

音高方向グラデーションにより音符を認識する ピアノロール画像の比較手法の検討

柴崎正浩[†] 鈴木泰山[‡] 米村俊一[†] 徳永幸生[†]

事例に基づく演奏表情生成システムの Kagurame Phase-III は、ピアノロール画像の比較による旋律の類似性を事例選択の指標として、人間らしい演奏の生成を行っている。しかし、これまでのピアノロール画像の類似性評価手法では、比較する音符同士のうち片方の音符の音高が半音以上異なると音符同士の相関が取れず、旋律の類似性を十分に評価できていなかった。そこで我々は以前に、音符の矩形の音高方向にグラデーションを追加したピアノロール画像と、音符がある画素と音符がない画素とを比較した際の比較評価手法を提案した。本稿ではそれに加え、ピアノロール画像の比較時において、一つの音符に対して相関を取った音符の個数の平均とその音符の長さの平均を旋律の類似性評価に取り入れる手法の提案を行う。さらに、以上3つの提案手法を導入することにより、より人間らしい演奏の生成が可能となったことを明らかにした受聴実験について述べる。

A Study of Piano-roll Images Comparison Method which Recognizes Notes by Pitch Direction Gradations

MASAHIRO SHIBASAKI[†] TAIZAN SUZUKI[‡]
SHUN-ICHI YONEMURA[†] YUKIO TOKUNAGA[†]

Kagurame Phase-III is one of the Case-based expressive performances rendering system. This system chooses an example data by note sequence similarity using piano-roll image. However, conventional similarity evaluation method of piano-roll images cannot recognize a note different in pitch. Therefore, the method was not able to evaluate note sequence similarity enough. There we in the past, we proposed new piano-roll images which have pitch direction gradations and an evaluation method which compares a white pixel and a black pixel. In this paper, we propose a new evaluation method and it is added to two proposed methods. And, we prove usefulness of these three method by audition experiment.

1. はじめに

現在の音楽情報科学の主要なテーマの一つに、コンピュータによる情緒や抑揚のある人間らしい表情の付いた演奏の自動生成がある[1][2]。コンピュータによって表情付けられた演奏をコンテスト形式で評価する試みである Rencon が 2002 年から定期的開催されるようになり、今までに演奏表情生成システムが数多く発表されている[3][4][5]。

その中で我々は、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-III の構築を進めている[6]。Kagurame Phase-III では「楽譜が似ていれば演奏表情も似ている」という仮定を基に、表情付けを行う対象曲と楽譜が類似した楽曲の人間の演奏を事例として用いる[7][8]。楽譜の類似性評価は楽譜情報 (MusicXML) から生成したピアノロール画像を比較することで行う。ピアノロール画像を用いることで、旋律の特徴量を抽象化・簡略化せずに比較でき、多くの特徴を反映した類似性評価が可能である[6][9]。

しかし、これまで用いていたピアノロール画像とその比較手法では、比較する音符同士のうち片方の音符の音高が半音以上異なると音符同士の相関が取れず、音符同士の比較評価ができない。そのため、旋律の類似性を十分に評価

できていないことが、我々のこれまでの実験から明らかとなった[10]。この問題により、人間の演奏とはほど遠い演奏が生成されていた。そこで我々は以前に、音符の矩形の音高方向にグラデーションを追加したピアノロール画像と、音符がある画素と音符がない画素とを比較した際の比較評価手法の提案を行った。しかし、この2つの提案手法を導入しただけでは、十分に解決できたとは言えなかった[10]。

そこで本稿では、さらにもう1つの手法を提案する。具体的には、ピアノロール画像の比較時において、一つの音符に対して相関を取った音符の個数の平均とその音符の長さの平均を類似性評価に取り入れる手法である。また、この3つの提案手法を導入した Kagurame Phase-III が、より人間らしい演奏の生成が実現できるのかを受聴実験によって明らかにする。

