

芸術的逸脱の生成モデルにおける 決定木のバックオフに基づく鍵盤楽器演奏の自動表情付け

奥村 健太^{1,a)} 酒向 慎司^{1,b)} 北村 正^{1,c)}

概要: 自動演奏表情付けの問題を考え、それを解くための一手法を提案する。提案手法では、既存のあらゆる演奏事例に対して演奏者自身の楽曲解釈を代替する情報を関連付けた生成モデルを使用することで、そのような演奏事例が生成されるためのルールの組み合わせを体系化する。また、生成モデルの持つ決定木の構造に従って適切な事例の選択するための指標を導入することで、既存の演奏事例に含まれる演奏者の特徴を忠実に反映した演奏表情を未知の系列に転写する。各種評価の結果から、提案手法により生成された未知の演奏は、既存の演奏に見られた人間性や音楽性を示す特徴を備えたものとなることが示された。

Automatic performance rendition for keyboard instruments based on decision tree backing-off in generative model of the artistic deviation

KENTA OKUMURA^{1,a)} SHINJI SAKO^{1,b)} TADASHI KITAMURA^{1,c)}

Abstract: We discuss the problem of generating a virtual musical performance with the characteristics of the existing ones. Our method systematizes the rules for generation of musical expression by using a generative model that associates the information to articulate the performers' own interpretation of musical compositions and characteristics found in their performance. It transcribes appropriate characteristics to any unknown music by investigating the whole samples of existing performances with a criterion according to the topology that is constructed with the generative models. The results of evaluations show that the humanic or musical features of existing performance are also found in the generated unknown performances.

1. はじめに

本稿では、自動演奏表情付けの問題を解くための一手法の提案を通じて、与えられた楽譜からいかにして表情の付いた演奏が作られるべきかを論じる。この問題において一般的かつ最大の関心は人間の演奏者がするような表情豊かな演奏を機械の上で自動的に行うことにあるが、演奏という行為のどの部分を自動化するのかによって課題や提案は多岐に亘る [1]。本稿で自動化の対象とするのは、人間の演奏者が楽譜から演奏を生成するまでの過程である。

その過程を自動化する場合において一般的に用いられているのが、人間の演奏者が実際に行った演奏の素片を事例

としてプールし、そのいずれかを選択して表情の生成に用いるアイデアである。それに基づくアプローチでは、有限な事例の演奏者に対する忠実さを保持しつつ、未知の楽譜に対して自然な滑らかさを持つ演奏表情を生成することが一般的なゴールとなる。生成に際して使用する事例の選択方法によってそれらを分類すると、複数の事例における統計量を用いるもの [2], [3], [4] と、単独の事例の特徴を直接転写するもの [5] が具体的な手法として挙げられる。前者はどのような楽譜に対しても比較的滑らかな演奏が得られるが、その特徴が平滑化されるため演奏者の特徴が十分に保持されない。逆に後者は事例の演奏者に忠実ではあるといえるが、本来連続していない事例が接続されるため、楽譜によっては演奏の滑らかさに不自然さが生じうる。これらの場合において解決すべき共通の課題は、未知の楽譜の内容に対しても演奏者らしい特徴を保持しつつ自然な滑ら

¹ 名古屋工業大学
Gokiso-cho, Showa-ku, Nagoya, Aichi, 466-8555 Japan
a) k09@mmsp.nitech.ac.jp
b) sako@mmsp.nitech.ac.jp
c) kitamura@mmsp.nitech.ac.jp

かさを与える表情を持つ事例をどのようにして適切に選択するかである。

演奏においてどのような表情が生じる傾向があるかは、楽譜から一意に得られる情報を関連付けた確率モデルによって説明できることがわかっている。それは表情の類似性を楽譜の情報によって体系化できる [6]。そこからは演奏者がどのような場合にどのような演奏をする傾向があるのかを特定できるため、未知の楽譜の演奏に対して付与すべき表情を推定する際にも有効な手法となると考えられる。本稿では事例の演奏者に対する忠実さを保持することを重視し、事例の特徴を直接演奏に転写する場合を扱う。しかし基本的な枠組みとしての提案手法は、先述のアプローチにおけるどちらの考え方にも適用可能なものである。

