音楽情報科学 42−10
（２００１．１０．２７）

HMMと音符n-gramを用いた音楽リズム認識

大槻 知史† 中井 満‡ 下平 博‡ 嵯峨山 茂樹†‡

†東京大学大学院情報理工学系研究科 (〒113-8656 東京都文京区本郷 7-3-1)
‡北陸先端科学技術大学院大学情報科学研究科 (〒923-1292 石川県能美郡辰口旭台 1-1)
E-mail: {t-otsuki,sagayama}@hil.t.u-tokyo.ac.jp, {mit,sim,sagayama}@jaist.ac.jp

本稿では，隠れマルコフモデル (HMM) を用いて，人間が鍵盤入力した演奏情報 (スタンダード MIDI ファイル) の音価の系列から，意図された音符列を復元推定する手法に，音符 n-gram を用いてより強い音楽制約を与える方法論を導入する．従来は，ある状態は直前の状態のみに依存するという仮定のもとで，音符状態の連鎖である楽曲の生成確率を与えていたが，その結果音楽的な音符列のモデルとしては拘束力が弱かった．本稿ではより制約の大きい n-gram 遷移確率を利用したモデルを提案し，そのための 2 種類の Viterbi 計算のアルゴリズムも提案する．また楽曲サンプルの音価の trigram・quadgram 連鎖確率を学習することで，より良い推定結果が得られることを示す．

Musical Rhythm Recognition Using HMM and Note N-gram

Tomoshi Otsuki†, Mitsuru Nakai‡, Hiroshi Shimodaira‡, and Shigeki Sagayama†‡

†Graduate School of Information Science and Technology, The University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656 JAPAN
‡Graduate School of Information Science, Japan Advanced Institute of Science and Technology
1-1, Asahi-dai, Tatsu-no-kuchi, Ishikawa 923-1292 JAPAN
E-mail: {t-otsuki,sagayama}@hil.t.u-tokyo.ac.jp, {mit,sim,sagayama}@jaist.ac.jp

This paper proposes the use of musical note n-gram in Hidden Markov Model (HMM) for rhythm recognition from musical performance signal recorded in the standard MIDI file format. In the previous paper, we used the bigram model with transition probabilities between adjacent notes which had rather weak constraint over musical note sequences. In this paper, we propose a more constrained model using n-gram transition probabilities in HMM, and two algorithms for n-gram Viterbi path calculation. Experimental results show that trigram and quadgram transition probabilities trained by music sample phrases give better performance.

1 序論

自動採譜，すなわち音楽演奏の音響信号から意図された楽譜を復元する逆問題は，多重音の分離，音色 (楽器)，音高，音価を特定するなどの多くの興味深い側面を持ち，単純に解決できる問題ではない．一方，多少の訓練を受けた人間ならば，簡単な音楽なら容易に楽譜化できる場合は多い．本稿では，そのような機能を工学的に実現することを目標に，自動採譜の一要素として音価の同定の問題に絞って考えたい．

この問題は，単独でも実際的な意味がある．たとえば，MIDI キーボード (以下,「鍵盤」という) で単旋律を演奏した場合を考えると，得られた MIDI 情報から，音色，音高の情報は正確に得られるので，音価，拍子，拍節，テンポなどが特定できれば楽譜

化が可能である．音楽の作・編曲，演奏，教育・出版などでは，楽譜の浄書や MIDI 演奏を目的に，楽譜をコンピュータへ投入する機会が多く，グラフィカルなインタフェースを備えたソフトウェアツールも普及している．さらに，鍵盤で演奏した音楽がそのまま楽譜化できれば大変便利である．また, MIDI ファイルの形でのみ存在する楽曲の楽譜化にも利用できる．

この音価を復元する手法としては，従来，閾値処理をベースとしてボトムアップ的に音符列を対応づける研究 [1, 2, 3, 4] が知られていたが，演奏者の意図を正確に復元することは難しかった．

我々は，同種の問題を扱っている連続音声認識分野の方法論 [5, 6] をヒントにし，この音符列復元の問題に対して Hidden Markov Model を用いたトップダウンアプローチによって，音符列推定 (リズム認識)，演奏テンポ推定，拍子・拍節認識が定式化できることを示した [7, 8]．ここでは楽曲の生成確率を与える際に，個々の音符の連鎖確率で表すモデル，および 1 小節単位の一連の音符の並びの連鎖確率で捉えるモデルの 2 種類を導入した．しかし，前者の場合は音楽的な拘束力が弱く，後者の場合は 1 小節のリズムパターンが膨大になって学習で得られなかった音符列は推定することができないために，リズムパターンが複雑な曲に対して本手法を適用することは困難であった.

