

フレーズ構造の自動解析に基づく テンポ変動曲線の生成によるピアノ曲の演奏表情付け*

☆伊東直哉, 深山覚, 齋藤大輔, 嵯峨山茂樹 (東大・情報理工)

1 はじめに

本稿では楽譜から表情豊かな自動演奏を生成できる演奏表情付けについて議論する. この技術は, 新作楽曲の演奏や, 演奏の著作権問題回避などに有用である. 自動演奏は, 楽譜に表記される音高と音長に従って音を鳴らすことによって, ひとまずは実現できるが, 人間が演奏しているようには聴こえない. 人間の演奏には楽譜に直接は書かれない“演奏表情”が含まれるからである. J. Sundberg [1] 以来, 演奏表情付けには様々な方法が提案されてきた. 次章で議論する音楽の階層性は GTTM [2] などの方法で記述された. 本稿では階層ごとの演奏表情を扱うことによる表情豊かな演奏生成の可能性について議論する.

2 演奏表情の階層性

演奏表情は楽譜での着目する時間単位の長さに応じて階層構造をなすとされる (Fig. 1) [3]. 楽曲全体, セクション, フレーズ, 個々の音符などの各階層ごとに, それぞれ異なる要因による演奏表情がある (Table 1). 広い時間幅に着目した上層の演奏表情には, 慣習や人間の意図や個性が反映され, 短い時間幅に着目した下層の演奏表情には, 楽器や演奏者の身体による物理的制約が反映される. 自動演奏システムにおいては, 楽曲スタイルなどの上層の演奏表情ほどユーザの意向を反映すべきであり, 完全自動化するより寧ろユーザが設定・選択する方が適している. 一方, 演奏者によって無意識に付与される下層の演奏表情ほど, 一音一音をユーザが人手によって表情付けすることは難しく, 工学的に最も自動化が有用である. 個々の音符や演奏記号に着目した下層の演奏表情は, [4] などにおいて研究が行われてきた.

本稿では, さらにひとつ上層のフレーズの階層をも着目した演奏表情の自動生成について議論する. フレーズの階層に着目した演奏表情は, テンポ変動などによって演奏として表現される [5]. これらの演奏表情を重ね合わせることで, より表情豊かな演奏の自動生成ができると考えられる.

3 フレーズ構造の自動解析

3.1 類似部分の検出による動機・フレーズ構造の解析

音楽理論では, まとまりのある音符列を動機 (モチーフ) と呼ぶ. 楽曲が作曲される際にはまず動機を用意し, それを変形して用いることで楽曲が構成されることが多いため, 一楽曲中には多くの動機とそれらが変形したものが存在することが多い. 同一の動機やその変形が使われた部分には何らかの類似性がある. Fig. 2 に例を示す. その類似部分を検出することによって動機とフレーズ構造を解析できる可能性がある.

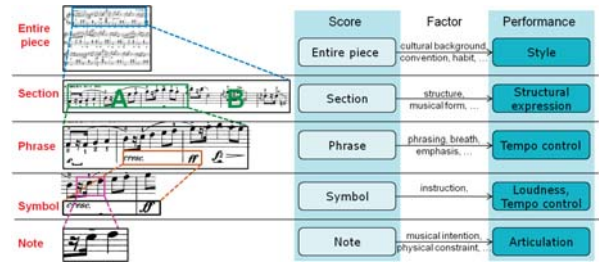


Fig. 1 音楽演奏の階層モデル



Fig. 2 同じ動機に基づく類似部分の例

Table 1 音楽演奏での着目箇所, 要因, 表情

楽譜での着目箇所	要因	演奏表情
楽譜全体	慣習, ...	スタイル
セクション	楽曲形式, ...	階層的表現
フレーズ	演奏者の呼吸, ...	テンポ変動
演奏記号	演奏指示, ...	音量変動, ...
音符	物理的制約, ...	アーティキュレーション

3.2 累積距離最小経路探索に基づく動機・フレーズ構造の解析法

本節では, 類似部分の検出に基づくフレーズ構造の解析法について議論する. 楽譜に対応する MIDI データの時刻 s と時刻 t の間の距離が定義でき $D(s, t)$ で表されるとする. 具体的には, 今回は時刻 t において MIDI ノートナンバー全ての音高について, 鳴っているだけ 1, そうでなければ 0 に値を持つ 127 次元の特徴ベクトルを \mathbf{x}_t とし, \mathbf{x}_s と \mathbf{x}_t の距離 $d(s; t)$ をハミング距離とした.