2. 手法の提案

楽器はその種類によって演奏方法だけでなく記譜体系も大きく異なるため、同一の枠組みをあらゆる楽器に対して同様に適用できるものとするのは非常に困難である。楽器単体で完結する独奏楽曲が豊富であることや、楽器の操作に関する情報を MIDI を介して容易に扱えることなどの利点を得るため、本提案では対象を鍵盤楽器に限定して話を進める。これによって、演奏内容は MIDI ストリームから note on/off や set tempo のようなイベントとして観測できる。また、演奏として観測した事例に対する楽譜の対照を容易にするため、楽譜の内容は全て MusicXML によって記述が可能であるとする。

2.1 問題の設定

演奏者が楽譜に対して演奏を生成する過程を自動化するために解くべき問題は、何らかの楽譜に対する演奏に付与されるのに最適な表情を持つ事例を、プールの中からいかに選択するかである。しかし、生成する演奏内容全体に対してどの事例を選択するのが最適であるかをあらゆる事例の組み合わせについて単純に検討するのは、計算コストの面から現実的ではない。

表情は、与えられた楽譜に対する演奏者の解釈に従って演奏に付与されていると考える。この仮定には演奏者が生成する表情は楽譜から全て説明できることが前提として必要であるが、実際には演奏者が楽譜をどのように解釈しているかを一般的に説明することは困難である。表情と楽譜を関連付ける提案は従来からなされているが、それらにおいて使用されてきたのは、専門家からの聴取によって得られた知見、各種楽曲解析手法の結果、演奏表現を形容する自然言語文などである [7], [8]。そのような情報も一意には決定せず、演奏者以外の人間の意図が付随するため、一般性において適当ではない。

先行研究 [6] から、高次元楽譜解釈ではなく、楽譜から音符とその近傍のみという局所的な範囲から一意に得られる

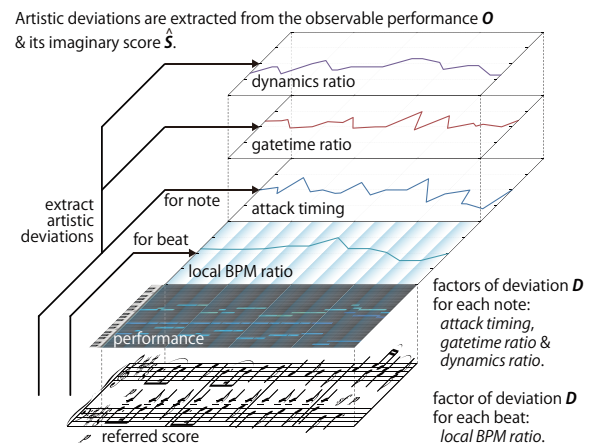


図 1 音符と拍に関する各種芸術的逸脱要素の抽出
Fig. 1 Extraction of factors for artistic deviation.

情報を使用することで、どのような音符列に対してどのような演奏表現が生じる傾向があるかを関係付けて説明できることがわかっている。そこで、この手法を演奏表情付けにおいて生成する演奏内容に対して転写する事例の選択に用いることを考える。この手法において演奏における表情と楽譜の情報の関係は、楽譜中の音符が示す音高や音価などの局所的な情報と、それに対して実際の演奏に付与された表情がどのような傾向として生じやすいかが確率モデルによって表現されており、それらは関連付けられた楽譜の情報の類似性によって木構造状に体系化されている。これを利用することで、演奏において何らかの類似した傾向を持つ事例がどのような楽譜の情報の下で生じやすいかを分類することができる。また、この木構造では演奏事例が一定の楽譜の情報によって階層的に分類されているため、部分的にのみ一致する楽譜の情報に対しても類似した傾向を持つ事例を特定することができるというメリットを持つ。それは未知の楽譜情報に対しても適切な事例を特定できる可能性があることを意味する。

