4R-9

二次元リズム木構造表現の確率モデルに基づくMIDI信号からの自動採譜

土屋政人† 落合和樹† 亀岡弘和†‡ 嵯峨山茂樹†

†東京大学大学院情報理工学系研究科 ‡NTTコミュニケーション科学基礎研究所

1 はじめに

自動採譜とは音楽音響信号からの楽譜推定を指し，即興演奏や音楽データベースの楽譜化等に利用することができる．この自動採譜課題は大きく分けると，音響信号から各楽音の発音開始時刻と基本周波数を推定する多重音解析の問題，そして多重音解析で得られた情報から音符列に変換するリズム解析の２つの問題がある．本稿では後者のリズム解析の問題を取り扱う．

リズム解析の問題の本質は，次の式に示すような，観測された実時間上の入力の長さから想定したテンポと音価長を推定しようとする際に起きる解釈の任意性に起因している:実時間 (sec) ＝音価 (Tick) ×テンポ (sec/Tick)．人間が音楽を聞いた時にその楽譜を想像することができるのは，我々は音楽として常識的なリズムやテンポに関する知識を持っており，そういった先験的な情報と各楽音の発音開始時刻という観測情報の両側面から総合的に解釈を行っていると考えられる．そこで，我々の研究室では楽譜を２次元木構造で表現し，自然言語処理の手法を用いて「リズムとテンポの自然さや起こりやすさ」を確率的に扱い，自動採譜問題を確率的逆問題として捉えたアプローチを提案してきた [1].

しかし，自然言語処理の観点から見た時，音楽の文法をどのように定義すればいいのかは全くの未解明である．そこで，本研究では音声認識分野とのアナロジーを考え，２次元木構造モデルに音符列を単語に模した「リズム語彙」を導入し，実験ではその効果を検証する．

2 二次元リズム木構造表現に基づく楽譜の生成モデル

採譜とは観測信号に最もよく適合するような楽譜を推定する処理であるが，楽譜が生成されるメカニズムおよび各楽音の発音開始時刻 (観測データ) が生成されるプロセスを確率モデルとして表現し，観測データがどういうメカニズム・プロセスで生成されたらしいかを推定することで自動採譜を行おうというのが本研究の基本戦略である．そこで，以下，楽譜がどのようなプロセスで生成されるか，楽譜から各音の発音開始時刻がどのようなプロセスで決定されるか，ということについて議論していく．音楽は時間方向に関して，リズムパターン，モチーフ，フレーズといったいくつかの階層構造を持ち，音高方向に関してもボイシングやパートといった階層構造が見られる．そこで，図１のようにトップダウン的に音符を時間方向に分割 (Time-Spanning) したり，音高方向にコピー (Synchronization) したりするのを繰り返すことで多くの楽譜を生成することができると考えられる．このプロセスは自然言語処理のPCFG を用いて定式化することができ，[3] に倣って確率的定式化を行うと図３のようになる．木構造の各ノードはまず第一の選択として分割するのをやめて（終端記号）音符を生成するか，分割して２つの子ノードを生成するかの選択を行い，そして，後者が選ばれた場合はさらにどのような分割生成規則 PR_n によって生成を行うか決定する．このプロセスをすべてのノードで終端記号が生成されるまで繰り返す．

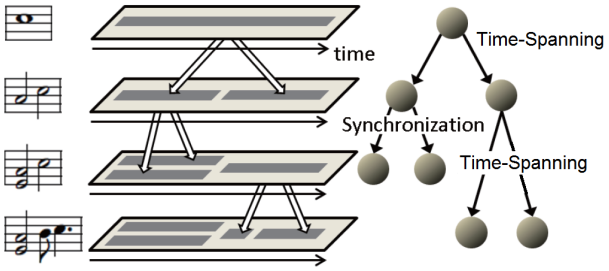

図 1: ２次元木構造表現による楽譜の生成

多くの楽曲は１曲の中で同じようなリズムを含むが，この性質がリズムとテンポの任意性を解消する手がかりになる可能性がある．そこで，図２に示すような小さなリズムパターンをまとめたリズム語彙 [2] を導入し，各リズム生成規則に適用確率を付与することによって同じリズムを繰り返し使う楽譜ほど高い確率値を持って生成されるように定式化した．木構造の事後分布推

Automatic Transcription of MIDI signals based on probabilistic model of 2-dimensional rhythm tree structure representation

Masato TSUCHIYA†, Kazuki OCHIAI†, Hirokazu KAMEOKA‡, Shigeki Sagayama†

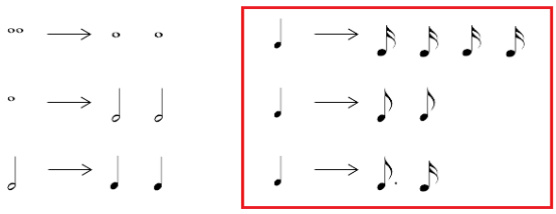

図 2: 定義した文法の一部.(赤枠がリズム語彙)

