時刻を用いて IOI を計算する．これにより，MIDI 信号から音長系列 $X = \{x_1, \cdots, x_T\}$ を得る（**図 9**）．

☆☆ 連続する音の音高差のみから判断するので，トリルが記符された音符に対し装飾音符（上の音）から開始するバロック音楽の奏法にも対応している．

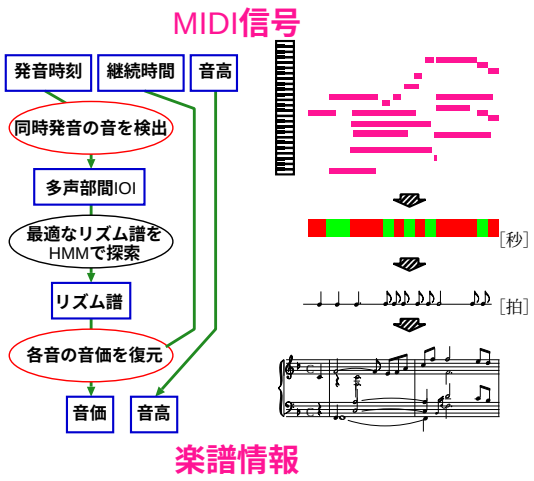

図 8 多声楽曲演奏（MIDI 信号）の採譜の処理手順
Fig. 8 Procedure of automatic rhythm transcription of MIDI signals.

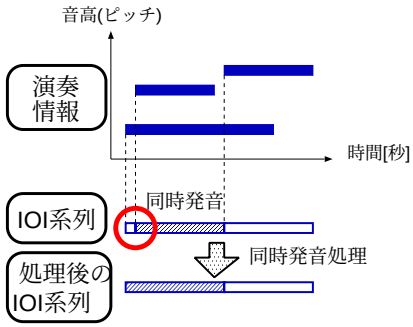

図 9 同時発音の検出
Fig. 9 Detection of simultaneous onset time.

Step 2：**リズム譜の推定**

リズムベクトルを用いた HMM において式 (6) により，音長系列 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 $Q = \{q_1, \cdots, q_T\}$ を得る．

Step 3：**各音の音価を推定（後処理）**

Step 2 で得たリズム譜と，MIDI 信号中の各音の音高と，その継続時間から，その音の音価を決定して楽譜を生成する．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，音の継続時間から音価を決定するため継続時間の量子化を行ったが，量子化の分解能は当該音のリズム譜中の音価に依存して与えた．すなわち，8 分音符以上ならば分解能は 8 分音符，8 分音符以下ならばその音長の音価を分解能とした．これは，継続時間を反映し過ぎて複付点音符などの複雑な音符を生成することを抑えるためである．

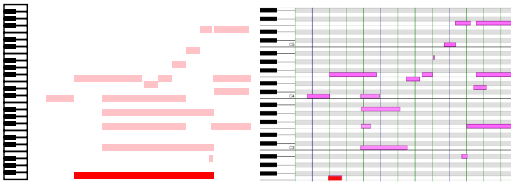

図 10 原楽譜（左）と実演奏（右）の音符の継続時間の違い（Traüumerei の冒頭部のピアノロール表示）
Fig. 10 Difference of duration (beginning of "Traüumerei").

3.3 評価実験

3.3.1 楽譜復元の原理的限界

リズム認識手法の性能評価を行うには，認識結果と比較に使用する客観的な「正解」の楽譜が必要である．今回は，既存の楽曲の演奏から楽譜復元による評価実験を行った．ここで，楽譜復元を行う場合に，原理的に避け難い誤りについて以下に述べる．

- オリジナルの楽譜とは異なるが等価な情報を与える楽譜を推定する「誤り」がある．これは，たとえば 3/4 で表される 3 拍子のリズムは 3/8 でも表現可能である☆ように，1 つの音楽情報を表記する方法が一意でないからである．2.1 節で述べた音価とテンポについての不確定性もこの例に含まれる．また，スタッカート奏法のように，楽譜にない短い休符が演奏では挿入される場合は，休符を記譜するかどうかに任意性がある．このような点で，オリジナルの楽譜とは異なるが等価な情報を持つ推定結果は，採譜の評価としては正解と見なされるべきである．
- 演奏がオリジナルの楽譜から大きく逸脱している場合は，楽譜の復元が困難である．本実験では，MIDI 信号に記録された継続時間（鍵盤を指で押している時間）が楽譜に記載されている音価に比べて極端に短く演奏される場合があった．トロイメライの冒頭のバスの第 1 音は，楽譜では 5 拍伸ばす（**図 10** 左の最低音）ことを指示されているが，ある実演奏では 1 拍程度の継続時間（図 10 右の最低音）で演奏されている．この継続時間からもとの楽譜の音価を復元することは難しい．なお，この演奏ではペダル奏法を用いることにより例のバスの音は 1 拍以上の長さの響きを保っているが，ペダルを踏んでいる区間を単純に継続時間に置き換えも問題は解決しない．なぜなら，ペダ