2.2 楽譜の情報に基づく演奏事例の生成条件の体系化

2.2.1 鍵盤楽器における演奏表現の特徴の獲得

MIDI ストリームからイベント群として観測される実際の演奏系列 $R = \{r_1, \dots, r_N\}$ の生成に際し、演奏者はその表情を付加する対象として楽譜から架空の楽譜系列 $\hat{S} = \{\hat{s}_1, \dots, \hat{s}_N\}$ を想定していると考えられる。 \hat{S} には R の全てのイベントと一対一に対応する演奏指示が存在すると仮定する。両者を対照することで、鍵盤操作を構成する複数の要素についての逸脱量系列 $D = \{d_1, \dots, d_N\}$ を得られる。この逸脱はその演奏に固有な特徴を構成する主要素（芸術的逸脱）であり、演奏者の芸術的な意図の下で楽器の操作に付与されている [9]。図 1 に示すように、鍵盤楽器の演奏は、音符と拍の異なる尺度で変動する逸脱要素群の並走によって構成されていると考えられる。提案手法で逸脱量系列として扱う要素を以下に示す。

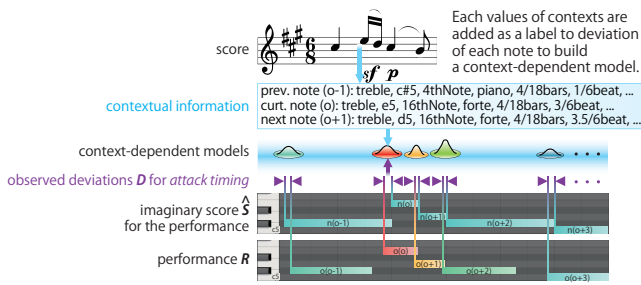


図 2 文脈依存モデルの構築 (音符の場合の例)

Fig. 2 Formation of context-dependent models for each note.

D^{note} 音符単位で観測される逸脱要素

D^{at} (attack timing) 拍の長さを基準とした打鍵時刻の逸脱量。

D^{gr} (gatetime ratio) 楽譜上の音価に対して演奏から観測された打鍵から離鍵までの継続長の比率。

D^{dr} (dynamics ratio) MusicXML 上の dynamics 値と演奏の MIDI ストリームにおける velocity 値の比率。

D^{ps} (pitch shift) 楽譜の指示する音高からの逸脱量を示す整数値。通常は 0。

D^{ks} (key strokes) その事例と楽譜上で同一の音符を基準として演奏された事例の総数。通常は 1。

D^{beat} 拍単位で観測される逸脱要素

D^{br} (local BPM ratio) 楽譜の指示に対して拍単位で変動を算出した局所的な BPM の比率。

ここで任意の事例における逸脱量の発生はその直前の逸脱量に依存するマルコフ性を仮定し、直前の逸脱量との差分を観測した系列 D' も扱う。直前の事例は後述する楽譜の情報から決定し、 $d'_n = d_n - d_{n-1}$ とする。演奏者は何らかの確率に基づいてこれらを生成していると考え、 d_n^{note} 、 d_n^{beat} をそれぞれ多次元正規分布とみなす。

尚、音符における架空の楽譜系列は演奏者によって厳密な楽譜とは異なる内容を持つ系列として想定されている可能性を考慮すると、 \hat{S} の一部は元の楽譜の演奏指示に対して挿入、欠落または置換の関係にあり、厳密には一致しない。 d_n がそれらのいずれかに該当する場合、pitch shift や key strokes が通常以外の値をとる。特に挿入の場合はその生成を具体的に指示する内容は元の楽譜には存在しないが、音高や打鍵位置、ならびに後述する楽譜の情報を参照することで、その生成に寄与した事例を一意に決定できる。

2.2.2 楽譜の情報から演奏事例を説明する生成モデル

提案手法では楽譜から一意に決まる局所的な情報を網羅的に用いて演奏者の意図を間接的に表現することを考え、図 2 に示すような生成モデルを個々の事例 d_n^{note} および d_n^{beat} の確率分布に対して独立して定義する (文脈依存モデル)。音符の場合、楽譜上で対応する多様な情報を獲得できる。文脈要素として使用する具体的な情報は先行研究 [6] に倣い、 d_n に対する d_{n-1} 、 d_{n+1} の決定は声部や和

