

# 出力合流隠れマルコフモデルに基づく 多声部音楽のリズム採譜

中村栄太<sup>1,a)</sup> 小野順貴<sup>1</sup> 嵯峨山茂樹<sup>2</sup>

概要：多声部のパッセージのリズム採譜を効果的に行う確率モデルに基づくアルゴリズムを提案する。リズム採譜は、演奏から拍子やリズムの構造など音楽的な情報を取り出すために必要である。これは、楽譜の構造とテンポ変動などを含む演奏の時間的構造を同時に扱う必要があるため容易ではないが、隠れマルコフモデル (HMM) を用いて効率的に行える。ポリリズムを持つ様な多声部のパッセージでは、これら両方の構造が複雑であり、リズム採譜はより高度な技術を要する。これらの構造が各声部内では比較的単純なことに着目し、各声部に対応した複数の HMM からの出力音符が合流して、多声部音楽を記述する確率モデルを提案する。声部を考慮しない従来の HMM を用いた結果比べ、採譜結果は良好であり、精度および計算効率の向上の可能性についても議論する。

## 1. はじめに

リズム採譜（あるいはリズム量子化）は、演奏音符から通常楽譜では量子化されている音価を推定する問題である。この自動化は機械による聴取の基本形であり、音楽情報検索や自動編曲、音楽データベースの作成などに応用でき、自動採譜問題の部分問題でもある [1]。リズム採譜と自動採譜は、音声音響信号からそのテキストを推定する音声認識と類似しており、テキストと楽譜、音声と音楽演奏がそれぞれ対応する。この類似に基づき、音声認識で発達していた確率モデル [19] を用いる方法によるリズム採譜のアルゴリズムが提案されており [5], [6], [7]、現在も研究されている [3], [4]。

多声部音楽では、楽譜にポリリズムの様な複雑な構造が現れ得る。この様な楽譜の演奏の時間構造は、声部間非同期性により複雑になり得る [12]。楽譜と演奏のマッチング問題における経験によると、この様な楽譜の処理にはモデルに声部構造を取り込むことが重要である [13], [14], [15], [16]。演奏される楽譜の声部情報は演奏音符自体においては明示的でないため、リズム採譜アルゴリズム内部において各々の音符の声部情報を推定する必要があり、また応用上使用可能な程度に計算量的に効率的でなければならない。

本稿では、声部構造を持つ確率モデルに基づくリズム採

譜のアルゴリズムを提案する。各々の声部は、文献 [6] と同様に隠れマルコフモデル (hidden Markov model, HMM) により記述され、全体のモデルは、これら複数の HMM からの出力を合流することで構築できる。ここで基礎となる、出力合流 HMM は、他の応用に関して文献 [8], [17], [18] で論じられているが、その基本的な結果もレビューする。

## 2. HMM によるリズム採譜

### 2.1 モデルの説明

本節では、まずリズム採譜のための HMM を用いた確率モデルについてレビューをし [6], [7]、モデルパラメータの学習についても次節以降で議論する。

演奏の過程は、意図された楽譜に基づき演奏音符を生成する確率過程として記述できる。楽譜は、各音符の音高  $p_n$  と楽譜時間  $\tau_n$  によって表され、音符数を  $N$  とすると  $(\tau, p) = (\tau_n, p_n)_{n=1}^N$  と記される。リズム採譜においては時間情報、即ち各音符の発音時刻  $t_n$  のみが本質的に重要であるが、これは背後にあるテンポ  $v_n$  によって制御されているとする。  $t = (t_n)_{n=1}^N (\tau, \tau p, v \text{ も同様})$  と記すことにすると、演奏モデルは、確率  $P(t, v | \tau, p)$  により表される。意図された楽譜を推定するためには、楽譜に現れる音符のモデルが必要であり、ここではこれを楽節モデルと呼ぶ。これにより確率  $P(\tau, p)$  が与えられる。以上により、確率  $P(t, v, \tau, p)$  が計算でき、さらに Bayes の公式より  $P(v, \tau | t, p)$  が得られるため、この確率の最大化によりリズム採譜を行える。

