

事例に基づく演奏表情生成システムにおける 旋律類似性の評価内容の視覚化

日野達也[†] 野池賢二 鈴木泰山^{††}
徳永幸生[‡] 杉山精[‡]

事例に基づく演奏表情生成システムの一つである“Kagurame Phase-II”では、楽曲を様々な長さの階層的な旋律断片に分割し、旋律断片の類似性を事例選択の指標としている。本稿では、本システムにおける旋律類似性の評価内容を視覚化して検証した結果について述べる。本システムでは、旋律断片の長さ、平均音高の比率、リズムパターン、音高分布について旋律断片の類似性を評価している。この評価の内容を画像によって視覚化した結果、本システムにおける旋律類似性の評価内容が直感的に理解できるようになった。さらに、旋律類似性の評価内容を視覚的に表現した画像から、画像比較による旋律類似性の評価可能性を検討した。

A Research of Visual Representation for Phrase Similarity Evaluation in A Case-Based Performance Rendering System

Tatsuya Hino[†] Kenzi Noike Taizan Suzuki^{††}
Yukio Tokunaga[‡] and Kiyoshi Sugiyama[‡]

“Kagurame phase-II” is one of case-based performance rendering system. This system divides target piece and example piece in performance case DB into various length and hierarchical phrase segments, and evaluates similarity of each segment to find appropriate segment from the DB. In this paper, we describe a method of visual representation for phrase similarity evaluation. As a result, we found that content of phrase similarity evaluation came to be intuitively understandable, and we found that phrase similarity evaluation is based on image information is possible.

1. はじめに

コンピュータによる表情の付いた演奏の自動生成は、音楽情報科学の主要なテーマの一つとなっている[1]。人間が行うような抑揚や情緒のある演奏を自動的に行う演奏表情生成システムが、1980年代から数多く発表されている[2][3][4][5][6]。また、これらの演奏表情生成システムが生成した演奏を評価する試みとして、人間による聴取実験によって演奏を評価するコンテスト Rencon[7]が2002年から行われている。

我々は、事例に基づく推論手法を用いた演奏表情生成システム“Kagurame Phase-II”[8][9]の構築を行い、人間が行うような演奏表情の生成を目指している。事例に基づく演奏表情生成では、人間によって行われた演奏を事例として、その演奏に見られる演奏表情を転写することで演奏データを生成する。Kagurameでは、事例を選択する際に、対象曲との類似性を評価し、楽曲の類似性が高い演奏を事例としている。そのため、楽曲の類似性評価の内容が、生成される演奏に大きく影響する。

しかし、楽曲の類似性を定量的に定義することは難しく、類似性評価の内容を正確に分析することは容易ではない。このことがKagurameにおける旋律類似性の評価手法を改善する妨げになり、演奏表情の質の向上を困難にしている。

そこで本報告では、Kagurameにおける演奏表情生成過程の根幹である「楽曲の類似性の評価」の内容を明らかにするために、システムで評価している楽曲の特徴量を画像で視覚的に表現した。

こういった特徴量の視覚化の例として、プログラミング学習の支援に関する試みがある。これは、変数の挙動や処理の概要を視覚化し、初学者のプログラミング理解を助けようとする試みである。Kagurameにおける旋律類似性評価の内容を理解する際にも視覚化は有効であると考え、類似性評価における特徴量を画像化することで、類似性評価の内容を人間が直感的に理解できるようになると考える。また、それによって類似性評価の内容を定量的に評価できるようになり、定量的に評価する方法を考えていく上での有効な手がかりになると考える。

以下、2章で事例に基づく演奏表情生成システム“Kagurame Phase-II”における演奏表情生成手法について概説する。3章でKagurameにおける旋律類似性の評価内容の視覚化について述べる。4章で旋律類似性の評価内容を視覚化した結果を述べ、5章で視覚化した結果について考察する。

[†] 芝浦工業大学大学院 工学研究科 電気電子情報工学専攻
Division of Electrical Engineering and Computer Science,
Graduate School of Engineering, Shibaura Institute of Technology