繰り返し現れる動機及びその類似部分は, 距離行列の中で斜めの線として現れる. それらを検出する問題は, 行列要素のたどり方を経路とみなし, 行列の上端と下端を経路の始端と終端に対応させれば, 経路探索問題に帰着できる. このとき, 行列の対角線は経路として意味をなさないため, 対角線近辺は経路探索が禁止される. また, 動機及び類似部分の順序は一定せず経路は連続的ではないので, 遠方へのジャンプを許さねばならない. さらに, 同一動機が異なる音高に移されて繰り返し使われることが多いので, 特徴ベクトル系列と, 音高を n 半音シフト (n は N 種類) した特徴ベクトル系列の間の距離行列 12 種類を使えば, 12 枚の距離行列間の遷移を許す累積距離最小経路探索問題として, 音高が異なる同一動機や類似部分が検出できる. 今回は, シフトの種類数を $N = 12$ とし, このようにして検出された動機を 1 小節以上 3 小節以下の長さにとまとめたものをフレーズとして解析した.

* Performance rendering on piano music by generating tempo curve with automatic phrase analysis by ITO Naoya, FUKAYAMA Satoru, SAITO Daisuke, and SAGAYAMA Shigeki (The University of Tokyo)

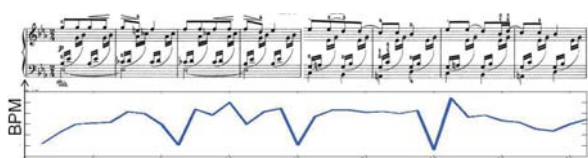


Fig. 3 実演奏のテンポの変動 (E.Grieg's Arietta Lyric Piece Op.12-1, Measure 1-8)

Table 2 主観評価に用いたピアノ曲

作曲者名	曲名	略記
E.Grieg	Arietta Lyric Piece Op.12-1	gri
F.Chopin	Etude No.3	cho
L.v.Beethoven	Sonata No.8 (2nd movement)	bee
B.Bartok	Romanian Folk Dances No.1	bar

4 フレーズ構造に基づくテンポ変動の生成

フレーズ構造に着目した演奏表情はテンポ変動に顕著に現れる. 実際, 音楽演奏理論において「ピアニストはフレーズの終わりなどでテンポを落とすことによって階層構造を表現している」などの知見 [6] があり, 実際の演奏の観察 (Fig. 3) においても確認できる. 我々は, フレーズ単位を加減速を表現する関数として, 加速と減速の程度を, 深さや鋭さを表す4つのパラメータ (a_1, b_1, a_2, b_2) によって表現し (式 (1), 式 (2)), 更に加速と減速が切り替わる中心時刻もパラメータ t_c として扱い, t_b, t_e はフレーズの始端時刻, 終端時刻として, テンポ変動曲線を,

$$f(t|t_b \leq t \leq t_c) = -a_1 \left(\frac{t-t_c}{t_b-t_c} \right)^{b_1} + 1 + a_1 \quad (1)$$

$$f(t|t_c \leq t \leq t_e) = -a_2 \left(\frac{t-t_c}{t_e-t_c} \right)^{b_2} + 1 + a_1 \quad (2)$$

と表した. これらのパラメータを変化させることによって, 多様なテンポ変動が生成できる. これにより生成されるテンポ変動と, 既存手法 [4] によって個々の音符と演奏記号の階層に着目して生成されるテンポ変動を加算することで, フレーズ構造をも着目した演奏表情を生成することができる.