最も簡単な楽節モデルでは、楽譜時刻の列は Markov モ

<sup>1</sup> 国立情報学研究所  
101-8430 東京都千代田区一ツ橋

<sup>2</sup> 明治大学  
164-8525 東京都中野区中野

a) enakamura@nii.ac.jp

デルにより記述でき、その確率は音高列とは独立と考えることができる。楽譜時刻に関する並進不変性を仮定すると、この Markov モデルは  $\nu_n = \tau_{n+1} - \tau_n$  を発音間音価としたとき、確率  $P(\nu_n | \nu_{n-1})$  により記述される。このモデルでは、音高は直接観測されるものと扱われるため、その確率分布はリズム採譜の結果に影響を与えない。

意図された楽譜が与えられた時、演奏音符は演奏モデルに基づいて生成されると考えることができる。これには演奏楽譜マッチングに使われているものと同様のモデルを用いられる [16]。まず、テンポは Markov 過程によって生成されるとする： $v_{n+1} = v_n + \epsilon_v$  ( $\epsilon_v$  は Gaussian ノイズとする)。ここでは、テンポは通常の beat per minute (BPM) などとは逆数である、比  $\delta t / \delta \tau$  に比例するものとして定義する。時間並進に対して不変になるように、発音間時間  $\delta t_n = t_{n+1} - t_n$  を用いて発音時刻を記述すると、 $\epsilon_t$  を Gaussian ノイズとして、 $\delta t_n = \nu_n v_n + \epsilon_t$  とモデル化できる。ここで、 $n$  は演奏音符の順序と対応するものとする、 $\delta t_n$  は非負である。

ここまで、リズム採譜のための最も簡単なモデルの一つを議論してきた。本稿では、簡単のため以下主にこの設定に絞って議論するが、多くの他のモデルが提案されている [3], [4], [6], [7]。例えば、楽節モデルとして高次の Markov モデルを考えることができる他、例えば小節当たりのリズムパターンを単位として用い、階層的なモデルを構築することもできる。また、多声音楽を扱う方法にも複数の可能性がある。単純には、音価の状態空間に  $\nu = 0$  に対応する状態を導入できるが、これでは隣り合う二つの和音間の音価の関係は 1 次の Markov モデルでは効果的に取り入れられない。あるいは、自己遷移を用いて和音内の音符を記述できる。この際には、自己遷移は和音内音符に対応する遷移と同じ音価の繰り返しに対応する遷移の両方を表すことになる。以下、このモデルを考える。

まとめると、今のモデルはペア  $(\nu_n, v_n)$  を状態空間として、発音間時刻を出力するような HMM である。遷移確率と出力確率の詳細は次節以降に詳しく述べる。

## 2.2 楽節モデルのパラメータ決定

楽節モデルの状態空間は、可能な音価の集合である。一般に、音価の選択肢の数が増えるにつれて、生成される楽譜断片の可能性は大きくなるが、正確な推定はより難しくなる。かなり一般の設定として、ここでは、全音符、2分音符、4分音符、8分音符、16分音符、32分音符の各基本形と3連符、付点音符、二重付点音符からなる、全部で24種類の音価を考えることにする。

音価間の遷移確率は、原理的には音楽データから得られる。パラメータ決定の効率を上げ、データのスパースネス問題を避けるため、以下音価のスケール不変な確率を考える： $P(\nu_{n+1} | \nu_n) \propto \text{Freq}(\nu_{n+1} / \nu_n)$ 。ここで、 $\text{Freq}$  は連続する2つの発音間音価の比の相対頻度を表す。主に古典音楽