^{††} 株式会社ピコラボ
Picolab Co., LTD

[‡] 芝浦工業大学 工学部 情報工学科
Department of Information Science and Engineering,
College of Engineering, Shibaura Institute of Technology

2. 事例に基づく演奏表情生成システム

本章では，“Kagurame Phase-II”で用いている，事例に基づく演奏表情生成手法について説明する。

2.1 システムの概要

“Kagurame Phase-II”の構成を図 2.1 に示す。Kagurame では，演奏表情生成に関する知識として演奏事例 DB を用いる。個々の演奏事例は演奏データと演奏曲情報からなる。演奏データは，人間による演奏を標準 MIDI 形式ファイル(SMF)で与える。演奏曲情報は，演奏データに対応する楽譜情報と，音楽的なまとまり（旋律断片）を示す境界情報からなり，楽譜情報は MusicXML で与え，境界情報は独自フォーマットの XML ファイルで与える。対象曲情報は演奏曲情報と同様の形式で与える。境界情報は，個々の演奏事例と対象曲を旋律断片に分割する際に用いる。図 2.2 に示すように，対象曲や演奏事例を様々な長さの旋律断片に分割し，旋律断片を対象に演奏事例を検索することで，事例データのスパースネス問題に対応し，事例を効率よく活用している。

入力を与えられると，まず，対象曲を構成する旋律断片ごとに，その断片に類似した演奏事例の旋律断片を演奏事例 DB から検索する。これによって，対象曲の旋律断片ごとに，類似した旋律断片の集合が得られる。この集合を参考事例集とする。

次に，参考事例集に含まれる事例（参考事例）について，重要度を評価する。重要度は，対象曲の演奏表情を生成する際に，各事例がどの程度参考になるかを表すスコアである。参考事例の重要度は，対象曲と事例との旋律断片の類似性から決定する。したがって，対象曲の断片と参考事例の断片とが類似しているほど，重要度が高くなる。また，参考事例の重要度の評価とともに，参考事例の演奏表情を分析する。演奏表情は，演奏データと楽譜とのずれという形で取り出すことができる。Kagurame では，演奏表情をテンポや音の強さなどの絶対的な数値ではなく，それを変化量の比率に変換した相対的な数値で扱っている。この相対的な変化量を，演奏表情の大局比率と呼んでいる。

それぞれの参考事例に対して，重要度の評価と，演奏表情の分析を行うと，重要度でスコア付けされた演奏表情の大局比率の集合が得られる。この集合における個々の演奏表情の大局比率を，重要度で加重平均して合成し，対象曲の演奏表情を生成する。最後に，生成した演奏表情を対象曲の楽譜情報に適用して，演奏表情のついた対象曲の演奏データの SMF ファイルを作成する。

