

# 楽曲構造に基づくテンポ包絡曲線の生成による 自動演奏表情付けと楽曲構造解析法の検討

伊東 直哉<sup>1</sup> 深山 寛<sup>1</sup> 中村 栄太<sup>1</sup> 齋藤 大輔<sup>1</sup> 嗟峨山 茂樹<sup>1</sup>

**概要**：楽曲構造に基づくテンポ包絡の生成によって人間が演奏しているような演奏を生成する演奏表情付け法を提案するとともに、その楽曲構造の自動解析を楽譜からの類似箇所検出問題として定式化した。本報告では、音楽学で言及されている演奏表情の階層性に基づいてフレーズ構造に基づく演奏表情に着目した。フレーズ構造に基づく演奏表情付けはテンポ包絡の生成問題に帰着される。テンポ包絡は各階層のテンポ包絡を重ね合わせてモデル化を行った。モデルと実演奏データをフィッティングしてモデルの妥当性を検討した。主観評価実験の結果、フレーズ構造に基づく演奏表情と下層の演奏表情を組み合わせることにより人間が演奏しているような演奏を生成でき、モデルの有効性が示された。また、演奏表情の階層性に基づき、演奏表情付けシステムの利用場面におけるユーザの介在について議論する。さらに演奏表情付けの自動化のためにフレーズ構造解析を扱うことによって、解析手法の自動演奏表情付けに対する有効性が示唆された。

**キーワード**：楽曲構造、演奏表情付け、テンポ変動、多項式フィッティング、構造解析、類似部分検出

## 1. はじめに

本報告では楽譜から人間が演奏しているような演奏を生成する演奏表情付け技術について議論する。この技術は、新作楽曲の演奏や、演奏の著作権問題回避、人間の演奏法の解明に有用である。演奏の自動生成は、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、その演奏は人間が演奏しているようには聴こえない。なぜなら、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。演奏表情は、意識的に付与される変動も、無意識的に付与される変動もある。これらの変動を計算機処理によって付与すれば、人間が演奏しているような演奏が生成可能と考えられる。

自動演奏表情付けの手法は、J. Sundberg [1] 以来様々な手法が提案されてきた。それらはルールベースの手法、事例ベースの手法、機械学習による手法に分けられる。ルールベースの手法 [2] では、どのような場合にどのように演奏すべきかというルールによって演奏生成を行う。事例ベースの手法 [3][4] では、どのような場合にどのように演奏されたかを検索し転写することで演奏生成を行う。専門的な音楽知識や大量のパラメータ探索が必要というこれらの手法の問題点を解決するために、機械学習によるモデルが多数提案された [5][6]。

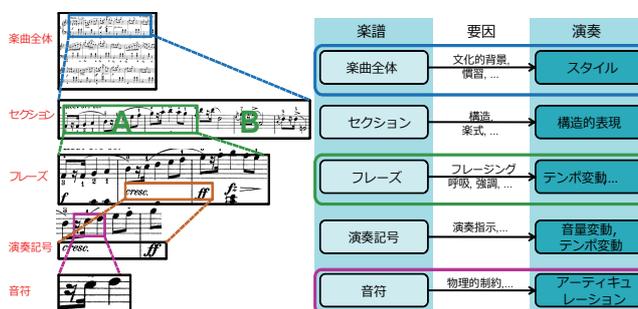


図 1 演奏表情の階層性。

さらに、インタラクティブにユーザの指示を受け付けるシステム [7] も提案された。ここでは、生成された演奏を部分的に精練し改良できる。

演奏表情を扱う方針として、音楽学で言及されている演奏表情の階層性が着目されている。音楽の階層性は GTTM [8] などの方法でも記述されている。ここで着目するのは、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点である [9](図 1)。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。

自動演奏システムにおいては、下層の演奏表情の自動化が最も有用である。楽曲スタイルなどの上層の演奏表情ほど文化的慣習や人間の意図や個性が反映されるので、システムが自動設定するのは困難である。一方、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煩雑であり、演奏者にとって無意識の楽器や身体による

<sup>1</sup> 東京大学  
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo  
113-0033, Japan