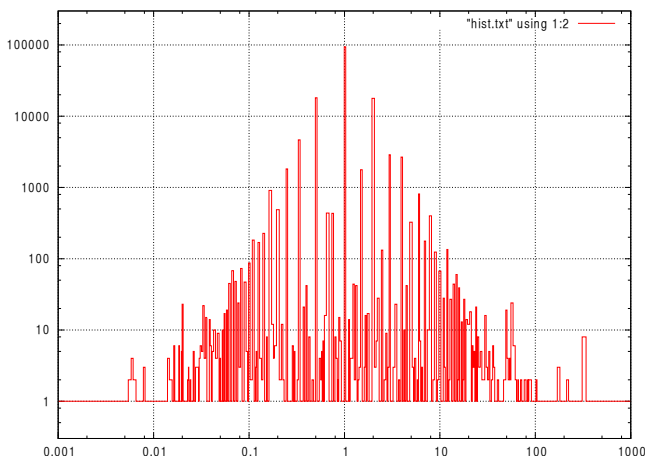


図 1 Relative frequency of the ratio of two succeeding IONVs.



図 2 Examples of homophonized representation of polyrhythmic passages.

のレパートリーから、任意に選ばれた音楽データから得られた確率を図 1 に示す。発音音価の全体の数は約 150,000 であった。

## 2.3 演奏モデルのパラメータ決定

演奏モデルのパラメータは、Gaussian ノイズ  $\epsilon_v$  と  $\epsilon_t$  の中に含まれている。ゼロ平均の仮定より、残るパラメータは各々の分散となる。これらのパラメータの値は演奏データから決定できるが、それには文献 [16] の結果を用いられる。演奏解析の結果、この文献では標準偏差  $\sigma_t = \sqrt{\langle \epsilon_t^2 \rangle}$  の値は、約 0.02 s と得られている。標準偏差  $\sigma_v = \sqrt{\langle \epsilon_v^2 \rangle}$  の値も演奏データベース [20] の解析より得られており、本稿での記法では、 $\sigma_v = 0.06 \text{ s/QN}$  である (QN は 4 分音符単位を表す)。

## 3. 出力合流 HMM によるリズム採譜

### 3.1 複雑な多声部楽節における問題

2 節のモデルでは、多声部楽節は Conklin のホモフォニー

化表現 [21] と同様の方法で扱われている。実際、発音イベントのみを考慮した場合には、一般の多声部楽節は等価なホモフォニー楽節によって表現できる。しかし、多声部音楽においては、複数声部が同時に重ねられた結果、通常の単旋律やホモフォニーに比べて複雑な構造をしばしば持っている。典型的な例は、しばしば 2 対 3 や 3 対 4 で現れるポリリズム楽節である。この場合、各声部は一般的なリズムを持っているが、ホモフォニー化表現におけるリズムは稀であることがある。このことは、効果的な楽節モデルでは、声部上を取り込む必要性があることを示唆している。

上のことに加えて、多声部楽節の演奏は、声部間非同期性により複雑な時間構造を持つ [12], [13], [14]。極端な場合には、声部間で楽譜時刻の異なる音符の順序入れ替えが起こることもある。楽譜・演奏マッチングの文脈で議論されているように、こうした効果は多声部楽節のホモフォニー化表現に基づくモデルによっては取り入れることは難しく、演奏モデルにおいても声部構造を組み込む必要があることを示唆している。

最近、多声部音楽演奏のこうした特徴を記述するための効率的なモデルが提案されている [8], [17]。次節でこのモデルの基礎的な結果を復習し、このモデルのリズム採譜への応用について 3.3 節で述べる。

### 3.2 出力合流 HMM の説明

以下に述べるモデルは、まずそれぞれの声部に対して HMM (パート HMM と呼ぶ) を考え、それら複数の HMM を結合させて、全ての出力を合流させたモデルを構築するという考え方に基づく。重要なのは、各時刻において観測される記号はどれか一つのパート HMM から出力され、この時その他のパート HMM は遷移および出力をしないということを仮定する点である。全体のモデルは、単純には HMM の積モデルとなるが、実際にはこの条件により効率的な推論アルゴリズムを持つことを示せる。後に議論するように、計算量の効率性を保ったまま、パート HMM 間にいくつかの相互依存性を導入することもできる。