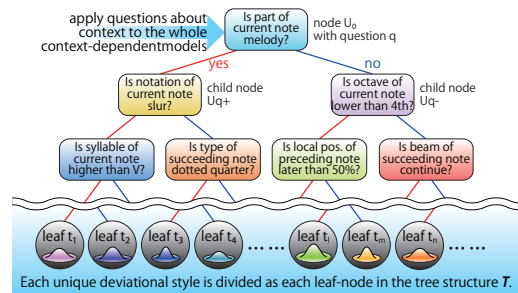


図 3 決定木による文脈依存モデル分類のための規則の体系化

Fig. 3 Systematization of the models with a decision tree.

音の構造を考慮する*1。それらは和声に関する情報であるといえ、主副旋律や伴奏などの複声部からなる系列とみなすことができる。一方、拍の場合は d_n の範囲に相当する楽譜上で、発音などの情報量が常に変動する。それらを律動に関する情報と考え、範囲内での打鍵の密度や、それらの位置を量子化したパターンなどを獲得できる。いずれの場合も d_n の生成にはその周囲を含む局所的な文脈が寄与すると仮定し、 s_{n-1} 、 s_n 、 s_{n+1} から得られる情報の組み合わせを付与する。

2.2.3 演奏事例の生成を制御する規則の体系化

文脈情報の組み合わせは無数に存在するため、その全ての学習を有限の学習データから充分に行うことは困難である。また、あらゆる文脈情報の組み合わせに対応するためには、有限な文脈依存モデルのいずれかを代替として使用できるような共有構造を示す体系化されたルールが必要となる。そのような共有構造は、Tree-based clustering[10]によって全てのモデルに対して文脈に関する網羅的な質問を用いたトップダウンな分類を行うことで構築できる。これによって分類された文脈依存モデルは、図 3 に示すような木構造 (決定木) 状に体系化される。

質問による分割が可能な限り、ほぼ全ての文脈依存モデルを個別にみなすための規則の組み合わせが楽譜の局所的な情報によって一意に定まる。その結果、任意のリーフノードには個別の事例における逸脱量が保持される。決定木の各ノードは分割に用いた質問を持つため、ルートノードからそれぞれの質問に従って木構造を辿ることで、文脈の内容が未知であってもそれに最も類似した文脈情報を持つリーフノードに到達できる。

2.3 最適な演奏表現を付与する事例の選択

既知の演奏表情 D から決定木 T を構築済みであるとして、それ以降の手順を図 4 に示す。生成する演奏内容は楽譜として与えられ、厳密な演奏指示の系列 S と楽譜情報の系列が得られる。 S に相応しい表情を備えた演奏系列 \hat{R} は、音符と拍について最適解となった表情の系列 \hat{D} を S に対して転写することで生成する。

*1 d_n が元の楽譜に具体的な演奏指示のない事例の場合、その生成に寄与した事例のモデルとほぼ全ての文脈情報を共有する。

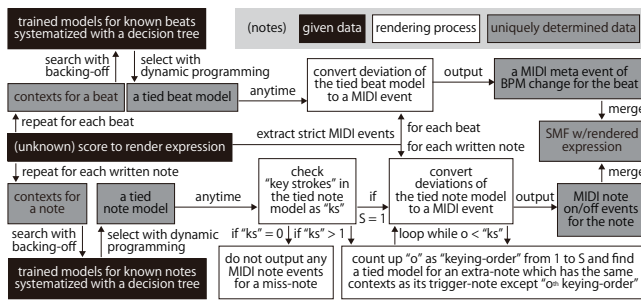


図 4 演奏を生成する楽譜系列に対する表情付けの手順

Fig. 4 Flow to render the performance expression to the score.