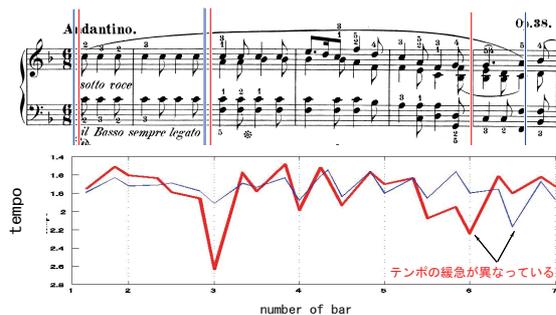


図 2 実演奏テンポ変動とフレーズ構造. (F.Chopin, Ballade No.2 Op.38, bars 1-6)

物理的制約が反映されるので大量のデータに共通している部分が多い。

本報告では、演奏表情の階層性に基づき、上層にユーザが介在した演奏表情付け法について述べる。

## 2. 階層的なフレーズ構造に基づく演奏表情付け

機械での自動化に有用な下層の演奏表情から順に扱い、より人間が演奏しているような演奏の生成を目指す。個々の音符や演奏記号に着目した下層の演奏表情は、[10]などによって研究が行われてきた。本報告ではさらにひとつ上層のフレーズに着目する演奏表情付けについて論じる。さらに複数のフレーズによって構成される大フレーズにも着目し、より長い時間単位の演奏表情を組み合わせることで、より人間が演奏しているような演奏が生成できるかを検証する。さらに、ユーザの介在が容易な部分について、人間と機械がいかに協調するかを検討する。

フレーズ構造に基づく演奏表情付け問題は、(1) フレーズ構造を楽譜からいかに解析するかという問題と、(2) フレーズ構造が与えられた場合にいかに演奏に反映させるかという問題に切り分けられる。(1)については、フレーズ構造解釈が演奏者によって異なる[11]ことなどから、人間が担当する方針も考えられる。(2)について、フレーズに基づく演奏表情は、呼吸の制約により大量のデータに共通している部分が多いと考えると、自動化には有用である。

本報告では主に(2)の問題について議論し、(1)の問題については限定的な場合において解析可能であることを示す。

### 2.1 テンポ変動によるフレーズの反映

人間はフレーズ構造をいかに演奏に反映しているかについて考える。音楽においてフレーズ構造とは、複数の音符から成るメロディのひとまとまりをあらわす単位のひとつである。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[12]。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[13]」などの知見がある。また、フレーズ構造解釈は演奏者によって変わりうる。図2は実演奏のテンポ変動の観察である。赤線と青線によって同一演奏者が別々のフレーズ構造境界で演奏したときのテンポ

変動が表されている。ここでテンポ変動によるフレーズ構造の表現が確認できる。フレーズ構造境界の違う演奏では異なるテンポの緩急が観察され、それを聴取する際に異なる印象を受けることを確認した。知覚に関して「人間は音量変動やテンポ変動によって表現されるフレーズ構造を知覚している[14]」とも言われている。よって、フレーズ構造ごとの演奏表情付けは、フレーズ構造ごとのテンポ変動を生成する問題に帰着される。音量変動についても重要であるがより顕著と言われるテンポ変動に今回は着目した。

フレーズ毎のテンポ変動がいかに生成されるかについて考える。保科らの演奏理論において、「フレーズでは、頂点に向かってテンポと音量が増加し、逆に頂点を過ぎればそれらは減少する[15]」とある。この知見に基づき、フレーズ構造ごとのテンポ変動生成は、頂点位置と加減速の形状をパラメータとした関数によって生成できると考えられる。

### 2.2 複数のフレーズによって構成される大フレーズ

演奏表情の階層性に基づく、より長い時間単位を捉えた演奏表情を扱うことでより人間が演奏しているような演奏表情が生成可能と考えられる。フレーズに比べてより長い時間単位である大フレーズは、複数のフレーズによって構成される[16]。ここで大フレーズに対してより短いフレーズを小フレーズと呼ぶ。テンポの緩急は小フレーズ毎に一定ではなく、他の小フレーズに比べてより顕著なテンポの緩急が付与される場合がある。

