

# 2次元LRパーサを用いた 音楽演奏MIDI信号からの自動採譜

高宗 典玄<sup>1,a)</sup> 亀岡 弘和<sup>1,b)</sup> 嵯峨山 茂樹<sup>2,c)</sup>

概要：本発表では、MIDI信号から楽譜を推定する問題を扱う。リズムとテンポの不確実性があるこの問題に対し、それぞれの生成過程を考え、統合的に解くことによりその解決を試みる。ここで、リズムの生成モデルとして2次元PCFGを用い、その解析アルゴリズムとして2次元LRパーサを提案する。実際に人間が演奏したいくつかのMIDI信号に対し採譜実験を行い、その結果を示す。なお、本発表は [1], [2] の内容をまとめたものである。

## 1. はじめに

自動採譜とは、人間が演奏した音楽音響信号から自動で楽譜を推定することであり、音声信号処理における音声認識の位置に対応する、音楽音響信号処理における重要な課題のひとつである。つまり、自動採譜ができるようになれば、単に楽譜がない曲の楽譜が得られるだけでなく、楽曲の検索や推薦といった様々な応用が期待できる。

自動採譜には大きく分けて2つの段階がある。1つは音響信号から音高や発音時刻、消音時刻を推定する多重音解析であり [3]、もう1つは多重音解析で推定された音高や発音時刻、消音時刻から演奏速度(テンポ)の変動や音価を推定するリズム解析である [4]。本発表では、後者のリズム解析の問題に焦点を当ててMIDIからの自動採譜を目指す。

## 2. 2次元LRパーサを用いたリズム解析

リズム解析の問題は、人間の演奏が多くの場合、テンポ変動や各音の発音・消音時刻のゆらぎを含んでいるため、非常に困難な問題となっている。観測上の時間においての音の長さは、楽譜上の音の長さやテンポに依存するため、どのような楽譜から演奏されたかどのようなテンポ変動を行ったかを一意に分解することは不可能である。リズム解析の問題の難しさはこのことに由来する。しかし、人間は多くの場合、そのようなテンポ変動や発音・消音時刻のゆらぎを含んだ演奏を聴いても、楽譜やテンポを推定することができる。これは、人間は楽譜やテンポに対して先験知識があり、その先験知識に沿ったような楽譜とテンポを推定しているためであると考えられる。そこで、本研究では先験知識として楽譜やテンポの確率的生成モデルを構築し、それを確率的逆問題として楽譜やテンポを同時に推定することを考える。

## 2.1 確率的生成モデル

### 2.1.1 楽譜の生成モデル

音楽を時間方向に見ていくと、モチーフやフレーズなどといった階層構造を持つことが分かる。この構造と、自然言語が単語や文節、文といった階層構造を持つことは似ているため、そのような階層構造を扱う自然言語処理で用いられているモデルである確率文脈自由文法 (Probabilistic Context-Free Grammar; PCFG) を利用できないかということが考えられる。一方、音楽には声部やパートといった、音高方向の階層構造も有しているため、単純にPCFGを音楽に適用することはできない。そこで、このような、時間方向と音高方向の二つの方向の階層構造を表すモデルとして、我々の研究室で提案された2次元PCFG [5] を用いることを考える。これを用いることで、起こりうるすべての楽譜に対し、少ない生成規則で生成確率を計算することができるようになる。ここで実際の楽譜で頻出するリズムが出現する確率を高く、あまり現れないリズムが出現する確率を低く設定することで、楽譜の尤もらしさを生成確率で表現できる。

### 2.1.2 テンポ・発音時刻・消音時刻の生成モデル

人間が音楽を演奏する際、テンポは必ずしも一定ではなく、緩やかな変動を伴っている場合がほとんどである。また、各音の発音時刻や消音時刻も物理的な制約や、演奏上の表情付けで楽譜上の位置からずれることがある。そこで、楽譜上のどの位置にいるかを表す有限状態空間を用いてテンポ変動や発音時刻・消音時刻のずれの生成をモデル化することを考える。このモデルは、各拍の時間を表したleft-to-rightのパスを複数持っており、また、このモデルにおいて、演奏の時間がある単位時間進むごとに状態を遷移し、楽譜上の発音時刻、消音時刻に対応する状態の前後の状態から発音や消音の指令を出力する。ここで、パス間の遷移確率は1拍の長さがどう変化しやすいか、つまりテンポ変動の起きやすさを表す。そこで、緩やかなテンポ変動が高い確率となるように、前の拍の長さと同じぐらいの長さのパスほど遷移確率が高くなるように設計する。