2.2 旋律類似性の評価手法

旋律類似性評価は，類似事例の検索と重要度の評価をする際に用いている。Kagurame では，以下の 4 種の旋律類似性評価式を用いて旋律の類似性を評価している。

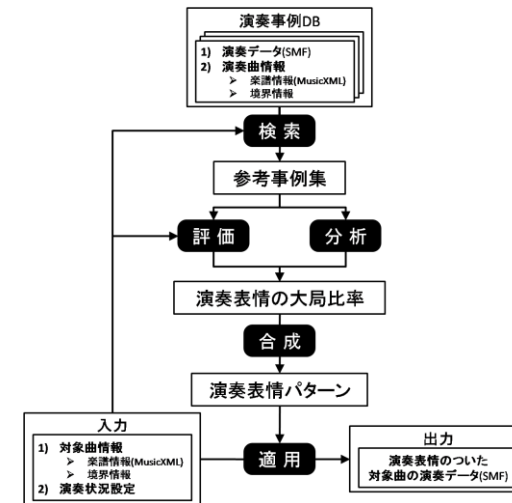


図 2.1 “Kagurame Phase-II”の構成

- 1) **Length Evaluator**
 旋律断片の長さを評価する。旋律断片の拍数の比率から長さの類似性 D_l を求める。
- 2) **KeyRate Evaluator**
 旋律断片の相対的な音高の高低を評価する。相対的な音高は，評価対象の旋律断片の，その断片を包含する旋律断片に対する音高の比率とする。この音高の比率を用いて音高の類似性 D_k を求める。
- 3) **Rhythm Evaluator**
 旋律断片のリズムパターンを評価する。リズムパターンは，旋律断片を 16 個の枠に分割し，枠ごとの音符の分布とする。このリズムパターンを比較してリズムの類似性 D_r を求める。
- 4) **Harmony Evaluator**
 音価を考慮した音の分布を評価する。分布は，旋律断片中の音符を階名で表した時のそれぞれの音の分布とする。音階の特徴の類似性 D_h を求める。

これら 4 つの旋律類似性評価式によって求めた旋律類似性を式(1)により正規化して合計し，指数関数に適用して事例の重要度とする。

$$\text{重要度 } R(s_i, t_j) = e^{-(A_l D_l + A_k D_k + A_r (1 - D_r) + A_h (1 - D_h))} \quad (1)$$

ここで、 $A = (A_i, A_k, A_r, A_h)$ は重み付け変数であり、それぞれ 0 以上の値をとる。この変数 A を変化させることで、同一の楽曲に対して、異なる演奏表情を持つ演奏を生成できる。変数 A の各値が小さいときには、重要度の低い参考事例も含めて、複数の事例をもとに演奏表情が生成される。一方、 A の値が大きくなるにつれて、演奏表情の生成に使用される事例は重要度の高いものに絞られていく。

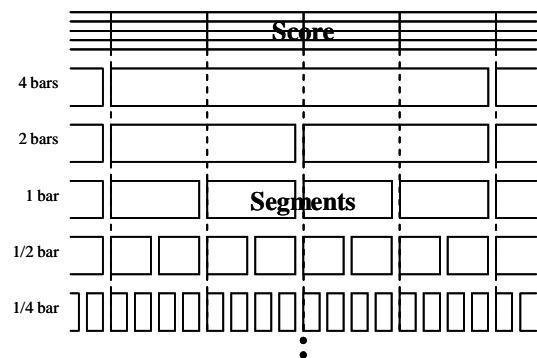


図 2.2 旋律断片への分割

3. 旋律類似性の評価の視覚化

Kagurame では旋律の類似性によって事例を選択し、その事例に基づいて演奏表情を生成する。この旋律類似性評価の内容によって選択される参考事例が変化するため、類似性評価の内容の改善が、人間らしい演奏を生成する上で重要であると考えられる。しかし、類似性評価の内容を定量的に定義することは難しく、定量的に類似性評価の内容を分析し、改善していくことは容易ではない。そこで、類似性評価に用いる旋律類似性評価式で評価している楽曲の特徴量を画像化する。

この画像を用いて類似性評価の内容を定量的に評価することを試みる。特徴量を画像化することで、類似性評価の内容を人間が直感的に理解でき、類似性評価の内容の定性的な評価が可能になると考える。定性的な評価が可能になれば、定量的に類似性の評価をする方法を考えていく上での有効な手がかりになると考える。本報告では、Kagurame で用いている 4 種の評価式のうち、Rhythm Evaluator と Harmony Evaluator の 2 種を例に、それぞれにおける旋律類似性の評価内容の視覚化について述べる。

3.1 Rhythm Evaluator

Rhythm Evaluator では、旋律断片を時間軸上で 16 個の枠に分割し、それぞれの枠を、各枠内で発音が始まる音符の音価の和によって重み付けしている。つまり、

Rhythm Evaluator で評価される旋律断片の特徴量は、要素数 16 のベクトルで表される。図 3.1 に示す旋律断片を Rhythm Evaluator で評価する際の特徴量を表すベクトルを表 3.1 に示す。