階層的なフレーズ構造に基づくテンポ変動は各階層のフレーズの影響が重ね合わされていると考えることで、より長い大フレーズを捉えた演奏表情をモデル化できる。

## 3. 階層的なフレーズ構造に基づくテンポ変動モデル

### 3.1 テンポ変動の重ね合わせによるモデル化

ある音符の楽譜時間(音価)を $s$ 、その音符が演奏された実時間を $\tau$ とする。楽譜時間と実時間の比であるテンポを $r_i(s)$ とする。 $i$ はフレーズのインデックスとする。

テンポ変動 $\tilde{r}_i(s)$ を、テンポを基準テンポ $\bar{r}_i$ で正規化した

$$\tilde{r}_i(s) = \frac{1}{\bar{r}_i} r_i(s) \quad (1)$$

と定義する。フレーズ毎のテンポ変動は、 $b_i, c_i$ をパラメータとする2次関数

$$\tilde{r}_i(s) = A_i + b_i(s - c_i)^2 \quad (2)$$

とすることで、頂点位置 $c_i$ と増減の形状 $b_i$ をパラメータに持つ関数としてモデル化できる。なお、 $A_i$ はテンポの正規化の制約によって決まる。フレーズ構造を各フレーズ構造境界の時刻 $t_i$ によって $\phi = (t_1, \dots, t_M)$ と定義する。フレーズの区間 $t_i \leq s \leq t_{i+1}$ の基準テンポはフレーズの始端から終端までの実演奏時間と楽譜時間の比

$$\bar{r}_i = \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} ds r_i(s) \quad (3)$$

である。

階層の深さを  $N_\phi$  , 階層  $p$  におけるフレーズ構造を  $\phi_p = (t_{1p}, \dots, t_{M_p})$  とする. 階層的なフレーズ構造に基づくテンポ変動  $w(s)$  は

$$w(s) = W \prod_{p=1}^{N_\phi} \tilde{r}_{i_p(i_{N_\phi})}(s) \quad (t_{i_{N_\phi}} \leq s \leq t_{i_{N_\phi}+1}) \quad (4)$$

とモデル化できる. テンポ変動をどう重ね合わせるかは自明ではないが, 最も単純に表現するために積によって記述した. 例えばいずれも 1 のテンポ変動を重ね合わせたテンポ変動は 1 であることから, 積の妥当性は指摘される. 上層のフレーズ構造境界は下層のフレーズ構造境界を含むので  $\phi_1 \supseteq \dots \supseteq \phi_{M_{N_\phi}}$  であり,  $w(s)$  は, 最も下層のフレーズのインデックス  $i_{N_\phi}$  ごとに場合分けされる.  $p \leq q$  とすると,  $i_p(i_q)$  を階層  $q$  でのインデックス  $i_q$  のフレーズを含む階層  $p$  でのフレーズのインデックス, と定義する.  $W$  を基準テンポとする.

なお, このモデルにおいてどの階層に対して最も正確なモデル記述がなされるかについては任意性がある.

### 3.2 テンポ変動モデルによる生成法

未知の楽曲に対していかにテンポ変動を自動生成には, 小フレーズと大フレーズに関するパラメータの平均値を利用すれば実現できる. また, 得られた分布を確率分布とみなせば, その確率分布に従ってパラメータを選択することでも実現できる. また, フレーズの連続性に特徴があると仮定すれば, 「あるパラメータが選ばれると, 次にこのようなパラメータが選ばれやすい」という情報を使って実現できる.

パラメータの分布をいかにして得るのかについて述べる. テンポ変動モデルに従って観測系列  $Y = \{(s_0, r_0), (s_1, r_1), \dots, (s_n, r_n)\}$  が生成されたと考えると, 観測系列が与えられたときのモデルパラメータの推定は, 逆問題である. これは, 観測系列  $Y$  に対しテンポ変動モデル  $w(s; \Theta)$  との誤差を最小にするようなパラメータ  $\Theta$  を推定する問題として定式化できる. また, 大フレーズの影響を差し引いた小フレーズでは同じテンポ変動が生成されるとするならば, より制約の強いパラメータ推定が可能である.