☆ 記譜法は各時代の楽曲様式や作曲家の習慣を反映し，作曲者は意図を持って楽譜を書いているのであって，3/4 と 3/8 は厳密には同一ではないが，このような背景は本稿では扱わない．

表 2 リズム認識評価実験に使用した評価データ
Table 2 Testing data for rhythm recognition experiments.

本稿での呼称	作曲者	曲名	データ数（演奏者数）	リズム譜の音価数
Fuga	J. S. Bach	平均律第 1 巻よりハ短調のフーガ BWV847	10 (5)	402
Sonata	L .v. Beethoven	ピアノソナタ 20 番ト長調 op.49-1 第 1 楽章前半	10 (5)	462
Träumerei	R. Schumann	組曲「子供の情景」op.15, No.7 トロイメライ（夢）	10 (5)	226

表 3 HMM の出力確率の学習に用いたデータ
Table 3 Performed MIDI data for training HMM output probabilities.

作曲者	曲名	収集データ数（演奏者）	リズム譜の音価数
C. Debussy	前奏曲第 1 巻より第 10 曲「亜麻色の髪の乙女」	4 (A,D)	242
R. Schmann	組曲「子供のためのアルバム」より「たのしき農夫」Op. 68 No.10	2 (A)	160

ルによって音を伸ばすことを意図した音の識別が必要であるからである．この例ではペダルによって音価分の響きを保とうとしている対象はバスであり，ソプラノを含まないことを推定する必要がある．この処理は Step 3 において行われるのが好ましいが，今回はこの処理は行わないことにした．

3.3.2 実験条件

5 人の演奏者が 3 曲を 2 回ずつ演奏した電子ピアノの演奏を 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，**表 2** に示す 3 曲である．5 人の奏者 A，B，C，D，E のうち，D，E は音楽大学を卒業しており，5 人とも Fuga，Sonata は学習したことがあり，Träumerei（トロイメライ）は今回の収録で初めて演奏した．Fuga 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．Sonata は装飾音（トリルやターン）が出現する．Träumerei 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と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にするためにテンポの変動が大きくなる傾向があるが，Träumerei はこの傾向があてはまる曲である．

Step 1 の同時発音の検出には実演奏を調べて $\delta_s = 0.04$[秒] と定めた．この値は，人間が同時発音を区別する限界といわれている値に近い．トリル，ターンの検出は音高差 $\delta_n = 2$（長 2 度），時間差 $\delta_t = 0.04$ を用いた．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 の遷移確率の学習を行うために，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）の統計を使用した．線形補間の係数 a_i の値は，予備実験により 10 種類の組合せの中から，高い性能の得られた値を選び，$a_0 = 0.4$，$a_1 = 0.1$，$a_2 = 0.1$，

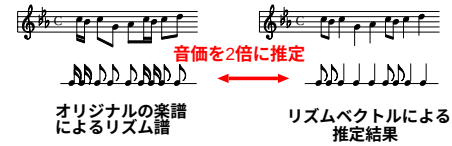

図 11 リズム譜の推定におけるテンポの多義性（正解に含める）
Fig. 11 Note values are estimated as half of that of original score.

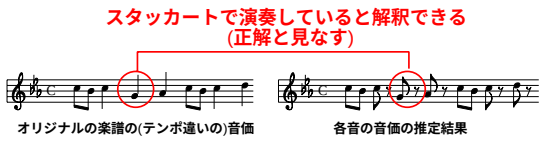

図 12 音価推定での正解音価判定の補正方法
Fig. 12 Estimation in different tempo (count correct).