## 2.2 解析アルゴリズム

### 2.2.1 2次元LRパーサ

PCFGを解析する手法には一般化LR法やCYK法等がある。しかし、これらの手法は時間方向の順序関係が重要であるので、2次元PCFGが生成する記号列の順序と観測される記号列の順序が異なることから、音高方向に拡張した2次元PCFGには単純に適用することができない。そこで、様々な並びを考慮に入れつつもビームサーチで確率が低い候補を切り落とすことで、計算量を削減するという

<sup>1</sup> 東京大学大学院情報理工学系研究科  
The University of Tokyo  
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

<sup>2</sup> 明治大学総合数理学部  
Meiji University  
4-21-1 Nakano, Nakano-ku, Tokyo, 164-8525, Japan

a) Norihiro.Takamune@ipc.i.u-tokyo.ac.jp

b) kameoka@hil.t.u-tokyo.ac.jp

c) sagayama@meiji.ac.jp

方針を考える。よって本研究では、時間的に順番に解析していき、解析途中で確率を評価することができる PCFG の解析手法である一般化 LR 法 [6] を応用することを考える。

一般化 LR 法において、解析器 (以下パーサと呼ぶ) の状態はスタックを用いて表現されており、現在入ってきた入力とスタックの先頭の状態からどのような解析を行うかという LR テーブルと呼ばれる規則をあらかじめ計算して保持している。そのため、一般化 LR 法のパーサは入力を逐次的に読み込みながらその LR テーブルに従い解析を行うので、非常に効率のよい構文解析手法である。このとき、どの行動を行うかが複数存在する場合はその分だけ仮説を増やしたのち、それぞれ計算する。このため、一般には仮説が指数関数的に増大し、計算が困難になるが、ビームサーチなどを行うことにより、近似的に計算が可能となる。

この一般化 LR 法を 2 次元 PCFG に応用するために、本研究ではスタックを複数内包している 2 次元 LR パーサを提案する。2 次元 LR パーサは、LR テーブルの生成のところで音高方向の分割規則に対しての処理を新たに導入し、解析を行う段階で、音高方向の分割に対応した複数のスタックを同期して処理を行うパーサである。2 次元 LR パーサに入力がきたときは、内包する複数のスタックのうちどれに割り振るか、また、新しくスタックを作成するかの仮説を増やし、それぞれ計算する。このとき、LR テーブルにより音高方向の分割規則に対応した処理が指定された場合、他のスタックのうち対応するものがあるかを調べ、対応するスタックがあったらそれらのスタックの統合を行う。

### 2.2.2 有限状態空間内を遷移する 2 次元 LR パーサ

2.2.1 節で示した 2 次元 LR パーサに入力されるシンボルは楽譜上の発音時刻、消音時刻であるべきであるので、2.1.2 節で示したテンポや発音時刻、消音時刻の生成モデルを考えると、次の解析アルゴリズムが考えられる。2 次元 LR パーサが楽譜上の位置を表す状態空間内を単位時間ごとに遷移していき、実時間上の発音時刻や消音時刻が観測されると、現在いる状態に対応する楽譜上の発音時刻、消音時刻を推定し、それを 2 次元 LR パーサの入力とし解析する。以上より楽譜の生成モデルとテンポ・オンセット・オフセットの生成モデルを統合したモデルの解析アルゴリズムを構築できる。

## 3. 採譜実験

提案手法の動作を確認するために、F. Beyer 作曲 Op. 101 の No. 51~59 と C. Debussy 作曲 “Arabesque No. 1” の 6~9 小節を人間が演奏した MIDI 信号の計 10 曲を解析した。

比較対象として、武田らによって提案された手法 [4] を用いた。この手法は、時間方向の 1 次元に射影した隠れマルコフモデルに基づく手法である。以下この手法を従来法と表記する。