## 4. 予備実験による階層的なモデルの妥当性の検証

### 4.1 実験条件

実際の楽曲のテンポ変動が提案モデルのパラメータとして具体的にどう表せるかを調べるために, 式 (4) に式 (2) を代入したテンポ変動モデルと実演奏データのフィッティングを行った. フィッティングは最小二乗法によって行った. 学習に必要な実演奏データは, 演奏表情データベース

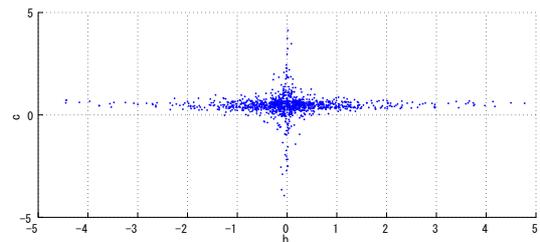


図 3  $N_\phi = 1$  の場合, 実演奏とモデルのフィッティングによって得たパラメータの分布.

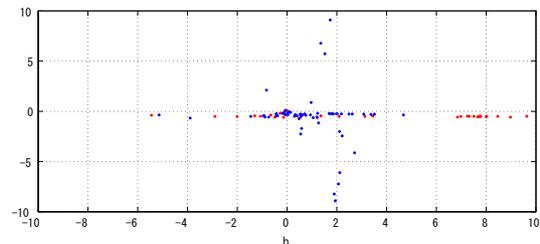


図 4  $N_\phi = 2$  の場合, 実演奏とモデルのフィッティングによって得たパラメータの分布.

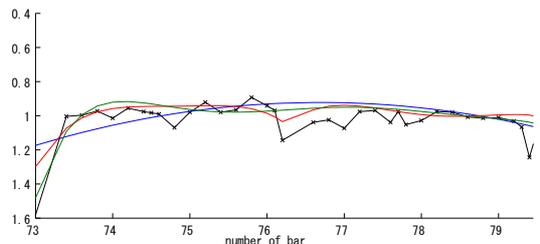


図 5 実演奏と 2 階層モデルのフィッティング例. (W.A.Mozart, Sonata K.331, 3rd movement, bars 73-80)

CrestMusePEDBv2.1, v2.2, v2.3, v2.4.1, v2.5, v2.6 を利用した. 階層的なフレーズ構造のラベルは, CrestMusePEDB-STRv1.0 を利用した.  $N_\phi = 1$  の場合はフレーズ構造のラベルが記述されている 91 演奏を選択し,  $N_\phi = 2$  の場合は 2 階層のフレーズ構造のラベルが記述されている 23 演奏を選択した. その中で, 大フレーズの個数は 33, 小フレーズの個数は 66 個であった.

$N_\phi = 1$  の場合, 式 (4) の積算の中身は, 式 (2) に一致する. 基準テンポ  $W$  も含めると, フィットするモデルパラメータは 3 つである.

$N_\phi = 2$ , 小フレーズの個数は大フレーズ毎に 2 つである場合, フィットするモデルパラメータは 7 つである. そのパラメータは, 階層 1 における 1 つ目のフレーズのパラメータ  $b_{11}, c_{11}$ , 階層 2 における 1 つ目のフレーズのパラメータ  $b_{12}, c_{12}$ , 階層 2 における 2 つ目のフレーズのパラメータ  $b_{22}, c_{22}$  と, 基準テンポ  $W$  である.

### 4.2 実験結果とモデルの妥当性の検証

$N_\phi = 1$  の場合の得られたパラメータ  $b, c$  の分布を図 3 に示す. 正規化された楽譜時間上の頂点位置を表すパラメータ  $c$  は, 0 から 1 の区間で密な分布となっており, 多くのフレーズに頂点が存在することを示している.