図 3.1 旋律断片の例

表 3.1 Rhythm Evaluator における図 3.1 の旋律断片の特徴量ベクトル

枠	1	2	3	4	5	6	7	8
音価の合計	1.25	0.25	0.25	0.25	1.0	0.0	1.0	0.0
枠	9	10	11	12	13	14	15	16
音価の合計	1.25	0.25	0.25	0.25	1.0	0.0	1.0	0.0

この旋律断片の特徴量を視覚的に表すために、枠の重みにしたがって色づけをし、横方向に時間軸を取って 16 本の縦縞によって表現した画像を図 3.2 に示す。色づけは、旋律断片中の音価の和に対する枠内の音価の和の比率に従って行った。色づけに用いた色は、図 3.3 に示す色相が 0 度-240 度の範囲の色である。枠内の音価の和の比率が 0 であれば色相 240 度の青となり、比率が大きくなるにつれて色相 0 度の赤に近づいていく。この画像によって Rhythm Evaluator では、時間軸方向の音の分布の類似性を評価していることが視覚的に理解できる。

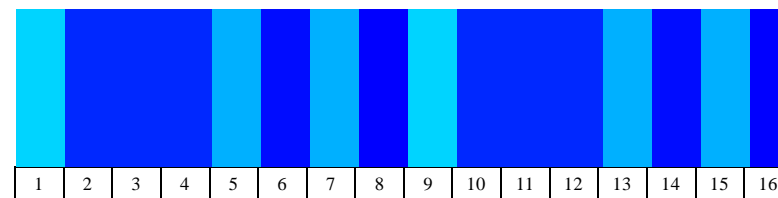


図 3.2 Rhythm Evaluator における図 3.1 の旋律断片の特徴量を視覚化した画像



図 3.3 色づけに用いた色

3.2 Harmony Evaluator

Harmony Evaluator では、旋律断片中の音符を階名で表し、オクターブの差異を取り除いた上で、各階名にその音価の合計で重み付けをする。つまり、Harmony Evaluator で評価される旋律断片の特徴量は、要素数 12 のベクトルで表される。図 3.1 に示す旋律断片を Harmony Evaluator で評価する際の特徴量を表すベクトルを表 3.2 に示す。

表 3.2 Harmony Evaluator における図 3.1 の旋律断片の特徴量ベクトル

音階	音価の合計	音階	音価の合計
I	1.5	IV#	0.0
I#	0.0	V	0.0
II	1.0	V#	0.5
II#	0.5	VI	0.5
III	2.0	VI#	0.0
IV	0.5	VII	1.5

この旋律断片の特徴量を視覚的に表すために、枠の重みにしたがって色づけをし、縦方向に下から音階を並べて 12 本の横縞によって表現した画像を図 3.4 に示す。色づけは、旋律断片中の音価の和に対する各階名の音価の和の比率に従って行った。色相の範囲は図 3.3 に示すとおりである。断片中で、どのオクターブにおいても音符のない階名の部分は色相 240 度の青となり、比率が大きくなるにつれて色相 0 度の赤に近づく。この画像によって、Harmony Evaluator では旋律断片中の音を 1 オクターブ内に集約して、その音高分布の類似性を評価していることが視覚的に理解できる。



図 3.4 Harmony Evaluator における図 3.1 の旋律断片の特徴量を視覚化した画像

4. 実験結果

4.1 Rhythm Evaluator による類似性評価内容の視覚化

図 4.1 は対象曲の旋律断片であり、図 4.2、図 4.3 は図 4.1 に示す対象曲に対して検索された事例の旋律断片である。それぞれ図の(a)は旋律断片に対応する五線譜、(b)は Rhythm Evaluator における旋律断片の特徴量を画像で表現したものである。それぞれの図の(b)を比較すると、縦縞の現れ方から図 4.2 の事例のほうが、図 4.3 の事例よりも図 4.1 に示す対象曲に類似している。また、図の(a)においても、16 分音符の分布の仕方などから図 4.2 に示す事例の方が対象曲に類似していることがわかる。つまり、Rhythm Evaluator による類似性評価と、それぞれの図の(b)に示した画像の類似性が一致していることから、これらの画像は、Rhythm Evaluator の評価内容を視覚化できているといえる。