Draw rule probability :
$\phi^T \sim \mathrm{Beta}(\phi^T; 1, \beta^T)$
$\phi^B \sim \mathrm{Dirichlet}(\boldsymbol{\phi}^B; 1, \beta^B)$
For each node in the parse tree :
$b_n \sim \mathrm{Bernoulli}(b_n; \phi^T)$
If $b_n =$ EMISSION
$S_r \sim \delta_{S_r, S_n}, \quad G_r \sim \delta_{G_r, G_n}$
If $b_n =$ BINARY-PRODUCTION
$PR_n \sim \mathrm{Categorical}(PR_n; \phi^B)$
$G_{n1} \sim \delta_{G_{n1}, \mathrm{L}(PR_n)}, \quad G_{n2} \sim \delta_{G_{n2}, \mathrm{R}(PR_n)}$
If PR_n が SYNCHRONIZATION 型の時
$S_{n_1} \sim \delta_{S_{n_1}, S_n}, \quad S_{n_2} \sim \delta_{S_{n_2}, S_n}$
If PR_n TIME-SPANNING 型の時
$S_{n_1} \sim \delta_{S_{n_1}, S_n}, \quad S_{n_2} \sim \delta_{S_{n_2}, S_n + \mathrm{Len}(G_{n_1})}$

図 3: 楽譜の 2 次元木構造生成モデルの確率的定式化
但し，$\mathrm{L}(\cdot), \mathrm{R}(\cdot)$ は分割規則の左・右の記号を返し，$\mathrm{Len}(\cdot)$ はその記号の楽譜上の長さを返す.

定と，文法適用確率の推定を交互に繰り返すことで，できる限り 1 曲の中で同じリズムが適用されるようにできる．今回は解析手法として動的計画法に基づく確率計算を行うため，リズム語彙を Chomsky 標準形に変換し，文法を記述した．

最終的な発音開始時刻はテンポ変動と演奏ゆらぎの影響を受ける．そこで，拍時刻 μ_d に 1 次マルコフ性を仮定し，$p(\boldsymbol{\mu}) \propto \prod_{d=2}^{D} \mathcal{N}\left(\mu_d; \mu_{d-1}, (\sigma^\mu)^2\right)$ のように前の拍位置から正規分布分のゆらぎを持つようにモデル化を行うことでなめらかなテンポ変動を表現することができる．さらに，演奏ゆらぎを表現するため，$p(\boldsymbol{\tau}|\boldsymbol{\psi}, \boldsymbol{S}) = \prod_r \mathcal{N}\left(\tau_r; \psi_{S_r}, (\sigma^\tau)^2\right)$ のようにテンポ変動を受けた後の本来の発音開始時刻から正規分布のゆらぎを持つように最終的な発音開始時刻を生成する．

各楽音の発音開始時刻を観測データとし，それをもとに以上の生成モデルにおいてなされた決定プロセスを推定したい．紙面都合上詳細は省略するが，変分ベイズ法に基づき木構造の事後分布および文法適用確率の事後分布を近似計算する反復アルゴリズムを導出した．詳細は [3] を参照されたい．

表 1: リズム語彙の導入による正答率の変化

楽曲	小節	bps	導入前	導入後
Mozart Piano Sonata No.15	1-4	80	89.3%	97.8%
	9-12		40.2%	57.9%
Bartok Roumanian Folk Dance No.1	19-26	80	62.5%	89.0%
Bartok Roumanian Folk Dance No.2	1-12	144	51.4%	86.6%

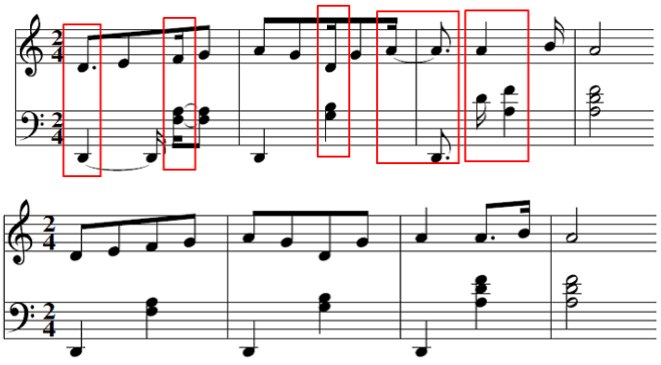

図 4: 上：リズム語彙なし，下：リズム語彙あり

3 MIDI からの採譜実験

楽譜を 2 次元リズム木構造で表現した効果を検証するために小規模な実験を行った．入力は実演奏を記録した MIDI ファイルで，演奏された元の楽譜を出力する．曲の長さと拍子，そして分割生成規則及びリズム語彙の辞書は手入力で与えた．比較対象はリズム語彙を導入していない，2 次元の PCFG モデルでの推定結果であり，各音符の開始位置の正答率をそれぞれの手法で算出した．その結果を表 1 に示す．採譜精度の向上が見られ，特に速いテンポの曲では大きく正答率が上昇した．出力された楽譜の一部を図 4 に示す．

4 結論

本稿では 2 次元リズム木構造を用いた楽譜のモデル化によって MIDI 信号からの採譜を行う方針の採譜精度について検証を行った．実験ではリズム語彙の効果によって採譜精度が改善する結果が得られた．今後はより効率的な採譜のための構文解析アルゴリズムの考案や音響入力からの採譜実験を行う予定である．

参考文献

[1] H. Kameoka et al, “Context-free 2D tree structure model of musical notes for Bayesian modeling of polyphonic spectrograms,” in Proc. ISMIR2012, in CD-ROM, Oct. 2012.

[2] H. Takeda et al, “Rhythm and tempo analysis toward automatic music transcription,” in Proc. ICASSP, vol. 4, pp. 1317-1320, Apr. 2007.

[3] P. Liang et al. “The infinite PCFG using hierarchical Dirichlet processes.” in Proc. EMNLP-CoNLL, pp. 688-697, Jun. 2007.