表 1 主観評価に用いたピアノ曲

作曲者名	曲名	略記	$N_l$	$N_g$
F.Chopin	Nocturne No.2	nct	9	7
F.Schubert	Moment Musical No.3	scb	15	10

表 2 各ピアノ曲に対して用意した演奏表情

	w/ conv	w/o hier	w/ hier	human
deadpan	×	×	×	×
w/o conv,w/o hier	×	○	×	×
w/o conv,w/ hier	×	○	○	×
conventional	○	×	×	×
w/ conv,w/o hier	○	○	×	×
w/ conv,w/ hier	○	○	○	×
human	×	×	×	○

$N_\phi = 2$  の場合、得られたパラメータの分布を図 4 に示す。赤で表現したのが大フレーズに基づくテンポ変動のパラメータ ( $b_{11}, c_{11}$ )、青で表現したのが小フレーズに基づくテンポ変動のパラメータ ( $b_{12}, c_{12}$ ), ( $b_{22}, c_{22}$ ) である。フィッティングの例を図 5 に示した。単純にパラメータを増やした場合 (緑線, パラメータ数が同じ高次多項式を用いたフィット曲線) と同程度の近似精度が得られることが分かった。

一般に、パラメータを増やすとフィッティング精度は高くなると言えるが、どのように増やすかは自明ではない。階層的なモデルに従ってパラメータを増やしたことで、単純にパラメータを増やした場合と同程度のフィッティング性能が得られた。この階層的なモデル化は人間が行う演奏表情付けに近い形でパラメータライズされている可能性があり、ユーザが介入しやすいモデルになっている点において妥当性が指摘される。一方で、フィッティング性能の向上がより人間らしい演奏生成に繋がるかは自明ではない。これについては、6 節で検証する。

## 5. フレーズ構造に基づく演奏表情付けに対する主観評価

### 5.1 実験条件

階層的なフレーズ構造に基づくテンポ変動モデルによってより人間が演奏しているよう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。評価に用いる楽曲は、演奏表情データベース CrestMusePEDB から拍子・テンポ・作曲家が異なり、階層的なフレーズ構造のラベルが付与された 2 つの楽曲 (表 1) を選択した。小フレーズの個数 ( $N_l$ ) 及び大フレーズの個数 ( $N_g$ ) は表 1 の通りである。テンポ変動のパラメータは、実演奏とラベルを用いてフィッティングした値を用いた。それぞれの楽曲に対して階層性の考慮の有無、従来法との組み合わせの有無によって 7 つの演奏を用意した (表 2)。従来法としては、既存手法 [10] による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付けを利用した。表情豊かな演奏の評価指標として、「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 人の被験

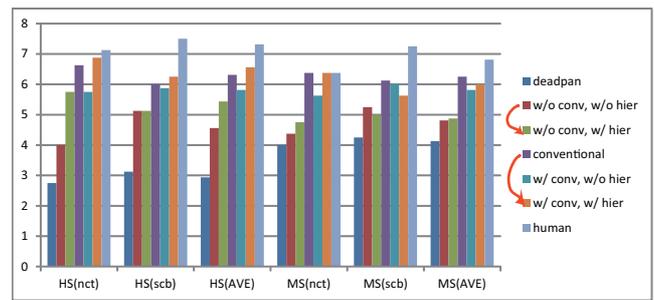


図 6 7 種類の演奏表情付けに対する評価値の平均

表 3 大フレーズ及び小フレーズのパラメータを定数倍した演奏毎の、最も表情豊かと評価した人数。

	0	$\Theta_l$	$2\Theta_l$	$4\Theta_l$	$4\Theta_g$
0	0	0	0	0	0
$\Theta_g$	0	0	0	0	0
$2\Theta_g$	0	0	2	4	0
$4\Theta_g$	0	1	2	1	0
$6\Theta_g$	0	0	0	0	0

者に対して 10 段階評価を行った。被験者の構成はピアノ演奏 10 年以上の経験者が 5 人と、ピアノ演奏の未経験者が 5 人であった。

### 5.2 階層的なフレーズ構造に基づくテンポ変動モデルの有効性の評価

2 つの評価指標 (HS, MS) についての評価値の平均を図 6 に示す。さらに HS, MS の全楽曲についての平均 (HS(AVE), MS(AVE)) も併せて示す。