$a_3 = 0.4$ とした．また，HMM の出力確率のパラメータであるリズムベクトルの分散の学習には，**表 3** に示す演奏から得られる IOI の統計を用いた．式 (4) におけるパラメータ値として $\alpha = 0.1074$，$\beta = 0.0608$ が推定された．ここでは，音価に依存する変動要因が主要項であるように，実験では音価に依存しない変動要因を小さくし，$\alpha = 0.1$，$\beta = 0.002$ とした．

3.3.3 実験結果

提案手法の評価のために，3.2 節で述べた Step ごとに正解率を計算した．本手法が MIDI データから抽出している情報の精度を評価するために，3.3.1 項で述べた楽譜記述の任意性を考慮して，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，**図 11** のようにすべての演奏が 2 倍の音価（倍テンポ）として推定された．複数の楽譜表現の例としては，Step 3 においてスタッカート奏法を考慮する．実際のフーガの推定結果では**図 12** のように音

表 4 市販ソフト Finale の量子化を用いたときの音価復元率［単位：%］
Table 4 Accuracy of note values using commercial software Finale [%].

演奏曲	音価復元率
Fuga (Bach)	45.4
Sonata (Beethoven)	18.8
Träumerei (Schumann)	14.4

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 分休符のように演奏したためであるので，正解と見なすことにした．以上をふまえて各 Step での正解率を次式で計算した．

- Step 1（同時発音の検出）
  同時発音処理の正解率 $= \frac{N'-D-I}{N}$
- Step 2（リズム譜）
  リズム認識率 $= \frac{n'-D-I-S}{n}$
- Step 3（演奏データの各音価）
  音価復元率 $= \frac{N'-D-I-S}{N}$
  I: 挿入誤り
  S: 置換誤り
  D: 削除誤り
  N': 実演奏 MIDI 信号の音数
  N: オリジナル楽譜の音価数
  n: オリジナル楽譜のリズム譜の音価数
  n': 演奏曲の多声部間 IOI の音価数

Step 1 に置ける誤りは，同時発音として処理する・しないの判定を誤った音の個数を表し，Step 2，Step 3 における誤りは音価の推定誤りを表す．誤りの個数は，各 Step で DP（Dynamic Programming）による自動計算による評価で得た挿入，脱落，置換誤りの合計値を用いている．本実験の Step1，3 では，評価データに含まれている演奏ミス（ミスタッチ）の影響☆を受けている．

参考のため，評価データから曲ごとに 1 つの演奏（演奏者 A，1 回目）を選び，市販ソフトの量子化機能を用いたときの演奏データのすべての音の音価の正解率を求めた．TEMPO=120 で演奏を記録した評価データの MIDI 信号を，量子化機能の最小単位を 16 分音符とし，3 連符の出現も許すと設定して量子化を行った．その結果は，**表 4** に示すように正しい音価がほとんど推定できなかった．これは，評価実験と同じ条件としてテンポの事前知識を用いていないため，適切でないテンポを基準にして処理を行っているためである☆☆．

表 5 電子ピアノによる演奏からの楽譜復元の正解率［単位：%］
Table 5 Accuracy of score recovering [%].

	Fuga (Bach)		
演奏データ	同時発音処理	リズム認識率	音価復元率
A，1 回目	96.9	93.0	89.7
A，2 回目	97.8	95.0	91.2
B，1 回目	97.6	95.0	93.0
B，2 回目	98.0	95.0	93.4
C，1 回目	98.0	94.5	93.0
C，2 回目	97.8	94.3	93.0
D，1 回目	96.3	93.5	90.8
D，2 回目	96.9	93.5	91.3
E，1 回目	98.0	92.5	93.3
E，2 回目	98.1	95.0	93.6
平均	97.5	94.1	92.2

	Sonata (Beethoven)		
演奏データ	同時発音処理	リズム認識率	音価復元率
A，1 回目	99.4	43.7	38.7
A，2 回目	97.6	42.2	38.8
B，1 回目	97.0	37.7	23.0
B，2 回目	97.2	42.9	39.6
C，1 回目	97.2	37.5	35.0
C，2 回目	97.4	41.6	39.7
D，1 回目	97.8	42.5	35.9
D，2 回目	97.8	43.3	35.6
E，1 回目	97.2	43.7	39.7
E，2 回目	97.9	40.9	42.2
平均	97.7	41.6	36.8

	Träumerei (Schumann)		
演奏データ	同時発音処理	リズム認識率	音価復元率
A，1 回目	90.1	70.5	52.2
A，2 回目	92.8	52.6	41.6
B，1 回目	94.3	64.2	55.7
B，2 回目	95.8	78.8	58.4
C，1 回目	91.3	71.1	49.6
C，2 回目	96.2	71.1	54.6
D，1 回目	90.1	76.1	53.2
D，2 回目	92.3	67.3	53.5
E，1 回目	90.4	49.6	51.2
E，2 回目	94.3	59.7	51.0
平均	92.8	66.1	52.1

