

# 確率文脈自由文法を用いた音楽演奏 MIDI データのリズム・テンポの推定\*

山本遼, 武田晴登, 酒向慎司, 嵯峨山茂樹 (東大・情報理工)

## 1 はじめに

MIDI キーボード等による音楽演奏の自動採譜は、楽譜や MIDI データの入力、楽譜のない演奏の保存・分析等において重要な技術である。テンポが未知かつ変動する演奏のリズムを推定する手法として、我々は演奏されるリズムパターンの確率的な偏りとテンポの連続的な変動を手がかりに、演奏データの IOI(Inter-Onset Interval) の局所的な比の時系列を HMM により確率的にモデル化し最尤推定を行う手法 [1] を提案し有効性を報告している。

我々が IOI を特徴量として利用してきた主な理由として、音楽演奏において演奏音の発音時刻と比べ消音時刻は意識的・無意識的な揺らぎが大きいことが挙げられる。しかし音楽演奏の背後には、その演奏の元となる楽譜の満たすべき厳密な構造が存在し、それは各音符の長さと密接に関連している。従って発音時刻の情報とともに揺らぎも含めた消音時刻の情報を用いることで、楽譜の構造を陽に利用した頑健なリズム推定が行える可能性がある。

今回我々は、多声部の演奏データを発音時刻の一次元の時系列に縮約せず、各音の発音・消音時刻を確率文法により直接モデル化し、最尤リズム・テンポを推定する新しいアプローチについての検討を報告する。

## 2 確率文脈自由文法によるリズム認識

### 2.1 問題の定式化

まず問題の定式化を行う。演奏データ  $Y$  は演奏音  $t = (t_{on}, t_{off})$  の集合であり、 $t_{on}, t_{off}$  はそれぞれ発音時刻・消音時刻からなる。演奏の音高の情報は本研究では利用しない。演奏データから推定するリズムとは、演奏各音の音符名とそれらの間の時間的構造を

分類	規則例 $A \rightarrow B C$	規則適用確率
の生成規則 [時間方向]	$s_{1/4} \rightarrow s_{2/4} s_{2/4}$ $s_{1/4} \rightarrow s_{1/4} s_{1/4}$ $s_{1/4} \rightarrow s_{3/4} s_{1/4}$ $s_{1/4} \rightarrow s_{1/8} s_{1/8}$	$P((B, h_B), (C, h_C)   (A, h_A), p)$ $= F_p(h_A, h_B, h_C)$ $= \begin{cases} 1, & \text{when } \left[ \begin{matrix} h_B^{on} - h_A^{on} = h_C^{on} - h_A^{on} \\ h_B^{off} - h_A^{off} = h_C^{off} - h_A^{off} \end{matrix} \right] \\ 0, & \text{else.} \end{cases}$
の生成規則 [音高方向]	$s_{1/4} \rightarrow s_{1/4} s_{1/4}$ $s_{1/4} \rightarrow s_{2/4} s_{2/4}$ $s_{1/4} \rightarrow s_{1/4} s_{1/4}$	$P((B, h_B), (C, h_C)   (A, h_A), q)$ $= F_q(h_A, h_B, h_C)$ $= \begin{cases} 1, & \text{when } \left[ \begin{matrix} h_B^{on} = h_A^{on} = h_C^{on} \\ h_B^{off} = h_A^{off} = h_C^{off} \end{matrix} \right] \\ 0, & \text{else.} \end{cases}$
演奏音の生成規則	$s_{1/4} \rightarrow \uparrow$ $s_{2/4} \rightarrow \uparrow$ $s_{1/4} \rightarrow \uparrow$	$P(\uparrow   (A, h_A))$ $= F_A(h_A, t)$ $= \begin{cases} 1, & \text{when } \left[ \begin{matrix} \alpha_{on}^{(A)} - \beta > 0 \\ \alpha_{off}^{(A)} - \beta > 0 \end{matrix} \right] \\ 0, & \text{else.} \end{cases}$

図 1 演奏生成文法の例と各規則の適用確率。例えば「時間方向」の生成規則により部分楽譜  $A$  から部分楽譜  $B, C$  が生成される時、 $h_B^{ON} = h_A^{ON}, h_B^{OFF} = h_A^{OFF}$  が成り立つ必要がある。さらに本研究では演奏者の意図するテンポは一定であると仮定したため、 $A, B, C$  の音価あたりの演奏時間は等しい必要がある。これらを「時間方向」の生成規則の適用確率として導入する。同様に「音高方向」の生成規則も導入する。「音高方向」の規則では、 $A, B, C$  の隠れ発音区間は完全に一致している必要がある。また部分楽譜からの演奏音生成は、演奏音の発音区間が隠れ発音区間から確率的に決まるとした。本研究ではこの確率を与える関数を、観測された演奏音に対する部分楽譜の隠れ発音区間が正規分布として導出できる形で与えた。