HS(AVE) を見ると、w/o conv, w/ hier の評価値は w/o conv, w/o hier の評価値よりも高い値であり、w/ conv,w/ hier の評価値は conventional の評価値よりも高い値であった。

階層的なフレーズ構造に着目することで、実際により人間が演奏しているような演奏が生成され、テンポ変動モデルは有効であることが分かる。また、従来法による演奏表情生成を組み合わせた場合についてもより人間が演奏しているような演奏が生成された。

## 6. ユーザが介入する演奏表情付けの検討

あるフレーズ構造が与えられたとき、「ここは 2 倍程度の変動を加えたい」など、変動割合の決定によるユーザ介入は比較的容易と考えられる。一方で、フレーズを解釈したテンポ変動の形状をユーザが決定するのは煩雑である。よってまずはテンポ変動割合の決定によるユーザ介入法について述べる。

ユーザが決定するテンポの変動割合は、モデルパラメータの定数倍を扱うことによって反映できる。テンポ変動のパラメータを大フレーズ及び小フレーズの影響ごとに分けて ( $\Theta_g, \Theta_l$ ) とする。これらを定数倍した演奏を用意し、最も表情豊かに感じる演奏を被験者 10 名に主観評価させた結果を表 3 に示す。

小フレーズ及び大フレーズについてパラメータをそれぞ



図 7 楽曲中に存在する同じ動機に基づく類似部分

れ何倍すると最も表情豊かに感じるかは、元のパラメータに比べ定数倍を用いた方が、人間が最も表情豊かに感じる事が示された。これは、表情豊かな演奏とは二乗誤差の基準でフィッティングしたパラメータをそのまま用いて生成するのではなく、それらのパラメータを定数倍して、いわば少し「大きさに」テンポ変動を付与する方がより人間が演奏しているらしく感じることを示している。

演奏表情の階層性に基づく、ユーザの介入が容易な部分については、機械による自動決定を行うより寧ろユーザの介入を許す方が適している。具体的にここでは、各階層のテンポ変動に対してそれぞれ重み係数が掛かって重ね合わされるテンポ変動生成を考える。予め妥当な重み係数セットを用意し、それらをユーザが決定するとすれば、あまりテンポ変動を付与したくない場合や極端に付与したい場合など、ユーザの好みによってテンポの変動割合が変更できる。ユーザは重み係数の簡単な選択を行い、機械はそれに基づき、モデルに従った形状を持つテンポ変動を自動生成する。このように、簡単な指示を人間が担当しそれに基づく煩雑な自動生成は機械が担当するという、人間と機械が協調する演奏表情付けが可能と考えられる。

## 7. フレーズ構造の自動解析

フレーズ構造をいかに解析するかについて、楽曲の生成プロセスに着目して自動解析する手法を検討した。フレーズ構造が楽譜から解析できれば、解析されたフレーズ構造に基づく演奏表情付けの自動化が可能となる。

### 7.1 類似部分の検出によるモチーフの解析

音楽理論では、まとまりのある音符列を動機(以下、モチーフ)と呼ぶ。さらに、複数のモチーフがまとまってフレーズが構成される。楽曲が作曲される際にはまずモチーフを用意しそれを变形して用いることで楽曲が構成されることが多いことから、一楽曲中には多くのモチーフとそれらが变形したものが存在することが多い。同一のモチーフやその变形が使われた部分には何らかの類似性がある。図 7 に例を示す。その類似部分を検出することによってモチーフとフレーズ構造を解析できる可能性がある。

### 7.2 累積距離最小経路探索に基づくモチーフ・フレーズ構造の解析法

#### 7.2.1 特徴量と距離行列

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  $i$  について、鳴っていれば 1、そうでなければ 0 に値を持つ 127 次元(ピアノの鍵盤数)の特徴ベクトル  $\mathbf{x}_t = [x_{i,t}] = [x_{1,t}, x_{2,t}, \dots, x_{127,t}]$  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。