図 4.1(a) 対象曲の旋律断片の五線譜

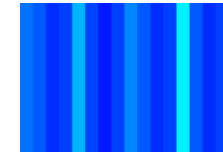


図 4.1(b) Rhythm Evaluator における図 4.1(a) の特徴量画像



図 4.2(a) 重要度第 1 位の事例の五線譜

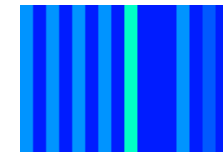


図 4.2(b) Rhythm Evaluator における図 4.3(a) の特徴量画像



図 4.3(a) 重要度第 10 位の事例の五線譜

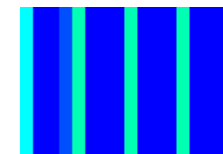


図 4.3(b) Rhythm Evaluator における図 4.3(a) の特徴量画像

4.2 Harmony Evaluator による類似性評価内容の視覚化

図 4.4 は対象曲の旋律断片であり、図 4.5、図 4.6 は図 4.1 に示す対象曲に対して Harmony Evaluator によって検索された事例の旋律断片である。それぞれ図の(a)は旋律断片に対応する五線譜、(b)は Harmony Evaluator における旋律断片の特徴量を画像で表現したものである。それぞれの図の(b)を比較すると、横縞の現れ方から図 4.5 の事例の方が、図 4.6 の事例よりも図 4.4 に示す対象曲に類似している。また、図の(a)においても、階名ソが多く現われている点や、旋律断片中の音高の分布の仕方から図 4.5 の事例のほうが対象曲に類似していることがわかる。つまり、Harmony Evaluator による類似性評価と、それぞれの図の(b)に示した画像の類似性が一致していることから、これらの画像は、Harmony Evaluator の評価内容を視覚化できているといえる。



図 4.4(a) 対象曲の
旋律断片の五線譜

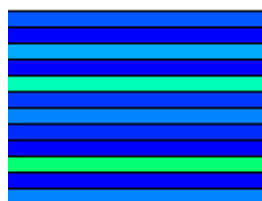


図 4.4(b) Harmony Evaluator
における図 4.1(a)の特徴量
画像



図 4.5(a) 重要度第 1 位
の事例の五線譜



図 4.5(b) Harmony Evaluator
における図 4.3(a)の特徴量
画像



図 4.6(a) 重要度第 2 位の
事例の五線譜

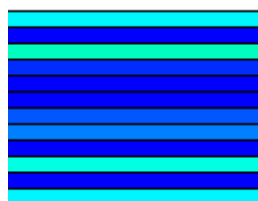


図 4.6(b) Harmony Evaluator
における図 4.3(a)の特徴量
画像

4.3 各評価式の評価内容の特徴

4.3.1 Rhythm Evaluator による長い旋律断片の評価

長さが 2 拍、4 拍、8 拍、16 拍の旋律断片の Rhythm Evaluator における評価内容を画像化したものを図 4.7 に示す。Rhythm Evaluator は旋律断片の長さにかかわらず、旋律断片を 16 分割してリズムパターンを分析する。つまり、旋律断片の長さが長くなるにつれて、相対的に Rhythm Evaluator の分解能が低くなる。これは、評価対象とする旋律断片の長さが長くなるほど、局所的なリズムの特徴よりは大局的なリズムの特徴を考慮したいためであり、Rhythm Evaluator では、これによって事例データのスパースネス問題に対応している。しかし、16 拍程度の長さの断片の評価で、すでに枠ごと

のリズムの特徴がほぼ同様の評価結果になってしまい、断片同士の差異をほとんど区別できなくなってしまう。これは、図 4.7 に示す、長さ 16 拍の断片の特徴量画像のように色の差異がわかりにくくなっていることから見て取れる。そのため、現在の Rhythm Evaluator は、扱う長さが長くなったときに意図どおりに機能していない可能性がある。