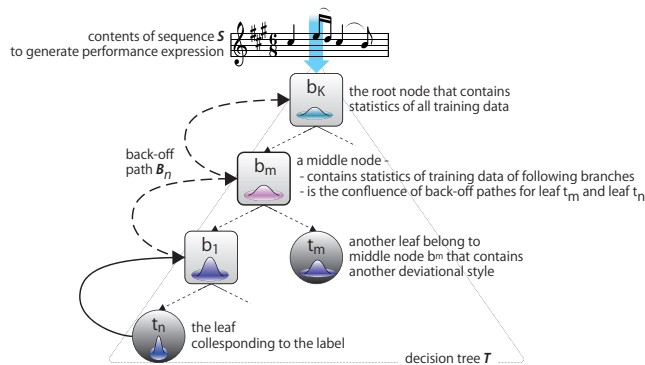


図 5 決定木のバックオフによる事例の探索

Fig. 5 Example of the decision tree back-off.

2.3.1 決定木のバックオフによる事例の探索

T において任意のリーフ t_n が保持するのは D 中の特定の文脈において発生した逸脱量であるのに対し、 T におけるルートノードは D 全体の大まかな傾向を保持している。任意の文脈を持つ演奏指示 s_n に転写するのに最適な逸脱量を持つのは t_n とは限らず、 t_n とルートノード中間にあるいずれかのノード b_m で分岐した先に存在するリーフ t_m である可能性が考えられる。その探索対象は T 全体に及ぶため、 S 全体に対してその選択の最適解を探索するのは現実的に困難である。そこで各リーフの評価指標として、決定木のバックオフ (図 5) のアイデアを用いる [11]。これによってルートノードから t_n に至る木構造を遡り、それ以外のリーフを木構造に従って段階的に選択候補に加えていくことができる。

s_{n-1} に対して t_l を選択した上で s_n に対して t_m を選択し、それらを連結する場合の評価は、連結する各事例と中間ノード b_m の統計量を基準とした確率密度関数によって行う。任意の逸脱要素について $P(t_l, t_m) = \frac{1}{\sqrt{2\pi}\sigma_m} e^{-\frac{(d_m^f - \mu_m^f)^2}{2\sigma_m^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_l} e^{-\frac{\{(d_m^f - d_l^f) - \mu_l^f\}^2}{2\sigma_l^2}}$ であり、 D^{note} の場合 $P(t_l^{note}, t_m^{note}) = P(t_l^{at}, t_m^{at}) \cdot P(t_l^{gr}, t_m^{gr}) \cdot P(t_l^{dr}, t_m^{dr})$ 、 D^{beat} の場合 $P(t_l^{beat}, t_m^{beat}) = P(t_l^{br}, t_m^{br})$ である。

2.3.2 生成する演奏内容に対する表情の転写

演奏指示系列 S 全体に対して最適な演奏表情の系列 \hat{D} は、 S の事例 s_{n-1} , s_n に対して t_l , t_m を連結した場合の

尤度を系列全体に対して最大化するように決定される。その相応しさは局所的な範囲で評価できるため、動的計画法を用いて現実的な計算コストで最適解に到達できる。尚、 S は \hat{S} と異なり実演奏系列が未知であるため、元の楽譜にない演奏指示のような情報は得られない。そのような事例の生成を促す演奏指示がある場合には、選択された事例が持つ逸脱量や、それが文脈情報を共有する事例を用いた補充によって、挿入、欠落、置換などへの対応を行う。

3. 手法の評価

提案手法の評価として、複数の楽曲または演奏者のデータを用いた生成演奏について、聴覚と視覚の両観点から評価及び検証を行った。生成モデルの学習データならびに生成演奏への転写候補としては、ピアノの名演奏の波形を特定の MIDI 音源において聴覚上で模倣したデータベース*2をもとに、市販の楽譜を再現した MusicXML と対照して抽出した音符の逸脱と拍の変動を使用した。

3.1 可聴化による評価

提案手法による表情を付加した演奏について主観評価を行い、その品質を検証した。同一の作曲者による楽曲は楽譜上で比較的類似した特徴を持つと仮定し、以下の3種類のモデルを学習した。

C-A F. Chopin's Etude No. 3, 4, 23, Mazurka No. 5, Nocturne No. 2, 10, Prelude No. 7, 20, Waltz No. 1, 3, 9 and 10, performed by V. Ashkenazy.

M-G W. A. Mozart's Piano Sonata K. 279 all of three movements, and Piano Sonata K. 310 the first movement, performed by G. Gould.