#### 7.2.2 累積距離最少経路探索問題への帰着

モチーフの变形は、(1) 繰り返される場合、(2) 变形される場合、(3) 異なる音高に移される場合が存在する。

繰り返し現れるモチーフ((1)の場合)及びその類似部分((2)の場合)は、 $d(s; t)$  を用いて構成される距離行列  $D_T$  の中で斜めの線として現れる。それらを検出する問題は、行を入力時間、列を参照時間、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、実際の楽曲では同一モチーフが異なる音高に移されて繰り返し使われる((3)の場合)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  $\mathbf{x}_{t,n} = [x_{1+n,t}, x_{2+n,t}, \dots, x_{127+n,t}]$  の間の距離

$$d_n(s; t) = D(\mathbf{x}_s || \mathbf{x}_{t,n}) = |\mathbf{x}_s - \mathbf{x}_{t,n}| \quad (5)$$

によって構成される距離行列を 12 種類用いれば、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、音高が異なる同一モチーフや類似部分が検出できる。今回は、音高のシフトはオクターブ以内で生じることが多いことから、シフトの種類数を  $N = 12$  とした。このようにして検出されたモチーフを 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。

### 7.3 階層的なフレーズ構造の性質と解析法

大フレーズ構造に関して小フレーズと同様の類似性があることから、その解析は類似部分検出問題と捉えることができ、距離行列に対する累積距離最小経路探索問題に帰着される。

ただし、大フレーズの解析の際に小フレーズの自動解析結果を利用できる。大フレーズの境界は小フレーズの境界の中から選ばれるので、遠方でのジャンプは小フレーズで境界とした箇所でのみ許される。また、大フレーズの経路は小フレーズの経路よりも連続しやすいため、遠方へのジャンプするコストは小フレーズのそのコストより高く設定される。

### 7.4 フレーズ構造解析結果の議論

提案する楽曲構造解析法の有効性を述べる。提案法によって自動で得たフレーズ構造を入力として、自動演奏表情付けを行った。提案法の評価は、自動演奏表情付けにお

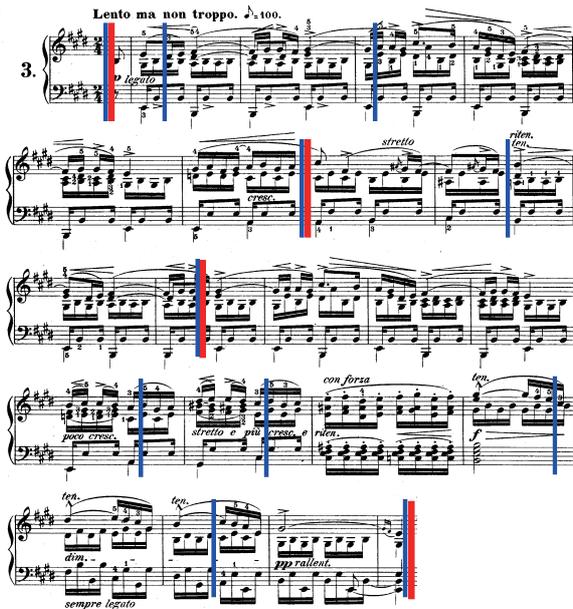


図 8 大フレーズ及び小フレーズの構造解析例。

ける有効性がひとつの基準となる。自動演奏表情付けに対する主観評価実験の結果、生成された演奏表情がより表情豊かであると示されたことから、演奏表情付けのための構造解析手法として有効と考えられる。演奏表情付けの評価に対する、構造解析結果の影響度は興味とするところである。これに対しては例えば、同じ演奏表情付け法で構造解析結果のみを変えて、自動演奏表情付けに対する主観評価がどう変わるかを考察することで検証できる。しかし、「構造解析結果を変える」方法は一意ではなく、音楽的な知識を基にいかにか構造解析結果を変えて演奏表情付けを行うかという点は十分な検討が必要である。

得られた構造解析例を図 8 に示す。青い線が小フレーズの解析結果で赤い線が大フレーズの解析結果である。得られたフレーズ構造から、類似部分が検出されていることが分かるが、正確な解析でない部分もある。正確な解析でない部分に注目すると、確かに類似度は高いが、フレーズ構造解釈としては不自然であるという部分が多い。これは、類似部分検出による手法の限界でもあり、ここに別の制約(小節線や和声境界)の必要性が示唆されている。また、大フレーズ構造境界には小フレーズの構造境界が選ばれ、類似部分が小フレーズとして解析されていることが分かる。今後は構造解析結果自体の評価として、ラベルを用いた客観評価が必要である。