2. Kagurame Phase-III の概要

2.1 Kagurame Phase-III の構成

Kagurame Phase-III の構成を図 2.1 に示す。Kagurame Phase-III では演奏表情に関する知識として演奏事例 DB を用いる。DB の個々の演奏事例は演奏データと事例曲情報からなる。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。事例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る。楽譜情報は MusicXML で与え、境界情

[†] 芝浦工業大学
Shibaura Institute of Technology
[‡] 株式会社ピコラボ
Picolab Co., LTD

報は独自フォーマットのXMLファイルで与える。システムに入力する対象曲情報は事例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。旋律断片への分割は対象曲や事例曲を1/2, 1/4, 1/8...と様々な長さに分割する。これにより、事例データのスパースネス問題に対応することができる。

システムに対象曲が入力されると、まず、対象曲を構成する旋律断片(対象断片)ごとに演奏事例を構成する旋律断片(事例断片)との類似度を算出し、類似度が閾値以上となる事例断片を『検索』する。これにより、得られた類似した事例断片の集合を参考事例集と呼ぶ。次に、参考事例集に含まれる事例断片について重要度を『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断片がどの程度参考になるかを表す。重要度は類似度と演奏状況の近さから算出するが、本稿では演奏状況は未使用であるため重要度は類似度と同値となる。また同時に、事例断片の演奏表情を『分析』する。演奏表情は各事例断片の演奏データから取り出すことができる。なお、演奏表情はテンポや強弱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う。この相対的な変化量を演奏表情の大局比率と呼ぶ。本稿での演奏表情のパラメータは、局所テンポ、音の強弱、音の発音長、発音タイミングのずれの4種類としている。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、演奏表情の付いた対象曲の演奏データを作成する。

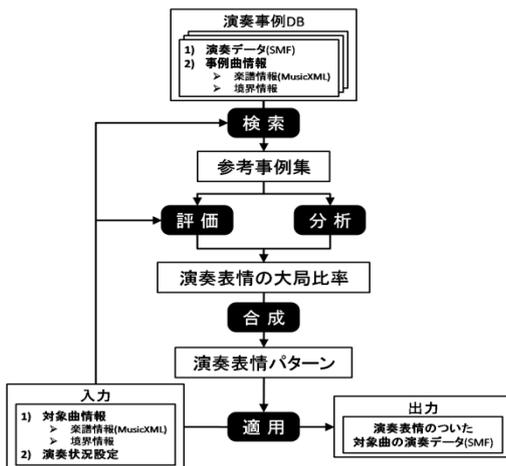


図 2.1 Kagurame Phase-III の構成

2.2 画像比較による旋律の類似性評価

類似度算出のために Kagurame Phase-III では楽譜画像の比較による類似性評価を用いる。これは、旋律の特徴量をより多く用いた類似性評価を行うためである。五線譜は演奏記号が多く煩雑であり、小節ごとの幅や段の配置などが様々であるためピアノロール画像を用いて比較する。図 2.2

にピアノロール画像と対応する五線譜を示す。ピアノロール画像の縦方向は音高、横方向は拍位置を表す。画像の高さは 88pixel、幅は旋律断片の長さに応じた幅となる。音符を表す矩形の高さは 1pixel、幅は 4 分音符を 32pixel としている。音符の拍位置に相当する画素を(R,G,B)=(255,255,255)、終端を(R,G,B)=(1,1,1)としたグラデーションをかけている。これは音符同士を比較する際に音符の拍位置の違いを評価するためである。音符が無い部分の画素は(R,G,B)=(0,0,0)としている。画像を比較評価する際、対象断片画像と幅が異なる事例断片画像は、対象断片画像と同じ幅に伸縮させてから評価する。また、音高が異なっても旋律の概形が類似していれば有効な事例であると考え、画像を音高方向に上下動させ、最も音符の矩形が一致する場合で評価する[9]。