本研究では，楽曲の生成確率の音楽的な制約を強めるために，状態遷移の確率を直前の状態の履歴のみから与えるのではなく，2 個以上の履歴を考慮した n-gram 遷移確率を用いて与える手法を提案する．

本稿では，まず n-gram を用いた HMM の学習・認識モデルを提案し，また，この n-gram 遷移確率を用いた Viterbi 探索を実行するための 2 種類のアルゴリズムについて説明する．次に実際の演奏の音価系列から元の楽譜の音符列に復元する実験結果を示し，特に 3 連符を含む演奏の復元に関しても述べる．

2　従来の音価列推定問題の定式化

2.1　連続音声認識と音価列推定問題の同型性

本稿では，楽譜上の音符の (整数比関係にある) 正規の長さを「音価」と呼び，それが演奏されて音の物理的長さとして観測されたものを「音長」と呼ぶことにする．これは，音声認識における音素と特徴量の関係に類似している．演奏は，意図された音価系列が揺らぎを持つ音長系列に変換される過程であるとみなす (図 1).

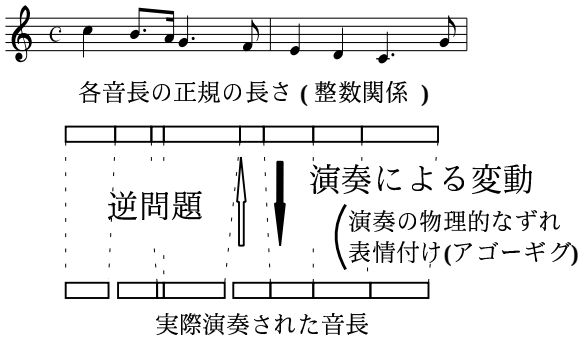

図 1: 逆問題としての音符列推定

表 1: 音声認識と音価系列認識の対応

	連続音声認識	音価系列認識
入力単位	文音声	楽曲
語彙	単語	リズムパターン
単位モデル	音素	音符
隠れ状態	音響イベント	
観測値	スペクトル列	物理的音長列

本問題はその逆問題として音長系列から音価系列を推定する問題と考えられ，連続音声認識とは表 1 のように同種の問題である．音声認識における音素を音符の音価に対応づけ，語彙や文法制約を音符列制約に対応づければ，問題を解くアルゴリズムも対応づけができる [7, 8].

2.2　逆問題としての音価列推定問題

本研究の目的は，演奏から得られた T 個の音長系列 $X = \{x_0, x_1, \ldots, x_T\}$ に対し，最もありそうな T 個の音価列 $Q = \{q_0, q_1, \ldots, q_T\}$ を対応付けることである．尤度最大の音価列 Q^* は連続音声認識の場合と同様に,

$$Q^* = \underset{Q}{\operatorname{argmax}} P(Q|X) \tag{1}$$

で与えられる．つまり，音長系列 X が与えられたとき，あらゆる仮説 Q の中で事後確率を最大化する音符列 Q^* を導出する仮説検定の問題となり，Bayes の定理より,

$$P(Q|X) = \frac{P(X|Q)P(Q)}{P(X)} \tag{2}$$

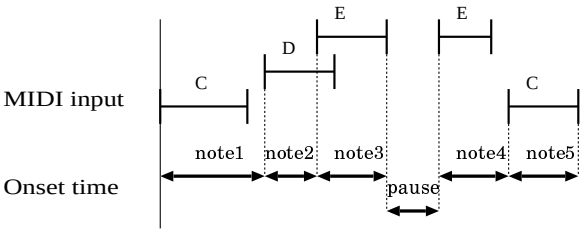

図 2: 「onset time 処理」による音長系列 X の導出

であり，ここで $P(X)$ は Q に依らないから，

$$Q^* = \operatorname*{argmax}_Q P(X|Q)P(Q) \qquad (3)$$

となる．

右辺の条件つき確率 $P(X|Q)$ は音長の伸縮変動モデルから得られ, 音価系列 Q が音価列パターンとして現れる事前確率 $P(Q)$ は, 音符列モデルから得られると考える [7, 8]．