以下、一般の HMM の出力合流 HMM について述べる。簡単のため、主に 2 つのパート HMM を合流させる場合を考える。\$a\_{ii'}^{(1)}\$ と \$a\_{jj'}^{(2)}\$ をパート HMM の遷移確率、\$b\_{ii'}^{(1)}(o)\$ と \$b\_{jj'}^{(2)}(o)\$ をその出力確率とする。ただし、\$i\$ や \$j\$ は HMM の状態、\$o\$ は出力記号を表す。ここでは、出力確率が現在と一つ前の状態の両方に依存する一般の場合を考え、また 2 つの HMM の状態空間は異なってもよい。

全体のモデルの状態は、ペア \$(i, j)\$ により表される。各時刻にどちらのパート HMM が遷移するかを示す変数 \$\eta = 1, 2\$ を導入し、出力合流モデルの状態空間は \$k = (\eta, i, j)\$ で表現される。パート HMM 間に相互依存性が無い時、これらは各時刻でどちらが遷移と出力を行うかを選択する確率過程によってのみ結びつけられる。この確率過程をコイン投

げの過程 (Bernoulli 過程) として、その確率値を \$\alpha\_1\$ と \$\alpha\_2\$ (\$\alpha\_1 + \alpha\_2 = 1\$) とすると、出力合流モデルの遷移確率は次の様に表される：

$$a_{kk'} = P(k'|k) = \begin{cases} \alpha_1 a_{ii'}^{(1)} \delta_{jj'}, & \eta' = 1; \\ \alpha_2 a_{jj'}^{(2)} \delta_{ii'}, & \eta' = 2. \end{cases} \quad (1)$$

出力確率は、遷移する方に選ばれたパート HMM の出力確率に従うとすると、次の様に書ける：

$$b_{kk'}(o) = P(o|k, k') = \begin{cases} b_{ii'}^{(1)}(o) \delta_{jj'}, & \eta' = 1; \\ b_{jj'}^{(2)}(o) \delta_{ii'}, & \eta' = 2. \end{cases} \quad (2)$$

式 (1) と (2) は、出力合流モデルがそれ自体 HMM であることを示している。各時刻で遷移しない方のパート HMM の現在の状態が、\$k'\$ の中に保存されており、それにより出力合流 HMM では声部構造が保たれることに注目されたい。

パート HMM 間には、次の様にいくつかの相互依存性を導入することができる：

$$a_{kk'} = \begin{cases} \alpha_1(k) a_{ii'}^{(1)} \delta_{jj'} \phi_{kk'}^{(1)}, & \eta' = 1; \\ \alpha_2(k) a_{jj'}^{(2)} \delta_{ii'} \phi_{kk'}^{(2)}, & \eta' = 2, \end{cases} \quad (3)$$

$$b_{kk'}(o) = \begin{cases} b_{ii'}^{(1)}(o) \delta_{jj'} \psi_{kk'}^{(1)}(o), & \eta' = 1; \\ b_{jj'}^{(2)}(o) \delta_{ii'} \psi_{kk'}^{(2)}(o), & \eta' = 2. \end{cases} \quad (4)$$

ここで、\$\alpha\_1(k) + \alpha\_2(k) = 1\$ であり、\$a\_{kk'}\$ と \$b\_{kk'}(o)\$ は適切な規格化条件を満たすものとする。相互依存項 \$\alpha\_\eta(k)\$ と \$\phi\_{kk'}^{(\eta)}, \psi\_{kk'}^{(\eta)}(o)\$ の応用上の例は、文献 [17] で議論されている。出力合流 HMM は 3 つ以上の声部に対しても一般化可能であり、\$\eta\$ と \$i\_\eta\$ の両方に関して、高次の Markov モデルを考えることもできる [8]。