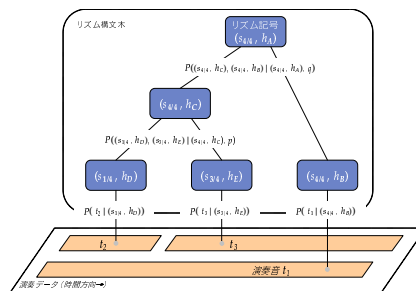


図 2 文法による演奏の生成図

表すものとする。音符間の構造には、主に音符と音符が「時間方向」に並ぶ(連続する)構造と、「音高方向」に並ぶ(和音となる)構造が存在し、これらは図 1 の「規則例」にあるような文脈自由文法により記述可能である。ここで  $S_x$  は音価  $x$  相当の音符の集合(以下部分楽譜)を表す。その他のリズムの重要な要素として休符の構造があるが本研究では扱わない。

リズムを  $X$  とすると、演奏データからのリズム推定問題は

$$X_0 = \arg \max_X P(X|Y) \approx \arg \max_X P(Y|X) \quad (1)$$

となる。

### 2.2 演奏生成文法

次にリズムからの演奏生成を確率的にモデル化する。シーケンサが楽譜を一定のテンポで正確に演奏する場合と比較して、人間の演奏は時間的な揺らぎを持っている。我々はこの揺らぎが演奏全体が時間的に伸縮するテンポ変動と、個々の音それぞれにおける独立な発音・消音時刻の揺らぎからなると仮定した。

別の言い方をすると演奏の背後に楽譜の機械的な演奏(正確なテンポで各音符を 100% の長さで演奏したもの)を奏者が意図するテンポどおりに時間的に伸縮した、演奏の意図とでも言うべき仮想的な演奏が隠れていて、実際の演奏はこの仮想的な演奏に対して各音が揺らぎを持ったものとして生成されるとするモデルである、とも言える。

上記のモデルを踏まえ、図 1 のように演奏の生成をモデル化した。楽譜の文法の各シンボル(部分楽譜と呼ぶ)には、その部分楽譜の演奏が意図される時間領域「隠れ発音区間」が存在するとし、これを  $h_A = (h_A^{on}, h_A^{off})$  により表す。ここで  $h_A^{on}, h_A^{off}$  は隠れ発音区間の発音と消音の時刻である。「時間方向」「音高方向」の並びの制約は、図の規則適用確率によって表現される。

### 2.3 尤度計算と最尤リズム探索

リズム構文木からの演奏が生成される尤度の計算を図 2 において説明する。図の 5 つの規則の各適用確率の積は  $P(t_1, t_2, t_3, (s_{1/4}, h_D), (s_{3/4}, h_E), (s_{1/4}, h_B), (s_{1/4}, h_C), (s_{1/4}, h_A))$  となる。これを  $h_A, \dots, h_E$  について周辺化すると  $P(t_1, t_2, t_3, s_{1/4}, s_{3/4}, s_{1/4}, s_{1/4}, s_{1/4}) = P(X, Y) \propto P(Y|X)$  となる。

\* Estimation of Rhythm and Tempo of MIDI Performance Using Stochastic Context-Free Grammar. by Ryo Yamamoto, Haruto Takeda, Shinji Sako, and Shigeki Sagayama (The University of Tokyo)

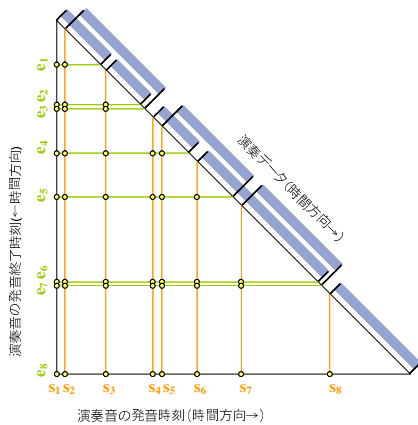


図3 最尤リズム探索のためのマトリクス

この計算は逐次的に行うことができる。観測として演奏音  $t_1, t_2, t_3$  が与えられたとき、演奏音の生成規則の適用確率関数  $F_r$  より、各部分楽譜  $B, D, E$  について隠れ発音区間  $h_B, h_D, h_E$  の尤度分布が得られる。次に  $h_D, h_E$  の尤度分布と  $F_p$  から部分楽譜  $C$  の発音区間  $h_C$  の尤度分布が得られる。このとき  $h_D, h_E$  の尤度分布は  $C$  より上位の適用確率と独立であるため周辺化を行うことができる。次に同様に  $h_B, h_C$  から  $h_A$  の尤度分布を求めることができ、最終的にこの  $h_A$  の積分値が構文木の尤度となる。このように逐次的に尤度分布算出と周辺化を繰り返すことで尤度計算が行える。また隠れ発音区間の長さからテンポの推定が行える。

通常確率文脈自由文法における最尤導出の探索には CYK アルゴリズム等が使われる。しかしこれらの手法は終端記号が一次元の系列であり木構造が交差ししないものにしか用いられない。演奏音にはこのような性質がないため直接このようなアルゴリズムを利用することはできず、探索範囲が膨大になってしまう。