2.3 音長の定義

まず，(演奏情報から得られる) 音長の定義を行う．図 2 に示すように，個々の音符が演奏される継続時間は，レガートやスタッカートなどのアーティキュレーションによって変動し，後続する音との間で，音が重なり合いあるいは空隙が生じることが多く，音符の音価に対応する音長の物理的観測量としては不適切である．従来どおり本稿では，便宜的に当該音の開始時刻から次の音の開始時刻までを音長として扱った．また，この音長が音の継続長よりある一定の値以上長い場合は，その間の無音区間を休符が存在するとみなした．以下の実験では MIDI キーボードから得られた標準 MIDI ファイルから，この処理 (以下「onset time 処理」[7, 8]) により，T 個の音長あるいは休符長の系列 $X = \{x_1, x_2, \ldots, x_T\}$ を抽出した．

2.4 音長の伸縮変動モデル

演奏者は，楽譜として意図した内部状態の系列 Q に相当する音価列から，その演奏者の音楽的な表現 (アゴーギグ)，演奏の癖，演奏のスキル不足などの原因で，同一の音価の音符でもその物理的音長が変動し，結果として揺らいだ音長時系列 X を出力する．単純化して考えるため，これらを確率変動と見なす．

音価 k が音長 x で演奏される確率密度を $b_k(x)$ と書く．そのパラメータは，演奏データから学習する

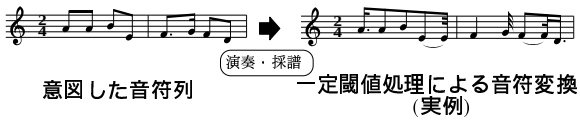

図 3: 常識に合わない楽譜

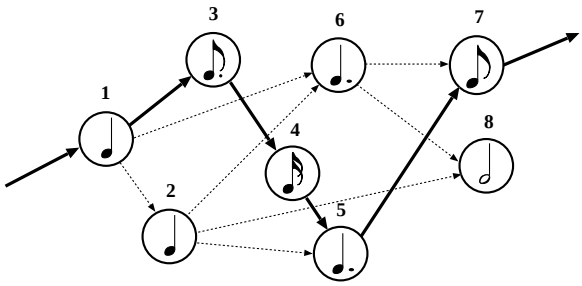

図 4: 音楽における状態遷移ネットワーク

ことができる．これは，人間の音楽経験に基づく音長の揺らぎの常識の形成に譬えられる．ネットワーク上の経路 Q は音価の列 (リズムパターン) に対応するので，Q が与えられた場合に音長系列 X が観測される確率密度を $P(X|Q)$ と書く．これは，音価列が Q であるような楽譜を演奏すると，各音符の長さの系列が X のようである確率 (密度) である．

音価 k が演奏される長さ x は，本来音価の履歴に依存すると考えるべきである (例えば 3 連符の場合，3 個の並びの中で 2 音目の分散が大きく，3 音目の平均音長が短くなるという実験結果が得られた) が，サンプルデータが十分でないことから，従来通り音価の種類のみの関数と考え，$b_k(x)$ を正規分布で近似した [7, 8]．

2.5 音価系列のモデル

音長に揺らぎがある演奏でも，聴き手には意図した音価の列 (さらに，時には伸縮の意図も) が伝わるのはなぜか．その理由の一つは，聴き手が出現しうる音符列に関する常識を持っているからであろう．たとえば図 3 右のような楽譜は理論上は可能ではあるが常識に合わない．そこで，聴き手や音楽家の常識をモデル化するために，本手法では音楽的な制約として音符の系列をモデル化する．これは音声認識における言語モデルあるいは文法に相当する部分である．

ここでは，言語モデルがしばしば状態遷移ネットワーク (有限状態オートマトン) で表現され，これを展開すると音素のネットワークとして理解できる

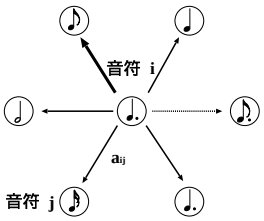

図 5: 2 音符連鎖モデル

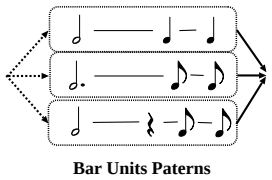

図 6: 小節パターンモデル

ことに倣い，音価の系列の生成源を確率的状態遷移ネットワークで表現する．図 4 に示すように，ネットワーク上のある状態遷移経路 $Q = \{q_0, q_1, \ldots, q_T\}$ は，あるリズムパターンを表現しており，その経路ごとにその生成確率 $P(Q)$ が存在している．しかし，それらのあらゆる音価列 Q について生起確率 $P(Q)$ を統計から得ることは実際上不可能であり，何らかの近似を行わざるを得ない．