類似の HMM が文献 [22] で提案されているが、重要な違いはここでのモデルでは各時刻においてただ 1 つの要素 HMM のみが遷移・出力を行うところである。この結果、導入可能な相互依存項の形も異なる。上で議論したように、この性質は多声部音楽に適切に応用されるために特に重要である。

### 3.3 リズム採譜のための出力合流 HMM

出力合流 HMM を用いたリズム採譜は、2 節で述べたモデルをパート HMM と用いることにより基本的に行える。3 つ以上のパート HMM を合流させることも出来るが、実用的に最も重要である 2 パートの場合について以下述べる。モデル構築に関して、次の 2 つに注意する。まず、2 つの声部のテンポには強い依存関係が期待されるが、これは両パート HMM に対して同一のテンポを仮定することにより効率的に組み込まれる。次に、声部構造に関して音高情報が重要なので、ここでは音高も明示的に

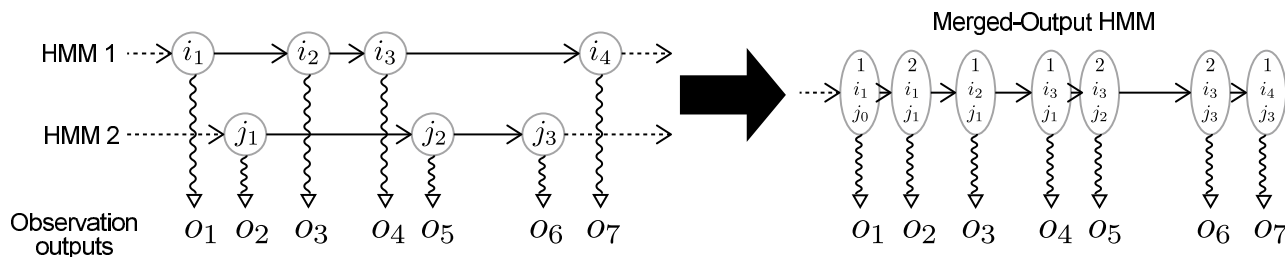


図 3 Schematic illustration of merged-output HMM.

モデル化する必要がある．以上 2 点に注意して，2 節の結果を用いると，リズム採譜のための 2 声部のモデルは次の様に記述される．各パート HMM の状態空間は，3 つ組  $(\nu^{(\eta)}, p^{(\eta)}, v)$  ( $\eta = 1, 2$ ) で表現され，出力合流モデルの状態空間は  $k = (\eta, \nu^{(1)}, p^{(1)}, \nu^{(2)}, p^{(2)}, v)$  により表現される．遷移確率  $P(k_{n+1}|k_n)$  は， $\eta_{n+1} = 1$  の時，

$$\alpha_1 P(\nu_{n+1}^{(1)}|\nu_n^{(1)})P(p_{n+1}^{(1)}|p_n^{(1)})P(v_{n+1}|v_n)\delta_{\nu_{n+1}^{(2)}, \nu_n^{(2)}}\delta_{p_{n+1}^{(2)}, p_n^{(2)}} \quad (5)$$

で表され， $\eta_{n+1} = 2$  の時も同様である．発音間時間の出力確率も  $\eta_{n+1}$  に応じて同様に与えられる．

推論の過程で声部分離を行うためには，音高に関する確率  $P(p_{n+1}^{(\eta)}|p_n^{(\eta)})$  を適切に与える必要がある．これは，文献 [8], [18] で行ったピアノ曲の分析結果を用いて行える．また，そこで議論されている様に，この音高に関する確率によって高い精度で，左右両手に対応する声部分離を行うこともできる．