## 8. おわりに

本報告では、楽曲構造に基づき人間が演奏しているような演奏を生成する技術及び楽譜から楽曲構造を解析する技術について論じた。これまでに提案された下層の演奏表情に着目した演奏表情に対し、より上層の演奏表情を組み合わせた手法によって、より人間が演奏しているような演奏が生成できることを確認した。演奏表情の階層性に基づき、演奏表情付けシステムの利用場面におけるユーザの介

在について議論した。また、演奏表情付けの自動化のために、フレーズ構造解析を楽譜から類似部分を探索する問題として扱った。

今後の課題としては、演奏表情付け技術に関しては、フレーズ構造に基づくテンポ変動を生成するパラメータを楽譜情報からいかに自動で得るかという問題、フレーズ構造をどのような特徴量と尺度によって得るかという問題について、今後のさらなる検討が必要である。構造解析技術に関しては、構造解析手法自体の評価として、ラベルを用いた客観評価が必要である。さらに、特徴量や距離尺度を変えた実験によって、適した特徴量の検討と考察が必要である。

## 参考文献

- [1] J. Sundberg, A. Askenfelt and L. Frydén, Musical Performance. A synthesis-by-rule approach, *Computer Music Journal*, vol. 7, no. 1, pp. 37–43, 1983.
- [2] 平賀瑠美, 五十嵐滋, 松浦陽平, “統合演奏視覚化システム,” 情報処理学会論文誌, vol.38, No.11, pp.2391–2397, 1997
- [3] 鈴木泰山, 徳永健伸, 田中穂積, “事例に基づく演奏表情の生成,” 情報処理学会論文誌, vol.41, no. 4, pp. 1134–1145, 2000.
- [4] K. Hirata and R. Hiraga, “Next Generation Performance Rendering –Exploiting Controllability,” *Proc. ICMC*, pp. 360–363, 2000.
- [5] R. Bresin, Artificial neural networks based models for automatic performance fo musical scores, *Journal of New Music Research*, vol. 27, no. 3, pp. 239–270, 1998.
- [6] S. Flossmann, M. Grachten and G. Widmer, “Expressive Performance Rendering: Introducing Performance Context,” *Proc. SMC*, pp. 155–160, 2009.
- [7] K. Hirata and R. Hiraga, “Ha-Hi-Hun: Performance rendering system of high controllability,” *Proc.the ICAD*, pp. 40–46, 2002.
- [8] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*, The MIT Press, 1983.
- [9] H. Keller, *Phrasierung und Artikulation*, Bärenreiter-Verlag Karl Vötterle KG, Kassel und Basel, 1955.
- [10] T. H. Kim, S. Fukayama, T. Nishimoto and S. Sagayama, “Polyhymnia: An automatic piano performance system with statistical modeling of polyphonic expression and musical symbol interpretation,” *Proc. NIME*, pp.96–99, 2011.
- [11] 長嶋洋一, 平賀謙, 橋本周司, 平田圭二, コンピュータと音楽の世界-基礎からフロンティアまで, 共立出版, 1999.
- [12] R. Aiello and J. A. Sloboda, Music Performance: Expression and the Development of Excellence, *Musical Perceptions*, pp. 152–169, Oxford University Press, 1994.
- [13] N. P. Todd, A model of expressive timing in tonal music, *Music Perception*, vol. 3, no. 1, pp. 33–58, 1985.
- [14] 大村 哲弥, 演奏法の基礎, 春秋社, 1998.
- [15] 保科 洋, 生きた音楽表現へのアプローチ:エネルギー思考に基づく演奏解釈法, 音楽之友社, 1998.
- [16] A. Tobudic, G. Widmer, Relational IBM in Music with a New Structural Similarity Measure, *Inductive Logic Programming*, vol. 2835, pp. 365–382, 2003.