ここで，q_t は時刻 t における内部状態であり，異なる音価の種類には，異なる状態が対応している．定義した音価の種類は，16 分音符の長さ及び 3 連 16 分音符を分解能とする音符が 20 種類，休符が 12 種類の計 $32(= P)$ 種類とした．

2.5.1 従来の音価系列のモデル

先行研究における音符 1 個を内部状態 1 個に対応させるモデル (図 5) では，$P(Q)$ を

$$P(Q) = \pi_{q_0} \prod_{t=1}^{T} a_{q_{t-1} q_t} \tag{4}$$

で与えてきた [7, 8] が，直前の音価のみに依存した音価列の構成法は，音楽のリズムパターンの制約としては拘束力が弱いと考えられる．(ここで，π_i は音価 i からフレーズが始まる初期確率，a_{ij} は音価 i から音価 j に遷移する確率である.)

一方，従来の小節を単位としたリズムパターンで曲を構成するモデル (図 6) では，童謡・民謡・歌曲といったジャンルの中で，テンポや拍子・小節線を含めた推定にも効果的であった [7, 8] が，一般のより複雑な音楽ジャンルに拡張する際には，実際に現れ得るリズムパターンの数が膨大であるため，全てのパターンの生起確率やそれらのパターン間の連鎖確率を与えることは現実的ではないと考えられる．

3 n-gram を用いた音価系列のモデル

3.1 n-gram 遷移確率を用いた音価系列のモデル

本稿では，状態の遷移確率を直前の音価だけでなく，$n-1$ 個前までの履歴も考慮する n-gram 遷移確率を用いるモデルを導入する．例えば状態 i，j の次に状態 k に遷移する trigram の遷移確率を a_{ijk} とすると，$P(Q)$ は

$$P(Q) = \pi_{q_0 q_1} \prod_{t=2}^{T} a_{q_{t-2} q_{t-1} q_t} \tag{5}$$

で与えられる．この多重 Markov モデルを用いることで，(4) 式を用いる場合よりも $P(Q)$ の精度が向上する．このモデルにより，従来の音符連鎖モデル (2-gram に対応) と同様にほぼ全ての音符列を構成でき，また n の値を大きくすることで小節モデルと同程度の音楽的制約を与えることができると考えられ，テンポ推定や拍子・小節線推定への拡張も可能であると考えられる．

3.1.1 HMM の遷移頻度の学習

n-gram 遷移確率の学習には，Web 上の携帯着信メロディ用の単音のクラシックフレーズのデータから得られる，全 130 曲，50000 音程度のサンプルデータを用いた．またジャズのサンプルデータも同様に得，n-gram 遷移確率のデータに大きな差があることを確認した．

3.1.2 サンプルデータからの学習及びスムージング

音声認識の場合は，パラメータ学習手法として，Baum-Welch のアルゴリズム [5, 6] 等が知られているが，音価 HMM の場合は HMM の状態滞留確率が 0 であるために，サンプル中の生起頻度の割合を得る単純な操作により，学習が可能となる．

ただし，遷移確率等を導出する際に，学習データが少ない場合の確率推定誤差を軽減するため，連続音声認識の場合と同様に，遷移・初期確率にスムージング [5, 6] を施した．たとえば音価の種類が i, j, k と遷移する確率 a_{ijk} の場合，以下のように，ijk と遷移する割合に，jk と遷移する割合，k の生起する割合，および定数を重み付けをして加えた．

$$a_{ijk} = \delta_0 + \delta_1 \frac{N_k}{N} + \delta_2 \frac{N_{jk}}{N_j} + (1-\delta_0-\delta_1-\delta_2)\frac{N_{ijk}}{N_{ij}} \quad (6)$$

ここで，$\delta_0, \delta_1, \delta_2 \ll 1$ は重み，N_x は x が生じた頻度，N は全学習サンプル数である．

4 n-gram モデルを実現するアルゴリズム

4.1 n-gram 遷移確率を用いた Viterbi アルゴリズム

前章までの議論により，例えば trigram 遷移確率を用いる場合には，

$$Q^* = \underset{\{q_0,q_1,\ldots,q_T\}}{\operatorname{argmax}} \pi_{q_0 q_1} \prod_{t=2}^{T} a_{q_{t-2} q_{t-1} q_t} \cdot b_{q_t}(x_t) \quad (7)$$