しかし演奏音  $t_1$  と  $t_2$  が「時間方向」の規則からの導出であれば、実際の消音-発音時刻間隔  $t_2^{on} - t_1^{off}$  はある程度小さいと考えられるため、閾値  $T$  により探索範囲を狭めることができる。「音高方向」の規則についても同様である。

以上から図3に示すような、演奏データの各音の発音・消音時刻の交点に候補部分楽譜を並べる構文解析用マトリクスにより効率的に演奏の構文解析が行えることを示す。この構文解析器は各部分楽譜を、その部分楽譜に含まれる演奏音の最初の音の開始時刻と、最後の音の消音時刻の交点にあたるマトリクスの成分に格納することで、冗長な探索を防ぐ。

最尤リズムを探索するアルゴリズムを図3の例に従い以下に示す。

1. 各演奏音について、各演奏音生成規則を用いてそれを導出する部分楽譜を、その演奏音の発音時刻、消音時刻に対応するマトリクス成分に入れる
2. 対象とするマトリクス成分  $(s_i, e_j)$  を、マトリクスの対角線側から  $(s_1, e_8)$  へ向かい動かしながら、3 から 6 までを実行
3. for  $k = j - 1$  to 1  
for  $l \in \{m \mid |e_k - s_m| < T_1, m > i\}$   
 $(s_i, e_k)$  成分にある部分楽譜と  $(s_l, e_j)$  成分にある部分楽譜を「時間方向」の文法で導出する部分楽譜を  $(s_i, s_j)$  に入れる
4. for  $l \in \{m \mid |e_j - e_m| < T_1, m < j\}$   
for  $l \in \{m \mid |s_i - s_m| < T_1, m > i\}$   
5, 6 を実行
5.  $(s_i, e_m)$  成分の部分楽譜と  $(s_n, e_j)$  成分の部分楽譜を「音高方向」の文法で導出する部分楽譜を  $(s_i, e_j)$  成分に入れる
6.  $(s_i, e_j)$  成分の部分楽譜と  $(s_n, e_m)$  成分の部分楽譜を「音高方向」の文法で導出する部分楽譜を  $(s_i, e_j)$  成分に入れる

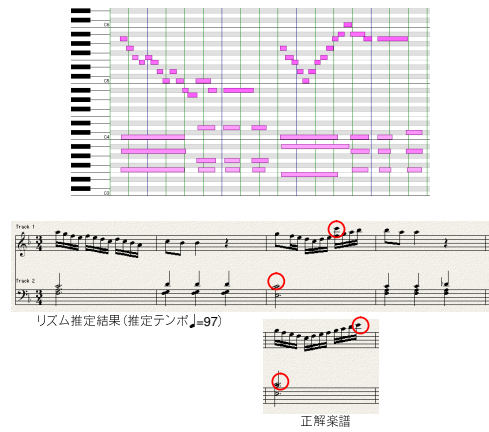


図4 入力演奏のピアノロール表示(上)とリズム認識結果(下)

7.  $(s_1, e_8)$  にある導出のうち、全演奏音を導出している最尤の部分楽譜をリズム推定結果とし、同時に隠れ発音区間の長さからテンポを推定する。

### 3 実験結果例

本手法の動作確認のため、以上のアルゴリズムを実装し Burgmuller のピアノ練習曲の演奏データにより、リズムとテンポの推定実験を行った。4/4 と 3/4 拍子の楽譜に対応した演奏生成文法を導入し、演奏生成文法の揺らぎを与えるパラメータは実験結果を元にヒューリスティックに決定した  $(\alpha_{on} \sim N(0, 0.01), \alpha_{off} \sim N(0, 0.1), \beta = 0.2)$ 。構文解析器における閾値  $T$  はこの演奏においては十分大きな 500ms とした。

演奏データのピアノロール表示と結果の楽譜出力を図4に示す。正解と比較すると第3小節の認識結果に誤りがある。これは正解楽譜における小節先頭と小節末の2音が、「時間方向」に並ぶと誤認識したものである。この結果は現在のヒューリスティックなパラメータの設定では演奏音の揺らぎのモデル化が不十分であることを伺わせる。今後これらの学習が課題となる。また上記の探索アルゴリズムは CYK アルゴリズム同様、演奏音の3乗に比例した計算量が必要であるため、今回の実験で用いた程度の音符数が計算量的に限界であった。効率的な枝刈りによる計算量削減も今後の課題である。

### 4 おわりに

楽譜の規則を包含する確率文法により多声部の演奏をモデル化し、最尤リズムを推定する手法を提案し、動作実験を行った。今後の課題として、演奏音の揺らぎの学習、探索における枝刈り、テンポの変動のモデル化、リズムに対する事前知識の導入、ピッチ情報の利用等が考えられる。

謝辞 本研究の一部は科学研究費補助金・基盤研究 B(課題番号 17300054) および科学技術振興機構 CREST プロジェクトの補助を受けて行なわれた。

### 参考文献

- [1] 武田ほか, “HMM を用いたリズムとテンポの反復推定による多声 MIDI 演奏のリズム認識,” 音講論(春), pp. 721-722, 2006.