図 2.2 ピアノロール画像と対応する五線譜

ピアノロール画像の類似性評価は式(1)(2)(3)を用いる。まず、画像相違度 D_{image} を式(1)より算出する。対象断片画像の各画素の明度を B_T 、事例断片画像の各画素の明度を B_S とし、 B_T と B_S の差分を算出する。この値に音符の拍位置の類似性を重視するために B_T と B_S の高い方の値をとる $Max(B_T, B_S)$ により重み付けをする。これを全ての画素で行い加重平均をとる。式(2)の画像類似度 S_{image} は画像相違度 D_{image} を 0 から 1 に正規化した値である。 w は画像伸縮と上下動分の重みであり式(3)より算出する。 $Shorter$ は短い方の旋律断片の幅、 $Longer$ は長い方の旋律断片の幅を表し、 $Height$ は画像の高さ、 $Overlap$ は画像を上下動した際の対象断片画像と事例断片画像の重なっている部分の高さを表す。 $Overlap$ の最大値は $Height$ と同じ 88pixel となるが、画像の上下動は 1 オクターブ分 (12pixel) までとしているため最小値は 76pixel となる。最終的に算出された画像類似度 S_{image} から事例断片を検索する。

$$\text{画像相違度} : D_{image}(t_i, s_j) = \frac{\sum |B_T - B_S| \cdot \text{Max}(B_T, B_S)}{\sum \text{Max}(B_T, B_S)} \quad (1)$$

$$\text{画像類似度} : S_{image}(t_i, s_j) = \left(1 - \frac{D_{image}}{255}\right)^w \quad (2)$$

$$\text{重み} : w = \frac{Longer}{Shorter} \cdot \frac{Height}{Overlap} \quad (3)$$

しかし、図 2.2 のピアノロール画像では、音符の矩形の高さを 1pixel としているため、図 2.3 のように対象断片画像の音符と音高が半音以上異なった事例断片画像の音符とは相関が取れなくなる。これでは、音符同士の比較ができず、旋律の類似性を十分に評価ができない[10]。

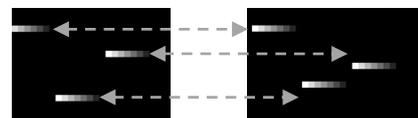


図 2.3 音符一つ分しか音符の相関が取れない例

3. 音高が異なる音符同士の相関の考慮

本章では第2章で述べたピアノロール画像の問題を解決するために提案した新しいピアノロール画像とその評価手法について述べる。3.1と3.2は以前に報告した手法であるため、本稿では詳細は割愛する[10].

3.1 音高方向グラデーションをかけたピアノロール画像

ピアノロール画像で表記された音符の矩形の音高方向にグラデーションを追加した新しいピアノロール画像(図3.1)を提案する。これにより、対象断片画像の音符と若干音高が異なる事例断片画像の音符との相関を取ることができ、音符同士の比較が可能となる。なお、グラデーションは上下方向の4pixel(4半音)分追加しており[9][10][13],グラデーションの終端の画素を(R,G,B)=(1,1,1)としている。

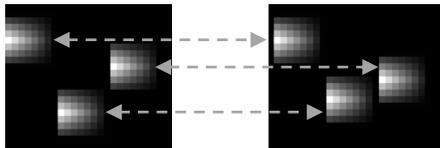


図 3.1 図 2.3 の例に音高方向グラデーションをかけたピアノロール画像

音高方向と拍位置方向ともにグラデーションをかけたため、音符の矩形の右端、上端、下端の明度が0に近い画素となる。ここでは、音符がある画素(明度が1以上の画素)と音符がない画素(明度が0の画素)とを比較評価した場合、明度の差分が小さくなり、類似性が高いと評価されてしまう問題がある。

3.2 音符がない画素との比較を考慮した比較手法

3.1で述べた新しいピアノロール画像の問題点を受け、音符がある画素と音符がない画素とを比較評価した際に画像類似度 S_{image} を下げるような重みを提案する。重みは式(4)(5)(6)を用いて算出する。 n は画素数を表す。式(4)の w_{sum} の項は、1から256の値をとり、 $\min(B_T, B_S)$ が小さいほど256に近づく。また、 $\min(B_T, B_S)=1$, $\max(B_T, B_S)=2$ の場合のように、両画素とも明度は0に近いが差分は小さい場合は、類似性が高いとみなす必要があるため、 $\max(B_T, B_S)$ で除算を行っている。そして、式(5)の w_{ave} により w_{sum} の平均値を算出し、式(6)の w_{norm} により w_{ave} を0から1に正規化する。この w_{norm} を画像類似度 S_{image} に掛け合わせる。音符がない画素とを比較評価した場合が多いほど w_{norm} は0に近づくため、画像類似度 S_{image} を下げることができる。