の右辺をを最大化すればよい．

ここでは式 (7) の右辺の尤度を最大にする状態パスを導出する問題が，従来の bigram を用いた単純 Markov モデルの Viterbi アルゴリズム [7, 8] に帰着できることを示す．

Viterbi 計算においては，各時刻 t における状態 q_t が音価 j になる経路の中で尤度最大になる経路の尤度 $\alpha_t(j)$ を逐次計算することによって，最終的に尤度最大となる経路を得るのであるが，今回導入する n-gram を用いた Viterbi 計算の中で例えば trigram の場合は，各時刻 t において 1 つ前の時刻の音価が j，現在の音価が k となる経路の中で，尤度最大となる経路の尤度 $\alpha_t(j,k)$ を逐次計算する．

すなわち，元の単純 Markov モデルの HMM の音価状態の集合を S としたとき，2 つの音の並びの直積集合 $S \times S$ を 2 重 Markov モデルの内部状態とした HMM を考えることに相当する．この時 $S \times S$ から 1 音ずらした $S \times S$ への遷移において，遷移前の $S \times S$ の状態の 2 番目の音価と，遷移後の $S \times S$ の状態の 1 番目の音価とが一致しているので，状態遷移は制限される．このため HMM の状態空間自体は大きくなるものの，ある状態からの遷移先の状態空間の大きさは，単純 Markov モデルの HMM と同じである．

同様の計算はより大きな n の n-gram Viterbi 系列を導出する際も適用でき，この Viterbi 計算の計算量は，保持する履歴を 1 つ増やすごとに高々定義した状態数 P 倍になる．このため，P の値が比較的小さい (= 32) 今回のケースでは有効である．

$t = 1$ について (初期化):

$(j,k) = (1,1),(1,2),\ldots,(P,P)$ について:

$$\alpha_1(j,k) = \log\{\pi_{jk} b_j(x_0) b_k(x_1)\}$$

$t = 2,3,\ldots,T$ について:

$(j,k) = (1,1),(1,2),\ldots,(P,P)$ について:

$$\alpha_t(j,k) = \max_i \alpha_{t-1}(i,j) + \log\{a_{ijk} b_k(x_t)\}$$

$$\beta_t(j,k) = \operatorname{argmax}_i \alpha_{t-1}(i,j) + \log\{a_{ijk} b_k(x_t)\}$$

終了:

$$[q_{T-1}, q_T] = \underset{[j,k]}{\operatorname{argmax}} \alpha_T(j,k)$$

トレースバック: $t = T-2, T-3, \ldots, 0$ について:

$$q_t = \beta_{t+2}(q_{t+1}, q_{t+2})$$

図 7: 3-gram Viterbi 計算のアルゴリズム

trigram の音符列モデルを用いる場合は，

$$\alpha_t(j,k) = \max_{\{q_0 \cdots q_t\}} \Pr\{q_0 \ldots q_{t-1} = j, q_t = k, x_0 \ldots x_t | M\} \quad (8)$$

となる．(M は音長の伸縮変動モデルと本稿で導入した n-gram モデルの統合モデル) $\alpha_t(j,k)$ の定義から，

$$\alpha_t(j,k) = \max_{i,j} \alpha_{t-1}(i,j) a_{ijk} b_k(x_t) \quad (9)$$

とできる．

この値を逐次計算することで，所望の状態系列が得られる．trigram の場合のアルゴリズムを図 7 に示す．ただし図 7 では対数尤度を用いている．

4.2 尤度 N 位までの音価列を得るアルゴリズム

次に，従来の bigram 遷移確率を用いて尤度 N 位までの状態系列を求めるアルゴリズムを与える．連続音声認識の場合には A*サーチ [10] を用いた手法等が知られているが，今回の音価 HMM のように，入力された音長系列と出力する音価状態列が 1 対 1 に対応する，すなわち状態滞留確率が常に 0 となる場合には，尤度最大のみを求める場合の計算時間に比べ高々N 倍に収まる高速なアルゴリズムが得られる．本稿の実験では，この N-best のアルゴリズム

$t=0$ について (初期化):

$i=1,2,\ldots,P$ について:

$n=1$ について: $\alpha_0(1,i) = \log\{\pi_i b_i(x_0)\}$

$n=2,\ldots,N$ について: $\alpha_0(n,i) = -\infty$

$t=1,2,\ldots,T$ について:

$j=1,2,\ldots,P$ について:

$n=1,2,\ldots,N$ について:

$$\alpha_t(n,j) = \max_{m,i}(n\text{-th})\{\alpha_{t-1}(m,i) + \log\{a_{ij}b_j(x_t)\}$$

$$[\gamma_t(n,j), \beta_t(n,j)] = \operatorname*{argmax}_{[m,i]}(n\text{-th})\{\alpha_{t-1}(m,i) + \log\{a_{ij}b_j(x_t)\}$$

終了: $n=1,2,\ldots,N$ について:

$$[\mathrm{tmp}_T(n), q_T(n)] = \operatorname*{argmax}_{[m,i]}(n\text{-th})\ \alpha_T(m,i)$$

トレースバック: $n=1,2,\ldots,N$ について:

$t=T-1,T-2,\ldots,0$ について:

$$\mathrm{tmp}_t(n) = \gamma_{t+1}(\mathrm{tmp}_{t+1}(n), q_{t+1}(n))$$

$$q_t(n) = \beta_{t+1}(\mathrm{tmp}_{t+1}(n), q_{t+1}(n))$$

図 8: N-best Viterbi 計算のアルゴリズム

を用いて，まず尤度 N 位までの解の候補を求めた後に，n-gram 遷移確率を用いて再ソートする手法による計算時間の節減の可能性も考察した．

以下アルゴリズムの説明を行う．ここでは，時刻 t に状態 j に遷移する経路の中で，尤度が n 番目に大きい経路の尤度を $\alpha_t(n,j)$ とする．$\alpha_t(n,j)$ の定義から，

$$\alpha_t(n,j) = \max_{m,i}(n\text{-th})\{\alpha_{t-1}(m,i)a_{ij}b_j(x_t)\} \quad (10)$$

となり，これを時刻および順位ごとに逐次計算すると，最終的に尤度が上位 N 位までの状態系列が得られる．ここで，$\max(n\text{-th})$ は n 番目に大きいものを表すとする．n 位の内部状態の系列 $\{q_0(n), q_1(n), \ldots, q_T(n)\}$ を得るアルゴリズムを図 8 に示す．

5 n-gram を用いた HMM による音符列推定実験

5.1 本実験の実験条件

MIDI データの入力には MIDI キーボード (YAMAHA CBX-K2) を用い，MIDI 音源 (YAMAHA MU-2000 TONEGENERATOR) を通して PC に入力した．また演奏収録ソフトとしては，YAMAHA XG-works を用いた．

表 2: 入力に用いたクラシック曲

曲目	作曲者	テンポ	3 連符
Ave verum corpus	Mozart	90	無
別れの曲	Chopin	90	無
諸人こぞりて	Hendel	90	無
アルルの女	Bizet	120	無
ボレロ	Ravel	90	無
アラヴァマ序曲	Barns	150	有
交響曲 2 番 3 楽章	Brahms	110	有
くるみ割り人形	Tchaikovsky	120	有

演奏データとしては，指定された曲を誤りなく演奏するのに何度も弾き直さなければならない者から，初見でほぼ誤りなく弾ける者まで幅広い演奏スキルをもつ，19 人の被験者のデータを用いた．

実験対象曲は，比較的鍵盤入力しやすい単音のクラシック曲の旋律を用いた．実験に用いた 8 曲を表 2 に挙げる．なお対象曲は全て学習外データである．また実験には，音符数の誤りがない演奏を用い，以下の 2 種類の演奏条件の下での入力を用いる．

演奏条件 1 演奏中にメトロノームを用い，テンポにできるだけ忠実な演奏

演奏条件 2 メトロノームを演奏前に聞き，演奏中はメトロノームを用いずにできるだけ一定のテンポを保つことを心がける演奏

また，認識率としては，各音長が正しい音符の長さに変換されているかのみを評価し，以下の式により与える．

$$\mathrm{accuracy} = \frac{T - \mathrm{sub} - \mathrm{del} - \mathrm{ins}}{T} \times 100\ [\%] \quad (11)$$

ここで，T は正解状態列の総音符・休符数，sub, del, ins はそれぞれ置換，脱落，挿入誤りが生じた数である [7, 8]．

5.2 実験結果

実験では，以下の 6 通りの認識率を比較した．すなわち，onset time 処理後に閾値で処理する場合 (THRD)，従来の bigram による HMM を用いる場合 (2-HMM), N-best Viterbi 計算によりまず尤度 N 位までの候補を得た後に，trigram・quadgram 遷移確率によって再ソートを行う場合 (3-SORT・4-SORT)，