M-P W. A. Mozart's Piano Sonata K. 279 all of three movements, Piano Sonata K. 310 all of three movements, Piano Sonata K. 331 second and third movements, and Piano Sonata K. 545 all of three movements, performed by M. J. Pires.

演奏表情の生成には表 1 に示す楽曲を用いた。いずれもモデルの学習に用いた楽曲とは多くの特徴が異なる。各モデルの学習に用いた楽曲の演奏および無表情の演奏を提示の上、表 2 に示す5段階で生成演奏を採点するものとした。音符と拍の各モデルに使用した楽譜の情報を持つ制約としての性能を個別に調査するため、拍または音符のみに表情を付与した場合も加えた3種類の演奏を評価対象としている。被験者は演奏経験を問わない20人の聴者である。尚、可聴化における再生条件などは提案の本質には含まれないため、本稿では議論の対象としない。

評価結果を図 6 に示す。評価点は半数以上の楽曲で4を超えており、概ね良好な結果が得られたと考えられる。

*2 CrestMusePEDB version 2, <http://www.crestmuse.jp/pedb>

表 1 主観評価に使用した楽曲群

Table 1 Compositions used in subjective evaluations.

Abbreviation	Composer	Title	Notes	Beats
bac-wtc123-p	J. S. Bach	Well-Tempered Clavier, BWV868 Prelude	418	72
bar-rou001	B. Bartók	Roumanian Folk Dances, Movement I, Stick Dance	373	101
deb-pld110	C. Debussy	Prélude Book 1, No. 10	987	270
pro-snt007-1	S. Prokofiev	Piano Sonata, No. 7	1247	454
rac-pld003-2	S. Rachmaninoff	Prelude in C-sharp minor, Op. 3, No. 2	1805	250
rav-sonatn-1	M. Ravel	Sonatine, Movement I, "Modéré"	1851	216
sat-gym001	E. Satie	Gymnopédies, No. 1	227	114

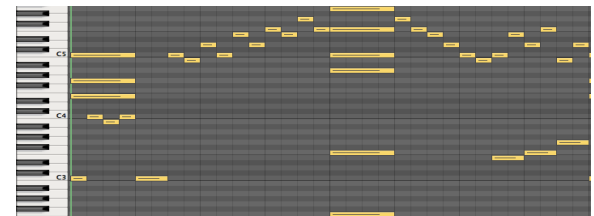
表 2 各主観評価の採点尺度

Table 2 Criteria for scoring in subjective evaluation.

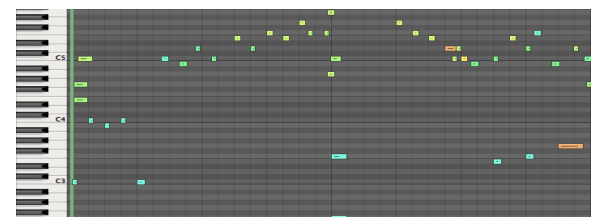
Subjective score	Requirement
5	very musical
4	generally musical
3	human-like
2	poor
1	random



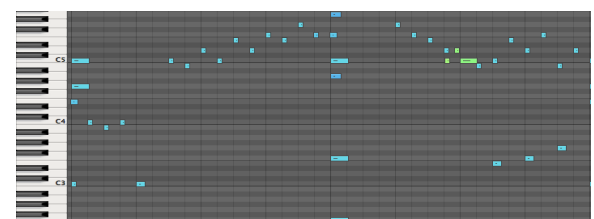
(a) Partial score to generate performances.



(b) MIDI stream of strictly converted score.



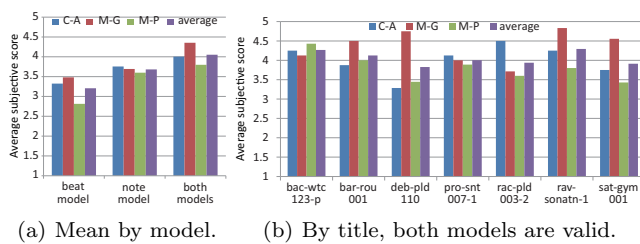
(c) Generated performance by M-G model.