$$w_{sum}(t_i, s_j) = \sum^n \frac{256}{\frac{\min(B_T, B_S)^{255}}{\max(B_T, B_S)} + 1} \quad (4)$$

$$w_{ave} = \frac{w_{sum}}{n} \quad (5)$$

$$w_{norm} = 1 - \frac{w_{ave} - 1}{255} \quad (6)$$

3.3 複数の音符と相関を取った音符を考慮した比較手法

我々は以前に、3.1と3.2で提案した手法をKagurame Phase-IIIに導入しても、必ずしも類似した事例断片画像を

選出することができないことを報告した[10]. この理由の一つとして、一つの音符に対して複数の音符と相関を取った場合の対処がされていないためと考えられる。例えば、一つの音価が長い音符に対して、連符などの音価が短い複数個の音符と相関を取ってしまう場合がある。特に音高方向にグラデーションを追加したため、この問題が起こりやすい。ここでは、音符の数や長さが異なる旋律を包含する旋律断片同士でも類似性が高いと評価されてしまう。

そこで、一つの音符に対して相関をとった音符の個数の平均とその音符の長さの平均を類似性評価に取り入れる手法を提案する。具体的には、対象断片画像の一つの音符に対して相関をとった事例断片画像の音符の個数の平均を NTS_{ave} , それら音符の長さの平均を LTS_{ave} , 事例断片の一つの音符に対して相関をとった対象断片画像の音符の個数の平均を NST_{ave} , それら音符の長さの平均を LST_{ave} として、一つの音符に対して相関を取った音符の個数の重み w_n を式(7), その音符の長さの重み w_l を式(8)によって算出する。この2つの重みを式(9)によって掛け合わせて w_{nl} を算出し、画像類似度 S_{image} に掛け合わせる。

$$w_n = \frac{\min(NTS_{ave}, NST_{ave})}{\max(NTS_{ave}, NST_{ave})} \quad (7)$$

$$w_l = \frac{\min(LTS_{ave}, LST_{ave})}{\max(LTS_{ave}, LST_{ave})} \quad (8)$$

$$w_{nl} = w_n \cdot w_l \quad (9)$$

例えば、図3.2の対象断片画像と事例断片画像を比較評価した場合を考える。対象断片画像の音符Aに対して相関をとる事例断片画像の音符の数は音符a, b, cの3つである。また、音符a, b, cの音符の長さは24pixelであるため音符Aに対して相関をとる音符の長さの総計は72pixelとなる。このような計算を全ての音符ごとに行うと、対象断片画像の音符と相関をとる事例断片画像の音符の総数は19個、それらの音符の長さの総計は456pixelとなる。対象断片画像の音符の数は10個であるため、対象断片画像の一つの音符に対して相関をとる事例断片画像の音符の個数の平均(NTS_{ave})は1.9個、音符の長さの平均(LTS_{ave})は45.6pixelとなる。事例断片画像の音符に対しても同様に計算すると、事例断片画像の一つの音符に対して相関をとる対象断片画像の音符の個数の平均(NST_{ave})は5.25個、音符の長さの平均(LST_{ave})は56pixelとなる。従って、 $\min(NTS_{ave}, NST_{ave})=1.9$, $\max(NTS_{ave}, NST_{ave})=5.25$ となるため、 $w_n=0.36$ となり、 $\min(LTS_{ave}, LST_{ave})=45.6$, $\max(LTS_{ave}, LST_{ave})=56$ となるため、 $w_l=0.81$ となる。最終的に算出される w_{nl} は0.29となる。