表 3: onset time 後の閾値処理 (THRD) による曲ごとの認識率と各手法 (2-HMM,4-SORT,4-HMM) の THRD に対する誤り削減率．単位:[%]，ただし上段がメトロノーム有り，下段がメトロノームなし

曲目	閾値	2HM	4SRT	4HM
Ave verum	98.4	0	0	**50**
corpus	97.7	0	0	0
別れの曲	97.9	-24	-24	-24
	94.5	20	20	20
諸人こぞりて	95.4	15	15	15
	92.0	23	23	**49**
アルルの女	86.3	0	0	**11**
	88.5	3	3	3
ボレロ	94.6	22	22	22
	92.4	25	25	25
アラヴァマ	81.0	**52**	**52**	33
序曲	77.0	**68**	65	35
Brahms	65.3	49	49	**61**
交響曲 2 番	74.2	81	81	**86**
くるみ割り	60.0	92	92	92
人形	46.0	91	91	91

trigram・quadgram による HMM を用いる場合 (3-HMM・4-HMM) の 6 通りである．結果の一部を表 3 に示す．また図 9〜図 11 に，n-gram の HMM による楽譜化の例を挙げる．

市販ソフトの場合は，単なる閾値処理を行っていると考えられ，例えば図 10 のような不適格な楽譜を出力するが，同じ演奏に対し，HMM を用いた場合は図 11 のようなほぼ正しい楽譜が得られた．なお表 3 における閾値処理 (THRD) の結果は，onset time 処理の後に閾値処理を行った場合の認識率である．

今回導入した quadgram を利用する場合，従来の単純 Markov モデルを用いる場合に比べ 1 曲を除いて認識率が上昇し，単純 Markov モデルの場合に対する誤り削減率は，0〜50 %となった．アラヴァマ序曲の認識率の低下は，3 連符を含んでいたためと考えられるが，この 3 連符の識別に関しては次章で考察する．

また，N-best Viterbi 計算を用いた後に n-gram 遷移確率により再ソートを行う場合は，現在のところ従来の単純 Markov モデルと同程度の認識率しか得られていないが，認識率が上昇した場合もみられた．今後の実験によりさらなる検証が必要である．

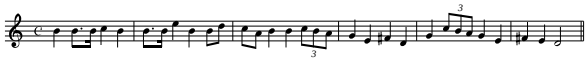

図 9: 「Brahms 」の演奏用の正しい楽譜

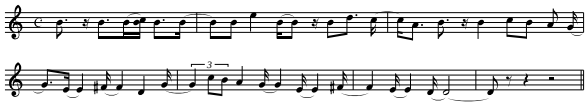

図 10: 「Brahms 」の演奏の XGworks による楽譜化

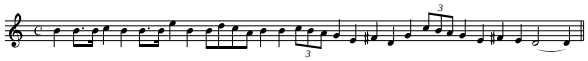

図 11: 「Brahms 」の演奏の HMM による楽譜化

6　3 連符リズムの識別

6.1　3 連符リズムの識別の難しさ

3 連符のリズムは，16 分音符を分解能とする音価の系列にない音価 3 個からなる特殊なリズムである．このため，市販ソフトのような閾値ベースの処理によって，3 連符リズムを正しく識別することは，困難であると考えられる．また，実際の演奏データの 3 連符の音長の揺らぎを調べると，分散が比較的大きくなっており，演奏の難しいリズムであるといえる．さらに，3 連符は通常 3 個連続して表れるため，n-gram 遷移確率に特異な傾向が現れる．

本章では，まず 3 連符に対する，HMM を用いたトップダウン的なリズム推定の有効性について述べ，さらに今回提案した n-gram モデルによる効果を考察する．

6.2　3 連符リズムの認識実験

本章では 3 連符の認識精度を調べるために，正しく 3 連符と認識できた個数 (9− 削除誤り) と誤って 3 連符と判断した個数 (挿入誤り) の両方の平均数を，前章と同じ 6 通りの手法で比較する．

6.3　考察

前章で挙げた 3 連符を含む曲に関しては，表 3 のように，HMM を用いた手法により認識率が大きく向上した．これは，3 連符の音価を閾値処理により認識することが原理的に困難なためと考えられる．

また，サンプル曲の中で「アラヴァマ序曲」以外

表 4:　3 連符の認識個数　単位:[個]

	2HMM	3SORT	3HMM	4SORT	4HMM
認識数	**7.9**	7.2	7.0	**7.9**	6.7
誤認識数	**0.4**	0.5	1.3	**0.4**	2.8