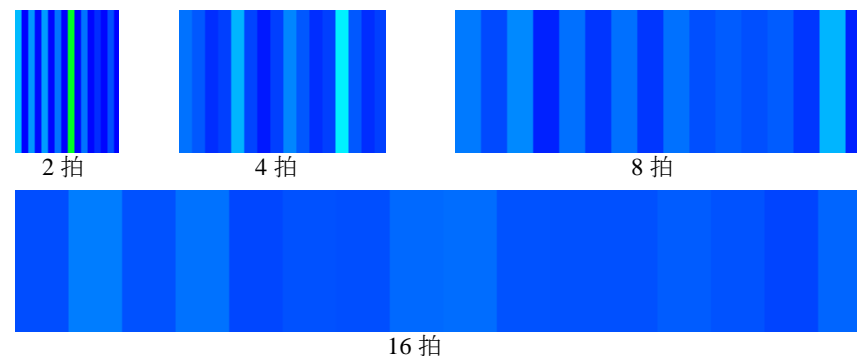


図 4.7 長さの異なる旋律断片の Rhythm Evaluator における特徴量画像

4.3.2 Harmony Evaluator による短い旋律断片の評価

長さが 2 拍、4 拍、8 拍、16 拍の旋律断片の Harmony Evaluator における評価内容を画像化したものを図 4.8 に示す。Harmony Evaluator では、旋律断片中の音高の分布を 1 オクターブに集約して評価する。そのため、旋律断片の長さが短くなるにつれて旋律断片に含まれる音符が少なくなり、Harmony Evaluator の効果が低くなると考える。

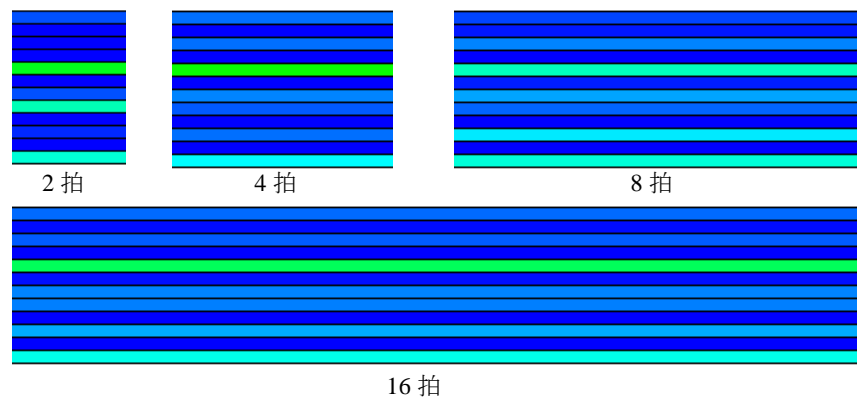


図 4.8 長さの異なる旋律断片の Harmony Evaluator における特徴量画像

5. 考察

5.1 特徴量の視覚化

実験結果から、Rhythm Evaluator, Harmony Evaluator とともに、五線譜どうしでの比較と画像どうしでの比較とで、それぞれの類似性が一致したことから、評価式の視覚化が適切に行われていることがわかった。また、人間が画像を見ても断片ごとの特徴量の違いや、類似性が見て取れ、特徴量を表す画像から旋律の類似性を評価できる見通しを得た。しかし、今回画像の色づけに用いた色相 0 度-240 度の範囲が視覚化に適切であるかなど、考慮すべき課題がある。