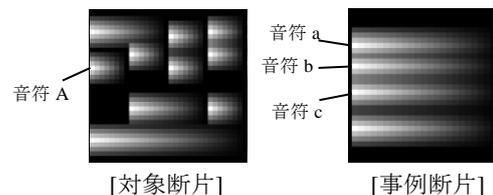


図 3.2 複数の音符と相関をとる音符の対処手法の例

4. 受聴実験

本章では、第3章で提案した3つの手法（以下、提案手法）を Kagurame Phase-III に導入して生成した演奏と、導入しないで生成した演奏で、演奏の人間らしさにどのような差異が生じるのかを明らかにする受聴実験について述べる。なお本稿では、従来通り提案手法を導入しないで生成する手法を従来手法と呼ぶこととする。

また我々は以前に、聴感による演奏表情の評価特性の分析を行い、テンポが演奏の印象に大きな影響を与える要素であることを報告した[11]。そして、テンポ記号の類似性評価手法の提案も行った[12]。そこで、今回の受聴実験では、テンポ記号の類似性評価を導入して生成した演奏も評価の対象とした。テンポ記号の類似性評価は式(10)(11)を用いる。 T_t は対象曲のBPM値、 T_s は事例曲のBPM値を表す。この差分を正規分布式に適用することでテンポ類似度 S_{tempo} を算出する。そして、画像類似度 S_{image} と掛け合わせることで、楽譜類似度 S_{score} を算出する。テンポ記号の類似性評価を導入して類似性評価を行う場合は楽譜類似度 S_{score} から事例断片を検索する。

$$\text{テンポ類似度} : S_{tempo} = e^{-\frac{(T_t - T_s)^2}{T_t + T_s}} \quad (10)$$

$$\text{楽譜類似度} : S_{score} = S_{image} \cdot S_{tempo} \quad (11)$$

4.1 実験手法

人間による演奏（見本演奏）を参考に Kagurame Phase-III で生成した演奏（評価対象演奏）を被験者に聴いてもらい、評価対象演奏の「自然さ」と「豊かさ」を評価してもらう。この指標は演奏表情付けコンテスト“Rencon”[5]で以前に用いられていた指標である。「自然さ」とは、どの程度違和感の無い演奏であるのか（ごこちなくないか、適切なテンポ・強弱であるか）を表し、「豊かさ」とは、演奏表情の付加の強さとその頻度を表す。この2つの評価が高ければ高いほど人間らしい演奏となる。従って、見本演奏は最も自然で豊かな演奏であると仮定して評価してもらう。被験者は図4.1に示した評価軸上に縦線を記入することで評価する。評価値は評価軸の左端からの距離とする。従って、評価値が大きいほど自然さは自然で、豊かさは豊かとなる。評価値の最大は6、最小は0とした。また、実験後にアンケート形式で「自然さ・豊かさの評価基準」「実験において難しいと感じた点」を記述式で回答してもらった。なお、実験前に評価手法に慣れてもらうため、練習として別の曲目を対象曲とした場合で3つの演奏を評価してもらった。



図4.1 実験で用いる評価軸

被験者は音楽経験者11名、未経験者11名の22名とした。対象曲は、Beethoven Piano Sonata No.8 Op.13-2「悲愴」冒頭8小節(BPM=30)、Chopin Nocturne No.10 Op.32-2 冒頭16

小節(BPM=60)(前奏は除く)、Grieg Lyric Pieces Op.12-1

「Arietta」冒頭12小節(BPM=60)とした。この3曲ごとに「従来手法で生成した演奏」（図4.2～図4.4では「従-無」と表記）、「従来手法にテンポ記号の類似性評価を導入して生成した演奏」（同様に「従-有」と表記）、「提案手法を導入して生成した演奏」（同様に「提-無」と表記）、「提案手法とテンポ記号の類似性評価を導入して生成した演奏」（同様に「提-有」と表記）の4種類ずつ（計12種類）を評価対象演奏とした。生成に用いた事例曲を表4.1に示す。()内の数値はBPM値である。試聴はヘッドホンを用いて行う。演奏を試聴する順序や評価する順序、試聴する回数は被験者の自由とする。ただし、見本演奏とそれぞれの評価対象演奏は、最低一回は最後まで聴くよう指示した。なお、被験者には試聴する曲名は伝えていない。