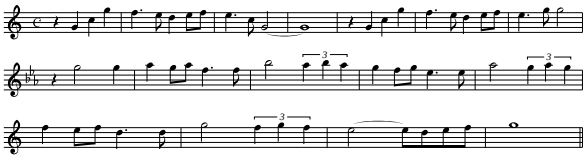

図 12: 「アラヴァマ序曲」の正しい楽譜

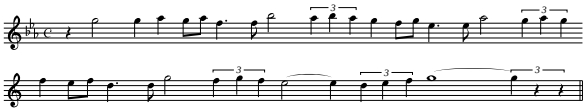

図 13: 「アラヴァマ序曲 (一部)」の 3 連符の誤認識例

の曲に関しては，HMM を用いることでほぼ正しく 3 連符を認識することができ，更に n-gram の導入により若干認識率が上昇した．

ここで,「アラヴァマ序曲」(図 12) についての 3 連符の認識精度の結果を表 4 に挙げる．

表 4 より，quadgram を用いて再ソートする場合に限り，単純 Markov の場合と同程度の認識率を挙げるものの，trigram や quadgram の HMM を用いる場合は誤挿入が多くなり，3 連符の認識精度が単純 Markov の場合に比べ低下することが分かった．この場合，16 分音符 3 個を 3 連符リズムとする誤認識や，正しい 3 連符 3 個の次の音も 3 連符とする誤認識 (図 13) が多く見られた．

3 連符の認識には，3 連符独特の演奏揺らぎを音長モデルに組み込む (3 連符の何番目の音かで，異なる平均や分散を与える等) 処理や，n-gram の n をさらに増やすことが有効であると考えられる．例えば 5-gram を用いる場合，3 連符音価 3 個の並びの後に 3 連符以外の音価に遷移しやすいことを表現できるため，精度の向上が期待できる．ただし，5-gram を導入するためにはさらに多くのサンプルデータが必要である．

7　結論と今後の課題

本稿では，まず n-gram 遷移確率を用いる音価列の学習・認識を提案し, n-gram を用いた Viterbi 計算のための 2 種類のアルゴリズムを提案した．これらを用いた結果，3 連符リズムを含めて意図した音符列の認識率に一定の向上が見られた．同様に n-gram モデルを，先行研究で効果的であったテンポ変動 HMM モデル [7, 8] に適用し，テンポへの追従や小節線推定の精度が向上することが期待される．また，長さ 0 の音符を考えることにより，単音の旋律のみでなく和音を伴った入力が扱える可能性もある．

今後は学習用の音楽統計データを十分に整備すると共に，n-gram の手法を，我々の目指すジャンルやスタイルを考慮したリズムパターンのモデル学習方法，リズムパターンに依存した音長伸縮特性を考慮した推定，ユーザのスキルや癖を学習するユーザ適応技術などに適用したい．

参考文献

[1] H. C. Ronguet-Higgins: Mental Processes, The Mit Press, 1987.

[2] 片寄, 井口: “知的採譜システム,” 人工知能学会誌, Vol.5, No.1, pp.59-66, 1990.

[3] 海野, 中西: “音楽情景分析における楽音認識と自動採譜,” インタラクション 99 予稿集, 1999.

[4] P. Desain, H. Honing: “Quantization of Musical Time; A Connectionist Approach,” Computer Music Journal, Vol. 13, pp. 56–66, 1989.

[5] 中川聖一: 確率モデルによる音声認識, 電子情報通信学会, 1988.

[6] L. Rabiner, B.-H. Juang: Fundamentals of Speech Recognition, Prentice-Hall, 1993.

[7] 齋藤 直樹, 中井 満, 下平 博, 嵯峨山 茂樹, “隠れマルコフモデルによる音楽演奏情報からの音符列推定,” 平成 11 年度電気関係学会北陸支部連合大会講演論文集, F-62, p.362, Oct 1999.

[8] 齋藤 直樹, 中井 満, 下平 博, 嵯峨山 茂樹, “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,” 平成 11 年情報処理学会音楽情報科学研究会資料, 99-MUS-33, pp. 27–32, Dec. 1999.

[9] 野池, 乾, 野瀬, 小谷: “演奏情報と楽譜情報の対からの演奏表情規則の獲得とその応用,” 情報処理学会音楽情報科学研究会, 97-MUS-26-16, pp.109-114, 1998.

[10] Frank K. Soong and Eng-Fong Huang: A Tree-Trellis Based Fast Search for Finding the N Best Sentence Hypotheses in Continuous Speech Recognition, IEEE 2977-7, 1991.