提案手法の評価実験結果として，5 人の演奏者 A，B，C，D，E の演奏の各 Step での正解率を**表 5** に示す．リズム認識率として 41.6〜94.1%，楽譜の音価の復元率として 36.8〜92.2%を得た．

☆ 各演奏者のミスタッチ（MIDI 信号の発音情報でオリジナルの楽譜と対応の付かないものの個数）の割合は，平均で A 0.4%，B 0.6%，C 1.6%，D 0.8%，E 0.7%であった．

☆☆ ただし，3.3.4 項で述べているテンポ違いについては正解として数えた．

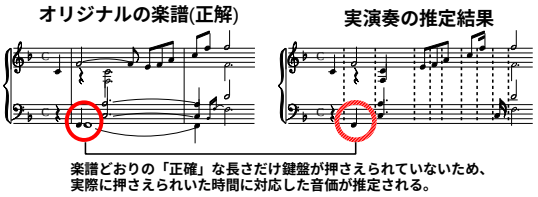

図 13 継続時間が短いために起こる音価推定誤りの例
Fig. 13 Misrecognized notes due to short durations.

3.3.4 実験結果についての考察

Fuga では，音価復元率が 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．Träumerei や Sonata の場合も，quantize による従来の楽譜推定よりははるかに適切な楽譜が得られたが，音価推定率自体は低い．この原因としては次の 2 つが考えられる．

1 つの原因は，3.3.1 項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 10 に示した Träumerei の冒頭部の演奏に対応する認識結果は，**図 13** 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の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 1 つは，装飾音について，本手法では十分に考慮していない点にある．Träumerei に見られる前打音，Sonata に見られるターンなど，これらの装飾音は Step 1 の閾値処理で同時発音として処理されることを期待していたが，実際の演奏では閾値よりも長く演奏されることが多く，今回適用した閾値処理では十分に対処できない．

音長の比を特徴量としているため，1 カ所音価の推定を誤ると，後続する音の音価の比を正しく推定しても音価の推定としては誤りになる．これは，本来ならほとんど一定であるはずのテンポが，音価の推定誤りを起こした後に音長比を正しく推定した場合には**図 14** 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．音価推定誤りの連鎖から起きるテンポ誤りは，本稿では詳細を述べないが曲全体のテンポを推定することにより検出可能である．

3.4 今後の展望

3.4.1 テンポの推定への応用

本手法は自動採譜を目的としているが，演奏曲のテンポの推定にも応用することができる．テンポとリズムには式 (1) の関係があるので，推定したリズム譜の音価をもとに演奏のテンポの時系列を求めることができる．リズムベクトルを用いた音価推定の誤り傾向を考慮したテンポの確率モデルを用いて，曲全体のテンポを推定することができる．

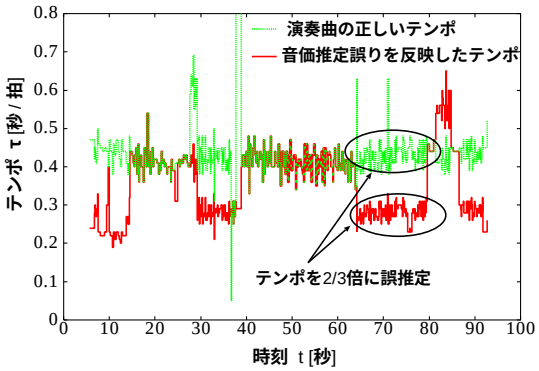

図 14 テンポの誤推定の例（演奏データ A1 の Sonata）
Fig. 14 Misrecognition of tempo caused by misrecognition of time values.

また，本稿ではリズム認識を音長から音価推定として扱ったが，式 (1) からリズム認識を音長からリズムとテンポへの分離と解釈を広げることができる．2 変数をそれぞれ適切な値に定めるために，リズムとテンポの推定を交互に繰り返す方法が考えられる．その場合，本稿で提案した手法は，推定のための初期値を与えることができる．予備実験として，評価実験で用いた Sonata の演奏に対して，提案手法を用いてリズムを推定した後にテンポ推定を行いさらにリズム譜の再推定を行ったところ，リズム認識率は最も性能の良いもので 92.1%，平均 60.6%となった．

3.4.2 複数のリズム譜を用いたリズム認識

本稿で提案したリズム譜は，複雑な多声部のリズム情報を一次元の時系列として扱えるという利点があったが，単声部にしたためにかえって複雑になるリズムも存在する．たとえば，八分三連符と八分音符が同時に演奏される場合は，2 つのリズムの重ね合わせであるリズム譜よりも，個別にリズム譜推定を行う方が好ましい．このような推定を行うには，演奏の各旋律に注目し，各旋律において同時発音処理を行いリズム譜を得て，そのリズム譜の間で同期をとりながらリズム認識を行えばよく，今後検討していきたい．