表4.1 生成に用いた事例曲

Chopin Etude Op.10-4 (176)	Chopin Prelude Op.28 No.15 (72)
Mozart Piano Sonata K545-1 (160)	Mozart Piano Sonata K545-2 (60)
Chopin Nocturne Op.32-1 (120)	Chopin Trauer Marsch (60)
Chopin Nocturne Op.55-1 (120)	Chopin Prelude Op.28 No.4 (60)
Burgmuller AveMaria (100)	Chopin Prelude Op.28 No.20 (40)
Chopin Prelude Op.28 No.7 (88)	Chopin Etude Op.10-3 (30)

4.2 実験結果と考察

4.2.1 有意差検定の結果と考察

対象曲が“Beethoven”の場合の結果を図4.2に、“Chopin”の場合の結果を図4.3に、“Grieg”の場合の結果を図4.4に示す。縦軸は評価値を表し、値が大きいほど自然さは自然で、豊かさは豊かとなる。横軸は演奏の生成条件を表している。黒塗りは音楽経験者、白塗りは未経験者を表す。有意差を表す実線は音楽経験者における生成条件間の有意差、破線は未経験者における生成条件間の有意差、二点鎖線は音楽経験有無間の有意差を表している。

図4.2～図4.4のaの有義差から、テンポ記号の類似性評価を導入して生成した演奏は、導入せず生成した演奏より自然な演奏であると評価されたことが示された。これは、テンポ記号が大きく異なる事例断片を排除することができたことから、極端なテンポのゆらぎが無くなり、ゆらぎが安定したためと考えられる。また、提案手法を導入することによって、自然さの評価が高くなるような有意差は認められなかった。このことから、cの有義差は提案手法ではなく、テンポ記号の類似性評価の導入によって認められたことが言える。以上より、テンポ記号の類似性評価の導入によって、より自然な演奏の生成が可能となったと言える。

一方、図4.2および図4.4のbの有義差から提案手法を導入して生成した演奏は、従来手法で生成した演奏より不自然な演奏であると評価されたことが示された。これは、提案手法は従来手法よりもより旋律が類似した事例断片を参照できたことから、様々なテンポの曲の事例断片を参照し、極端なテンポのゆらぎが発生したためと考えられる。また、図4.4中のhの有義差からテンポ記号の類似性評価を導入しても提案手法を導入した演奏は不自然であると評価された。テンポ記号の類似性評価の導入は自然さの評価を高め

る効果があることが a と c の有意差から考えられるため、この楽曲を対象曲とした場合は、提案手法よりもテンポの類似性評価手法を重視して導入すべきであることが示唆された。一方、図 4.2～図 4.4 中の d の有意差のように、提案手法を導入して生成した演奏は従来手法より豊かな演奏であると評価されたことが示された。これは、より旋律が類似した事例断片を参照できるようになり、より適切な演奏表情の適用ができたため、又は極端なテンポのゆらぎが発生したためと考えられる。また、テンポ記号の類似性評価を導入することによって豊かさの評価が高くなるような有意差は認められなかった。このことから、e の有意差はテンポ記号の類似性評価ではなく、提案手法の導入によって認められたことが言える。以上より、提案手法の導入によって、より豊かな演奏の生成が可能となったと言える。

図 4.3 および図 4.4 中の f の有意差から音楽経験者と未経験者で自然さの評価に違いがあることが示された。これは、後述するがアンケートの結果から音楽経験者はテンポのゆらぎが適切であるのかを重視して評価する人が多く、未経験者は曲全体のテンポが適切であるのかを重視して評価する人が多かったためと考えられる。また、図 4.3 中の g の有意差から音楽経験者と未経験者で豊かさの評価に違いがあることが示された。後述するがアンケートの結果から豊かさの評価基準は音楽経験者と未経験者で大きな違いは見られなかったことから、音楽経験で得た何らかの知識や音楽の聴き方が、無意識に豊かさの評価基準になったため、評価に違いが出たと考えられる。