#### 4. リズム採譜アルゴリズム

3.3 節で説明したモデルを基にリズム採譜アルゴリズムを構築できる．このモデルの状態空間において，テンポに対応する次元は基本的には正の実数で張られる連続空間であるが，通常の音楽で用いられるテンポの値はある区間に限られており，これを十分に細かな離散値の集合で置き換えることにより，Viterbi アルゴリズムを用いて最尤推定が可能である．5 節の性能評価の際には，0.2 – 2.0 (sec/四分音符) の区間内で，対数上で等間隔に 50 個のテンポの値を取っている．なお，文献 [8] に議論されているように，出力合流 HMM では Viterbi アルゴリズムに工夫をすることにより，大幅に計算量を削減できる．音価の種類を  $N_{nv}$ ，音高の数を  $N_p$ ，上記テンポの離散値の数を  $N_v$  とすると，モデルの状態数は， $N = 2N_{nv}^2 N_p^2 N_v$  である．通常の Viterbi アルゴリズムでは，尤度の更新に  $N^2 = 4N_{nv}^4 N_p^4 N_v^2$  回の確率計算が必要であるが，出力合流 HMM の特殊な遷移確率行列の形により，これを  $16N_{nv}^3 N_p^3 N_v^2$  とした効率化が可能である．さらに，テンポに関して，値が大きく異なる状態間の遷移確率を小さな定数とすると，さらなる計算量の削減ができる [9]．以上の工夫により削減を行っても，計算量は大きく，実用上は課題が残る．

前節で議論した大きな計算量の主な原因は，上のアルゴリズムでは，全ての音符に関してそれが両方の声部に含まれる可能性を残したまま推定を行っていることにある．このためには，例えば 88 音からなる音高の状態空間を導入する必要があるためである．一方，ピアノ曲などではポリリズムの様な複雑な多声部構造は，両手パートにまたがり存在することが多く，ピアノ曲の両手パートに対応する声部分離は，比較的に精度高く行えるため [8]，予め声部分離を行った上で，多声部のリズム採譜をすることも有効であると考えられる．声部分離をすることは，各音符に対して  $\eta$  を決定することに対応するが，この結果音高の状態空間を考慮する必要も無くなり，この時，状態空間が  $(\nu^{(1)}, \nu^{(2)}, v)$  で表せるような出力合流 HMM によりリズム採譜を行える．この際の Viterbi 更新の計算量は， $N_{vn}^3 N_v^2$  (テンポに関してはさらに上述の削減が考えられる) となり，非常に効率的なアルゴリズムが得られる．このアルゴリズムでは，前処理の声部分離の結果を用いて，各声部ごとにほぼ独立にリズム採譜を行うが，両者が一つのテンポに対して結び付けられることにより，声部間の音価の関連付けが保たれる．次節では，声部構造を考慮しない HMM によるリズム採譜との比較のため，ここで述べたリズム採譜アルゴリズムを用いる．

#### 5. 性能評価

本稿で提案するリズム採譜の有効性を確認するため，実際にアルゴリズムを実装し，リズム採譜の性能評価を行う．特に，提案するアルゴリズムによるリズム採譜に必要な処理時間の測定と，そのリズム採譜の結果と声部構造を考慮しない HMM による結果の比較，そしてより高精度なリズム採譜を行うための議論をする．

##### 5.1 処理時間の測定

前節で述べた，声部分離および出力合流 HMM によるリズム採譜に要する計算機の処理時間を測定する．どちらも，主な処理時間は動的計画法を用いた尤度計算により生じ，入力音符数にほぼ比例する．約 1000 音符の演奏データを用意し，各処理に必要な時間を測定したものを入力音符数で割り，1 音符当たりの処理時間を測定した．計算機環境は，CPU: 1.7 GHz Intel Core i7, Memory: 8 GB,

OS: Mac OSX 10.9 であった。10 回の測定の結果、平均処理時間は、声部分離に関して  $0.192 \pm 0.004$  ms/音符、リズム採譜に関して  $25.4 \pm 0.2$  ms/音符であった。同様の声部構造を取り入れた 2 次元確率文脈自由文法によるリズム採譜アルゴリズム [11] では、処理時間は大まかに 30 s/音符程度であり、比較すると本稿で提案するアルゴリズムは非常に高速であることが分かる。