4. おわりに

本稿では，テンポが未知である多声楽曲の MIDI 演奏を対象としたリズム認識手法について述べた．多声楽曲のリズムを扱うためにリズム譜を導入し，多声部間 IOI からリズム譜を推定する問題を，テンポに依存しない特徴量と音価の n-gram 文法を含んだ HMM における事後確率最大化問題として定式化した．実際

にリズム認識を行うために，HMM における探索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，リズム認識率として 41.6〜94.1%，楽譜の音価の復元率として 36.8〜92.2%を得た．

参 考 文 献

1) 片寄，井口：知能採譜システム，人工知能学会会誌，Vol.5, No.1, pp.59–66 (1990).
2) 柏野，中臺，木下，田中：音楽情景分析の処理モデル OPTIMA における単音の認識，電子情報通信学会論文誌，D-II，Vol.J79-D-II, No.11, pp.1751–1761 (1996).
3) Desain, P. and Honing, H.: The Quantization of Musical Time: a Connectionist Approach, *Comp. Mus. J.*, Vol.13, No.3, pp.56–66 (1989).
4) Hamanaka, M., Goto, M., Asoh, H. and Otsu, N.: Learning-Based Quantization: Estimation of Onset Times in a Musical Score, *Proc. SCI 2001*, Vol.X, pp.374–379 (2001).
5) Cemgil, A., Kappen, B., Desain, P. and Honing, H.: On tempo tracking: Tempogram Representation and Kalman filtering, *Jounal of New Music Research* (2000).
6) Raphael, C.: Automated Rhythm Transcription, *Proc. ISMIR*, pp.99–107 (2001).
7) Rabiner, L. and Juang, B.-H.: *Fundamentals of Speech Recognition, Prentice-Hall* (1993).
8) 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，情報処理学会研究報告，99-MUS-33, pp.27–32 (Dec. 1999).
9) 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).
10) Takeda, H., Saito, N., Otsuki, T., Nakai, M., Shimodaira, H. and Sagayama, S.: Hidden Markov Model for Automatic Transcription of MIDI Signals, *Proc. MMSP* (2002).
11) 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，情報処理学会研究報告，2002-MUS-46, pp.23–28 (2002).
12) Viterbi, A.J.: Error bounds for convolutional codes and an asymptotically optimum decodeing alogorithm, *IEEE Trans. Inf. Theory*, Vol.IT-13, pp.260–129 (1967).

(平成 15 年 7 月 10 日受付)
(平成 16 年 1 月 6 日採録)

武田 晴登（学生会員）

2001 年慶應義塾大学理工学部卒業．2003 年東京大学大学院情報理工学系研究科修士課程修了．現在，同大学院同専攻博士課程に在籍．音楽情報処理に興味を持つ．

西本 卓也（正会員）

1993 年早稲田大学理工学部卒業．1995 年同大学大学院理工学研究科修士課程修了．1996 年京都工芸繊維大学工芸学部助手．2002 年東京大学大学院情報理工学系研究科助手．音声インタフェース，音声対話システムの研究に従事．日本音響学会，電子情報通信学会，人工知能学会，ヒューマンインタフェース学会会員．

嵯峨山茂樹（正会員）

1948 年，兵庫県生まれ．1972 年東京大学工学部計数工学科卒業．1974 年同大学大学院工学系研究科計数工学専攻修士課程修了．同年日本電信電話公社に入社，武蔵野電気通信研究所にて音声情報処理の研究に従事．1990 年 ATR 自動翻訳電話研究所音声情報処理研究室長として自動翻訳電話プロジェクトを遂行．1993 年 NTT ヒューマンインタフェース研究所にて音声認識・合成・対話の研究開発に従事．1998 年北陸先端科学技術大学院大学情報科学研究科教授．2001 年東京大学大学院工学系研究科のち情報理工学系研究科教授．博士（工学）．1990 年発明協会発明賞，1994 年日本音響学会技術開発賞，1995 年情報処理学会山下記念研究賞，1996 年科学技術庁長官賞（研究功績者表彰）および電子情報通信学会論文賞等を受賞．日本音響学会，電子情報通信学会，IEEE，ヨーロッパ音声通信学会（ESCA），AVIRG 各会員．IEEE Trans. Audio and Speech Processing の Associate Editor.