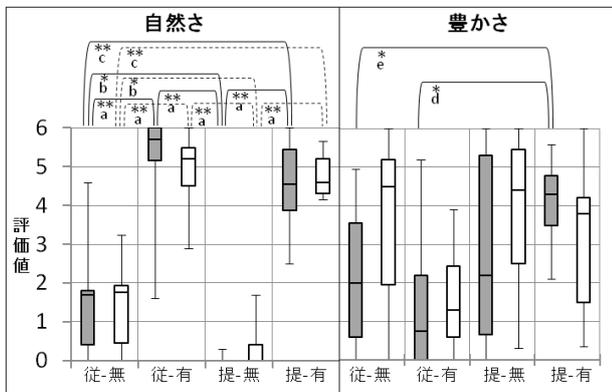


図 4.2 “Beethoven” の実験結果 **:p<0.01 *:p<0.05

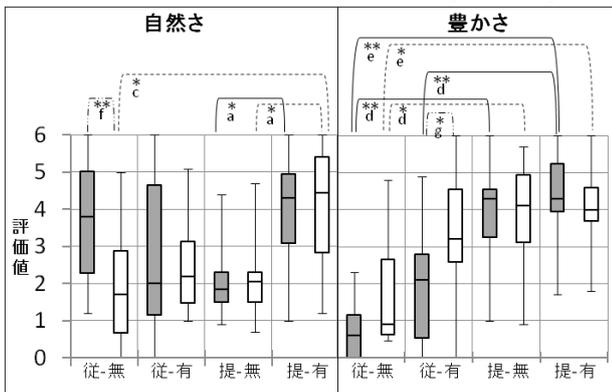


図 4.3 “Chopin” の実験結果 **:p<0.01 *:p<0.05

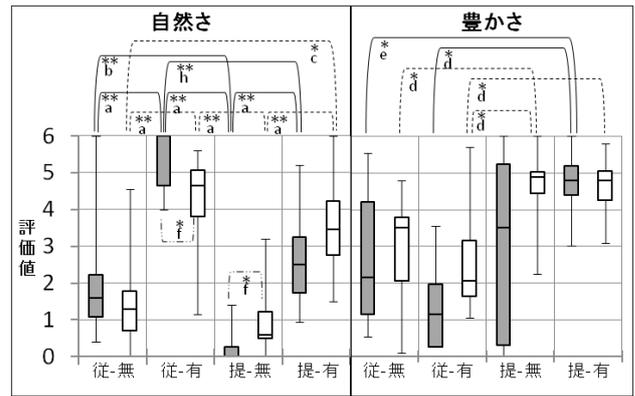


図 4.4 “Grieg” の実験結果 **:p<0.01 *:p<0.05

4.2.2 アンケートの結果と考察

自然さの評価基準を問うアンケートに対する回答のうち、回答者数が多かった上位 3 つまでの回答を表 4.2 に示す。音楽経験者は「テンポのゆらぎ」と回答した被験者が最も多く、未経験者は「曲全体のテンポ」と回答した被験者が最も多かった。このことから、自然さの評価基準は音楽経験の有無によって異なることが示唆された。「アレンジとしてテンポを揺らがせているのか、単に不自然なのかの違いがわからない」という回答も得られていることから、音楽未経験者はこの点からテンポのゆらぎを評価基準にしなかったと考えられる。被験者全体では、多くの被験者がテンポに関する回答をした。テンポ記号の類似性評価の導入が自然さの評価を高くしたことを踏まえて考察すると、テンポ記号の類似性評価の導入は、テンポのゆらぎを安定させ、対象曲に適切な曲全体のテンポにさせる効果があると考えられる。

表 4.2 自然さの評価基準のアンケート結果

経験者	未経験者	全体
テンポのゆらぎ(5名)	曲全体のテンポ(5名)	曲全体のテンポ(9名)
曲全体のテンポ(4名)	自然か自然でないか(2名)	テンポのゆらぎ(6名)
発音タイミングのずれ(4名)	テンポのゆらぎ(1名)	発音タイミングのずれ(5名)