## 5.2 リズム採譜の結果の比較

いくつかの楽節の人間による実演奏を用いて、リズム採譜の性能評価を行った。用いた楽節は、Brahms の Klavierstücke Op. 118 No. 2 の中間部の始めの 8 小節、Chopin の幻想即興曲の Allegro 部分より 4 小節、Debussy の Arabesque No. 1 より主題の 4 小節であり、いずれも 2 対 3 又は 3 対 4 のポリリズムを含む楽節である。結果の一部を図 4 に示す。図の採譜結果における上下段は、声部分離の結果を表しているが、元のピアノ譜の上下段と良く一致していることが確認できる。また、一部音価の推定を誤った箇所が見られるが、全体的に高い精度でリズム採譜が行われていることが分かる。本研究では、少数の演奏データに対して性能評価を行ったが、より多くのデータを用いた系統的な性能評価は、今後の課題である。

## 5.3 議論

前節で確認した通り、出力合流 HMM を用いたアルゴリズムによって、ポリリズムを含む楽節において比較的高い精度でリズム採譜が行えることが分かった。採譜結果を観察すると、音価の推定誤りが起こった箇所に置いては、2 つの声部の音価の累積値が異なり得ることが分かるが、出力合流 HMM を用いたアルゴリズムでは、これらが等しくなるような制約が入っていないためである。アルゴリズムにこうした制約を取り入れることにより、さらに高精度なリズム採譜ができる可能性がある。

また、場合により採譜結果の音価が、正解の楽譜に比べ全体的に定数倍された結果になることが起こった。リズム採譜を応用する状況にも依るが、音楽的には音価の絶対値よりも相対関係の方がより本質的であり、これは問題とならない状況も多いであろう。系統的にリズム採譜の性能評価を行う際には、隣り合う音価の比に基づく評価など、こうした状況を踏まえた評価手法も考察する必要がある。

最後に、本研究では、楽節モデルは音価に関する 1 次の Markov モデルをもとにしているが、小節あたりのリズムパターンを導入する等のモデルの改良も、さらなる精度向上の可能性を与える。一般的に、音楽はフレーズや動機の繰り返しなどの構造を持つが、こうした楽曲の構造をリズム採譜に用いることも、今後の大きな方向性の一つとして考えられる。

## 6. まとめ

本稿では、ポリリズム楽節のような複雑な構造を持つ多声部音楽の楽譜および演奏に対しても精度よくリズム採譜を行うために、声部構造を取り入れた確率モデルによるアルゴリズム構築を議論した。このようなモデルは、出力合流 HMM を用いて作られるが、計算コストを軽減するために、声部分離とリズム採譜を 2 段階に分けて行う手法を提案した。性能評価の結果、この手法により、声部を考慮しない従来の HMM に比べ、ポリリズム楽節の演奏に対して良好なリズム採譜を行えることを確認した。また、同様に声部を考慮した、2 次元確率文脈自由文法を用いた手法に比べ、高速な処理が可能であることも確認した。

本稿で提案した手法の系統的な性能評価試験の他、声部ごとの音価の累積値が揃うような制約を組み込むことが今後の課題である。また、より高精度なリズム採譜を行うためには、楽曲のフレーズや動機構造を利用して音価の推定を行うことが重要であると考えられる。