(d) Generated performance by M-P model.

図 7 同一楽曲の演奏における音符の逸脱の振舞い

Fig. 7 Behavior of deviations for note.



(a) Mean by model.

(b) By title, both models are valid.

図 6 主観評価結果

Fig. 6 Scores of subjective evaluations.

M-P に比べて学習データの少ない M-G でも総合的な評価は M-P を上回っており、今回設計したモデルとしての性能の獲得に必要なデータは M-G 程度でも不足ではなかったといえる。演奏者モデルごとの評価を比較すると、M-P が総合的に低い評価となっていることが目立つ。音符の表情のみを付与した場合の結果は他の 2 モデルと近いのに対し、そこに拍の表情も追加した場合の評価値の向上幅に他の 2 モデルとの差異が大きいため、拍の表情のみを付与した場合の評価が特に低いことが影響しているように見える。両群において演奏の人間性や音楽性への貢献度合が同等であるとは限らないが、今回拍のモデルに使用した楽譜情報の組み合わせでは特に M-P の演奏表情を捉えることが難しく、改善の余地がより大きいと考えるべきであろう。

3.2 可視化による評価

次に、生成された演奏について可視化による観測を行い、演奏者が持つ特徴的な表情を逸脱量として転写できているかを検証する。先の評価に用いた M-G および M-P で学習データとして共有されていた楽曲である W. A. Mozart's Piano Sonata No. 1, K. 279, The first movement の演奏

を生成した結果を 2 名の演奏者間で比較する。 D_{note} の各要素については、生成した MIDI ストリームのピアノロールを図 7 に楽譜と併せて示す。 D_{beat} については、表情として得られた *local BPM ratio* の変動の軌跡を図 8 に示す。

図 7 から、転写された逸脱量は大きかには類似しており、楽曲に制御されているように見えるが、音符の持つ文脈を考慮して比較すると演奏者による相違がある。譜例に従うと、 *trill* の装飾記号により記譜外の発音が想定される箇所では音高、発音順ともいずれも一般的に挿入される音符の組み合わせが再現されているほか、 *arpeggiate* の装飾記

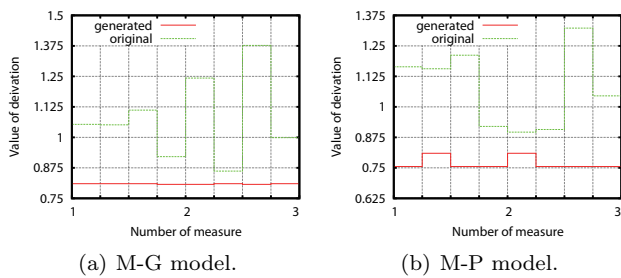


図 8 同一楽曲の演奏における *local BPM ratio* の軌跡
Fig. 8 Trajectory of deviations for beat in *local BPM ratio*.

号を持つ和音の構成音では *attack timing* に時間差があり、発音が低音側から行われることも演奏表現として自然である。また、特に M-G において大譜表の上下段によって明確に異なる傾向で *dynamics ratio* の制御が行われている。これらから、音符について関連付けた楽譜の情報は比較的良好にその特徴を説明できているといえる。しかし付加される表情は全般的に画一的な値に留まる傾向があり、生成結果において実際に選択されたリーフの種類は多くなかったと考えられる。特に M-P の生成演奏は比較的単調だが、先の評価における結果からはそれが演奏者固有の特徴であるとも断言できない。この傾向は *local BPM* の変動ではより顕著で、選択されたリーフが非常に限定的であったことがわかる。転写された逸脱量はいずれも実際に演奏された事例のいずれかのものであるが、結果的に生成された演奏の表情は統計的に平滑化されたと言えるようなものに近くなっている。モデル化の際に逸脱量に関連付けた楽譜の情報はその拍内の律動に関するものであるが、実際の演奏における任意の逸脱量を捕捉し再度呼び出すには、それらの組み合わせでは不十分であったといえる。また、生成する演奏内容に対して表情として転写する事例は、自己と直前に適用する事例が決定木の構造に従って持つ尤度に基づいて選択される。多様な変動を取るような逸脱量を持つリーフの選択を困難にした理由として、事例が持つ文脈情報と同様に、それも局所的な情報に対して強く依存していることが考えられる。