また、提案法において、状態遷移する単位時間を 30 ms、1 拍の長さの最小値を 0.3s、最大値を 1.5s とし、2 次元 PCFG の各文法適用確率は人手で適当に設定した。

楽譜の推定結果を評価する際に、テンポの定数倍の誤りや開始位置の誤りを取り除くため、次で定義される最小平均誤差を用いた。

$$J = \frac{1}{I} \min_{a,b} \sum_{i=1}^I (y_i - ax_i - b)^2 \quad (1)$$

ただし、正解の楽譜の音符  $i$  の発音時刻の位置は  $y_i$ 、各手法の対応する音符の発音時刻の位置は  $x_i$ 、 $I$  は各楽譜の音符の数である。各曲に対する平均最小誤差とすべての曲に対して音符数の重み付き平均した値を表 1 に示す。多くの場合、従来法を上回る結果が得られた。しかし F. Beyer 作曲 Op. 101, 55 番については悪い結果となった。これは、実際の譜面にあるリズムが今回用いた文法で表しづらいものがあつたか

表 1 各曲ごとの最小平均誤差

曲名	提案手法	従来法
Beyer, No. 51	$1.33 \times 10^{-4}$	$6.85 \times 10^{-2}$
Beyer, No. 52	0	$1.85 \times 10^{-2}$
Beyer, No. 53	0	0
Beyer, No. 54	0	0
Beyer, No. 55	$2.99 \times 10^{-2}$	0
Beyer, No. 56	0	$1.64 \times 10^{-1}$
Beyer, No. 57	0	$9.35 \times 10^{-1}$
Beyer, No. 58	0	0
Beyer, No. 59	0	$1.05 \times 10^{-2}$
Arabesque	$1.68 \times 10^{-3}$	$9.10 \times 10^{-2}$
総平均	$6.71 \times 10^{-3}$	$1.17 \times 10^{-1}$



(a) 正解楽譜



(b) 提案手法による採譜結果



(c) 従来手法による採譜結果

図 1 C. Debussy 作曲 “Arabesque No. 1” の 6~9 小節の正解楽譜 (a)、提案手法による採譜結果 (b)、従来法による採譜結果 (c)。

らと考えられる。そのため、文法の拡張を行うことで改善が期待できる。また、C. Debussy 作曲 “Arabesque No. 1” の 6~9 小節の正解となる楽譜とそれぞれの手法の採譜結果を Fig. 1 に示す。この曲は 2 声部あり、上パートが三連符、下パートが八分音符となっているため、声部の分離を同時に行う本手法では非常に正解に近い結果を得られた。

## 4. まとめ

本研究では、リズムとテンポの不確定性を含むリズム解析問題に対し、リズムやテンポ、発音時刻、消音時刻の生成過程を考え、統合的に解くことによりその解決を試みた。そして、リズムの生成モデルとして 2 次元 PCFG を用い、その解析アルゴリズムとして 2 次元 LR パーサを提案した。実験により、多声部な曲に対し、本手法の採譜結果を示した。

謝辞 本研究は JSPS 科研費 26240025 の助成を受けたものです。

## 参考文献

- [1] 高宗ら, “2 次元 LR パーサによる音楽演奏 MIDI 信号からの自動採譜,” 音講論 (春), No. 3-5-4, pp. 1039-1042, 2014.
- [2] 高宗ら, “2 次元 LR パーサに基づく実演奏 MIDI 信号からの自動採譜,” 情報処理学会研究報告, Vol. 2014-MUS-104, No. 7, 2014.
- [3] Benetos, E., et al. “An efficient shift-invariant model for polyphonic music transcription,” Proc. of MML. 2013.
- [4] 武田ら, “確率モデルによる多声音楽演奏の MIDI 信号のリズム認識,” 情処論, Vol. 45, No. 3, pp.670 - 679, 2004.
- [5] Kameoka, H., et al. “Context-free 2D tree structure model of musical notes for Bayesian modeling of polyphonic spectrograms,” Proc. of ISMIR. 2012.
- [6] Tomita, M. “Efficient parsing for natural language: a fast algorithm for practical systems,” Vol. 8. Kluwer Academic Pub, 1985.