謝辞 本研究の一部は、文部科学省/学術振興会科学研究補助費 課題番号 (23240021) および課題番号 (25880029) から補助を受けて行われた。

## 参考文献

- [1] A. Klapuri and M. Davy, *Signal Processing Methods for Music Transcription*, New York: Springer, 2006.
- [2] P. Desain and H. Honing, "The quantization of musical time: a connectionist approach," *Comp. Mus. J.*, **13**(3), pp. 56–66, 1989.
- [3] A. T. Cemgil, P. Desain and B. Kappen, "Rhythm quantization for transcription," *Comp. Mus. J.*, **24**(2), pp. 60–76, 2000.
- [4] C. Raphael, "Automated rhythm transcription," *Proc. ISMIR*, pp. 99–107, 2001.
- [5] 齋藤直樹他, "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," 情処研究報告 SIGMUS-99(106), pp. 27–32, 1999.
- [6] 大槻知史他, "隠れマルコフモデルによる音楽リズムの認識," 情処論文誌, **43**(2), pp. 245–255, 2002.
- [7] H. Takeda *et al.*, "Hidden Markov Model for Automatic Transcription of MIDI Signals," *Proc. MMSP*, pp. 428–431, 2002.
- [8] 中村栄太, 齋藤康之, 嵯峨山茂樹, "出力合流並列隠れマルコフモデルとその多声鍵盤音楽の楽譜追跡・両手部分離への応用," 情処研究報告 (SIGMUS) 2013-EC-27, 2013.
- [9] E. Nakamura, *et al.*, "Outer-product hidden Markov model and polyphonic MIDI score following," *JNMR* **43**(2), pp. 183–201, 2014.
- [10] H. Kameoka *et al.*, "Context-free 2D Tree Structure Model of Musical Notes for Bayesian Modeling of Polyphonic Spectrograms," *Proc. ISMIR*, pp. 307–312, 2012.
- [11] 高宗 典玄, 亀岡弘和, 嵯峨山茂樹, "2 次元 LR パーサに基づく実演奏 MIDI 信号からの自動採譜," 情処研究報告 SIGMUS-104-8, 2014.
- [12] C. Palmer and C. van de Sande, "Units of knowledge



The image displays a musical score in 3/4 time with a key signature of three sharps (F#, C#, G#). It is divided into two main sections: 'Original score' and 'Transcribed score'. Each section contains two staves (treble and bass clef). The 'Original score' shows a piece of music with various rhythmic patterns, including triplets. The 'Transcribed score' shows the same piece but with several red annotations: a red note in the treble staff, a red '3' above a triplet in the bass staff, and a red bracketed triplet in the treble staff. The score is presented in three systems, with the first system showing the initial notation and the subsequent systems showing the continuation of the piece.

図 4 出力合流 HMM を用いたリズム採譜の結果 .

- in music performance,” *J. Exp. Psych.*, **19(2)**, pp. 457–470, 1993.
- [13] P. Desain, H. Honing and H. Heijink, “Robust score-performance matching: Taking advantage of structural information,” *Proc. ICMC*, pp. 337–340, 1997.
- [14] H. Heijink, L. Windsor and P. Desain, “Data processing in music performance research: Using structural information to improve score-performance matching,” *Behavior Research Methods, Instruments, & Computers*, **32(4)**, pp. 546–554, 2000.
- [15] B. Gingras and S. McAdams, “Improved score-performance matching using both structural and temporal information from MIDI recordings,” *JNMR*, **40(1)**, pp. 43–57, 2011.
- [16] E. Nakamura *et al.*, “A stochastic temporal model of polyphonic MIDI performance with ornaments,” to appear.
- [17] E. Nakamura *et al.*, “Merged-Output Hidden Markov Model for Score Following of MIDI Performance with Ornaments, Desynchronized Voices, Repeats and Skips,” *Proc. ICMC*, 2014.
- [18] E. Nakamura, N. Ono and S. Sagayama, “Merged-Output Hidden Markov Model for Piano Fingering for Both Hands,” *Proc. ISMIR*, 2014.
- [19] L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proc. IEEE*, **77(2)**, pp. 257–286, 1989.
- [20] M. Hashida, T. Matsui and H. Katayose, “A new music database describing deviation information of performance expressions,” *Proc. ISMIR*, pp. 489–494, 2008.
- [21] D. Conklin, “Representation and discovery of vertical patterns in music,” in A. Smaill (eds.), *Music and Artificial Intelligence*, Lecture Notes in Artificial Intelligence, Springer, pp. 32–42, 2002.
- [22] Z. Ghahramani and M. Jordan, “Factorial Hidden Markov Models,” *Machine Learning* **29**, pp. 245–273, 1997.