5.2 旋律類似性の評価手法

特徴量を画像による表現を用いた視覚化によって、次のことが顕著になった。

まず、Rhythm Evaluator において長い旋律断片の場合には、16 分割した枠ごとの音符の分布に差がなくなり、断片ごとの違いが現れにくいことがわかった。これは、旋律が長くなるほど、細かなリズムを考慮しない評価によって、データスパースネスに対応するという意図が Rhythm Evaluator に反映されていることを示している。また、人間が長い範囲で楽譜を見るときは局所的な違いを考慮しないことと同様であると考えられる。しかし、長さが 16 拍程度の断片でこのような評価をしてしまうと、Rhythm Evaluator の効果が期待できない。したがって、Rhythm Evaluator に評価する断片の拍数に応じて変化する枠の重みを導入することによって、いままでよりも有効に機能させることができると考える。また、楽曲全体にわたって 8 分音符や 16 分音符などの細かい音符が出現する楽曲を評価する場合に枠ごとの差異が少なくなることが考えられる。このような場合にも断片ごとの違いを評価するために、音価に対する重み付けを調整可能にし、音符が密になっている枠を強調することが有効であると考えられる。

次に、Harmony Evaluator において、短い旋律断片の場合には断片中の音符が少ないため、音高の分布を評価するには適さないと考えられる。したがって、4 拍以上の長い旋律断片に対する類似性評価には、その評価値をより反映させるために重みを付け、4 拍未満の短い断片については、評価値を全体的に低くすることで、Harmony Evaluator をより有効に利用できると考える。

あるいは、Length Evaluator による評価値を Rhythm Evaluator や Harmony Evaluator による類似性評価の際に使用することで、断片の長さに応じた適切な評価が期待できる。

6. おわりに

本報告では、事例に基づく演奏表情生成システム “Kagurame Phase-II” における、旋律類似性評価の評価内容を視覚化について述べた。4 種の旋律類似性評価式のうち、まず、Rhythm Evaluator と Harmony Evaluator の評価内容を画像で表現することで視

覚化した。この視覚化によって評価手法の短所が顕著になり、その改善方法について検討できた。

今後は、Length Evaluator, KeyRate Evaluator についても同様に視覚化を行い、評価内容とその妥当性を明らかにする。また、視覚化に用いた画像を旋律類似性評価に用いることによって、画像を用いた視覚的な旋律類似性の評価が可能であるかどうか検討していく。さらに、より人間の直感に近い類似性評価の実現を目指すために、ピアノロール画像を用いた旋律類似性評価についても検討していきたい。

このように評価内容の視覚化を活用することで、旋律類似性評価の内容を改善し、生成される表情つき演奏の質を向上させていきたい。

参考文献

- 1) 平賀瑠美: 音楽の表情付け, bit 別冊, コンピュータと音楽の世界, pp.270-282, (1998)
- 2) Widmer, G: Inductive learning of general and robust local expression principles, Proceedings of the 2001 International Computer Music Conference, International Computer Music Association, pp.322-329, (2001)
- 3) 野池賢二, 豊田健一, 片寄晴弘: コーパスベース表情付けシステム COPER の基礎機能の実装とその評価, 情報処理学会研究報告, 2008-MUS-59, Vol.2005, No.14, pp.67-70 (2008)
- 4) 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)
- 5) 寺村佳子, 大熊秀治, 谷口雄作, 牧本慎平, 前田新一: ガウシアンプロセスによる名演奏の学習, 情報処理学会報告, 2008-MUS-78, vol.2008, No.127, pp.79-84 (2008)
- 6) Sebastian Flossmann, Maarten Grachten, and Gerhard Widmer: Experimentally Investigating the Use of Score Features for Computational Models of Expressive Timing, Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC 10), pp.218-223 (2008)
- 7) 橋田光代, 片寄晴弘, 平田圭二, 北原鉄朗, 鈴木健嗣: 演奏表情付けコンテスト ICMPC-Rencon 開催報告, 情報処理学会研究報告, 2008-MUS-78, vol.2008, No.127, pp.67-72 (2008)
- 8) 鈴木泰山, 金子雄介, 徳永幸生: 事例に基づく演奏表情生成アルゴリズムの分析, 情報処理学会研究報告, 2005-MUS-59, Vol.2005, No.14, pp.49-54 (2005)
- 9) 金子雄介, 鈴木泰山, 徳永幸生: 事例に基づく演奏表情生成システムにおける演奏類似性と聴覚評価, 情報処理学会研究報告, 2005-MUS-59, Vol.2005, No.14, pp.43-54 (2005)