5 表情付けに対する主観評価

5.1 実験条件

提案法によってより表情豊かな演奏を生成できるかを検証する目的で, 主観評価実験を行った. 4つの楽曲 (Table 2) について, 次の4種類の演奏表情に対し被験者に10段階の評価値を付けさせた: (1) 演奏表情なし (deadpan), (2) 既存手法 [4] によって個々の音符と演奏記号の階層にのみ着目して生成される演奏表情 (conventional), (3) 提案法によってフレーズ構造解析結果に基づいて生成されるテンポ変動を従来法によるテンポ変動に加算したもの (proposed), (4) 人間の演奏家による演奏 (human). 人間が演奏しているように感じさせる面と音楽として良いと感じさせる面があると考えられるので, 評価指標として, 人間が演奏しているように感じるか (Human-likeness Score; HS) と良い音楽であると感じるか (Musicality Score; MS) の2項目を用いて10段階評価を行った. テンポ変動曲線のパラメータセットは6組用意した. 10人の被験者に対して実験を行った.

5.2 実験の結果と考察

2つの評価指標 (HS, MS) の楽曲のについての評価値の平均を Fig. 4 に示す. proposed の評価値は, 6組

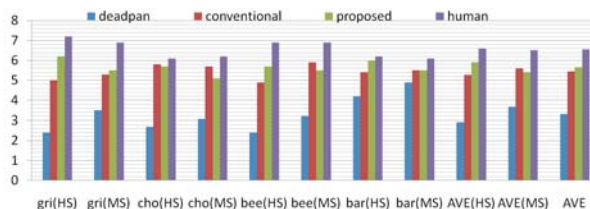


Fig. 4 4種類の演奏表情付けに対する評価値の平均

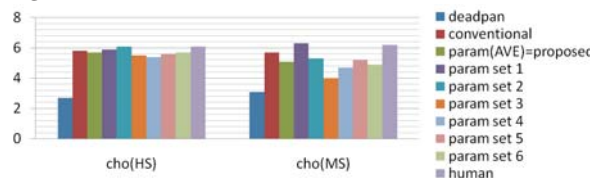


Fig. 5 ある楽曲 (cho) のパラメータ設定ごとの評価値

のテンポ変動曲線のパラメータセットに対して生成した演奏の評価値の平均である. さらに, HS, MS についての評価の全ての楽曲についての平均 (AVE(HS), AVE(MS)) と, AVE(HS) と AVE(MS) の平均 (AVE) をも併せて示す. AVEを見ると, proposed の評価値は human の評価値よりは低いが, conventional の評価値よりも高い値であった. これは多くの場合において提案法によって従来法より表情豊かな演奏が生成されたことを示している.

2つの評価指標についての評価を比較した. AVE(HS, MS) を見ると, 提案法の評価値は, 従来法の評価値に比べて HS は高い傾向があり MS は低い傾向があった. 提案法によって人間が演奏しているように感じさせる傾向が増加し, 音楽として良いと感じさせる傾向が減少したことを示している. これは, パラメータセットによっては音楽として良いと感じさせないテンポ変動の生成が行われたことが原因と考えられる.

ある楽曲 (cho) のパラメータセットごとの評価値を Fig. 5 に示す. 実験に用いたすべての曲について, HS と MS の両評価指標でもともに従来法より評価値が高いパラメータセットがあった. 曲に依って, 高い評価値が得られるパラメータセットは異なっていた. 曲ごとに適切なパラメータセットを自動生成する方法は今後の課題である.

6 結論

本稿では, 音楽演奏の階層性に基づき, 個々の音符と演奏記号に対する演奏表情と, フレーズ構造に基づくテンポ変動を共に扱える演奏表情付け法について議論した. フレーズ構造の解析は累積距離最小経路の探索問題として定式化でき, テンポ変動は, 加速と減速がパラメータライズされた曲線によって生成できた. 主観評価実験により, 従来法に比べより表情豊かな演奏が生成できることを確認した.

参考文献

- [1] J. Sundberg et. al., Computer Music Journal, vol. 7, no. 1, pp. 37-43, 1983.
- [2] F. Lerdahl et. al., The MIT Press, 1983.
- [3] H. Keller, Kassel und Basel, 1955.
- [4] T. H. Kim et. al., in Proc. NIME, 2011, pp. 96-99.
- [5] R. Aiello et. al., Oxford University Press, 1994, pp. 152-169.
- [6] N. P. Todd, Music Perception, vol. 3, no. 1, pp. 33-58, 1985.