4. おわりに

本稿では、鍵盤楽器演奏の自動演奏表情付の問題を解くため、芸術的逸脱の生成モデルにおける決定木のバックオフによって事例選択を行う手法を提案した。検証の結果、生成された演奏における演奏表情は、学習データとなった演奏が持つ特徴的な傾向を捉え、再現しうる性能を持つことがわかった。提案手法によって生成される演奏の構築に貢献するのは、元となった演奏の演奏者自身が持つ知識を代替する情報である。そこに各種楽曲構造解析手法やヒューリスティックなパラメタ調性などの演奏者以外の人的要因が関与しないことが、演奏者が演奏を生成する過程の自動化において重要である。

提案手法では音符と拍の概念を独立に捉え、各々に異なる内容の情報を楽譜から抽出して逸脱量に関連付けている。実施した各評価の結果からは、今回使用した楽譜の情報では、特に拍の変動に関連した説明を行うためには改善の余地が大きいといえる。その説明的的確さは、文脈要素としてどのような情報を使用するかに依存する。しかしここで文脈要素といえるものは前提次第で楽譜以外からでも際限なく定義できるため、究極的な解は存在しない。何を制約に演奏表現を説明するかはこの問題を解くための手法において本質的な課題であり、本手法においても今後最も具体的に検討すべき事柄の一つである。また、提案手法の生成演奏が演奏者の個性を保持したものであるかを検証することも今後の課題として重要なものである。音楽的により充実した見地からの検証を行うため、生成課題として更に多様な演奏者や楽曲に対応する必要があると考えている。

参考文献

- [1] Kirke, A. and Miranda, E. R.: *Guide to Computing for Expressive Music Performance*, SpringerLink : Bücher, Springer London (2013).
- [2] Grindlay, G. and Helmbold, D.: Modeling, Analyzing and Synthesizing Expressive Piano Performance with Graphical Models, *Machine Learning Journal*, Vol. 65, No. 2-3, pp. 361–387 (2006).
- [3] Flossman, S., Grachten, M. and Widmer, G.: Expressive Performance Rendering: Introducing Performance Context, *Proc. Sound and Music Computing (SMC) Conference*, pp. 155–160 (2009).
- [4] Kim, T., Fukayama, S., Nishimoto, T. and Sagayama, S.: Performance Rendering for Polyphonic Piano Music with a Combination of Probabilistic Models for Melody and Harmony, *In Proceedings of Sound and Music Computing (SMC) Conference* (2010).
- [5] Widmer, G.: Machine Discoveries: A Few Simple, Robust Local Expression Principles, *Journal of New Music Research*, Vol. 31, pp. 37–50 (2002).
- [6] Okumura, K., Sako, S. and Kitamura, T.: Stochastic Modeling of a Musical Performance with Expressive Representations from the Musical Score, *In Proceedings of the 12th International Society for Music Information Retrieval (ISMIR) conference*, pp. 531–536 (2011).
- [7] Widmer, G. and Tobudic, A.: Playing Mozart by Analogy: Learning Multi-Level Timing and Dynamics Strategies, *Journal of New Music Research*, Vol. 32, pp. 203–216 (2003).
- [8] Tobudic, A. and Widmer, G.: Learning to Play Like the Great Pianists, *Proc. the International Joint Conference on Artificial Intelligence*, pp. 871–876 (2005).
- [9] Seashore, C. E.: *Psychology of Music*, Dover books on music, music history, Courier Dover Publications (1938).
- [10] Odell, J. J.: The Use of Context in Large Vocabulary Speech Recognition, PhD Thesis, Cambridge University (1995).
- [11] Kataoka, S., Mizutani, N., Tokuda, K. and Kitamura, T.: Decision-tree backing-off in HMM-based speech synthesis, *In Proceedings of the 8th International Conference on Spoken Language Processing (ICSLP/INTERSPEECH 2004)*, Vol. 2, Jeju Island, Korea, pp. 1205–1208 (2004).