豊かさの評価基準を問うアンケートに対する回答のうち、回答者数が多かった上位 3 つまでの回答を表 4.3 に示す。音楽経験者、未経験者で回答に大きな差は見られなかった。このことから、豊かさの評価基準は音楽経験の有無には関係ないことが示唆された。また、被験者全体では、多くの被験者がテンポのゆらぎと強弱に関する回答をした。提案手法の導入が豊かさの評価を高くしたことを踏まえて考察すると、提案手法の導入はテンポや強弱のゆらぎを増幅させる効果があると考えられる。

表 4.3 豊かさの評価基準のアンケート結果

経験者	未経験者	全体
テンポのゆらぎ(6名)	テンポのゆらぎ(8名)	テンポのゆらぎ(14名)
曲全体の強弱(5名)	曲全体の強弱(4名)	曲全体の強弱(9名)
強弱のゆらぎ(4名)	強弱のゆらぎ(2名)	強弱のゆらぎ(6名)

実験において難しいと感じた点を問うアンケートに対する回答のうち、回答数が多かった上位 3 つまでの回答を表 4.4 に示す。音楽経験者、未経験者で回答に大きな差は見られなかった。また、被験者全体では、多くの被験者が豊

かさの評価が難しいと回答をした。その理由を記述した被験者の回答を図 4.5 に示す。総括すると、不自然な演奏の豊かさをどのように評価すべきかがわからないため豊かさの評価が難しくなると考えられる。これは、普段不自然な演奏を耳にすることが無いためと、本人が思う「豊かさ」と実験における「豊かさ」に齟齬があることが起因していると考えられる。これは、図 4.2 と図 4.4 の音楽経験者の結果のうち、自然さの評価が極端に低い演奏における豊かさの評価が、分散していることから推測できる。

表 4.4 実験において難しいと感じた点のアンケート結果

経験者	未経験者	全体
豊かさの評価(8名)	豊かさの評価(4名)	豊かさの評価(12名)
ショパンの曲(3名)	ショパンの曲(2名)	ショパンの曲(5名)
自然さの評価(2名)	自然さの評価(2名)	自然さの評価(4名)

- ・下手な演奏でも定義上豊かにしなくてはならないため
- ・自然ではないと感じた上で豊かさがどの程度なのかがわからない
- ・テンポが崩れていると評価しにくい

図 4.5 豊かさの評価が難しい理由(抜粋)

4.3 実験のまとめ

表 4.5 に従来手法との有意差検定の結果をまとめる。従来手法にテンポ記号の類似性評価を導入しただけでは、“Beethoven” と “Grieg” の楽曲の自然さを高めるだけであった。しかし、本稿で提案した手法の導入によって “Chopin” の楽曲も自然さを高め、全ての楽曲の豊かさを高めることができ、従来手法よりもより人間らしい演奏の生成が可能となったと言える。しかし “Grieg” の楽曲については自然さを低めてしまった。テンポ記号の類似性評価が自然さの評価を高めることが示唆されたことから、“Grieg” の楽曲に対しては提案手法よりもテンポ記号の類似性評価を重視して導入すべきであると考えられる。従って、楽曲ごとにテンポ記号の類似性評価と提案手法に適切な重みを付加する必要があると考えられる。

表 4.5 従来手法との有意差検定結果のまとめ

		Beethoven		Chopin		Grieg	
		経	未	経	未	経	未
テンポ 評価導入	自然さ	<<	<<			<<	<<
	豊かさ						
テンポ 評価と 提案手法導入	自然さ	<<	<<		<		<
	豊かさ	<	<	<<	<	<	<

<<:p<0.01 で従来手法より評価向上 <:p<0.05 で従来手法より評価向上
 空欄:従来手法と有意差なし 経:音楽経験者 未:音楽未経験者

テンポ記号の類似性評価の導入によって自然さを高めることができた。一方で、アンケート結果から、テンポのゆらぎや曲全体のテンポを中心に自然さを評価することが示唆された。このことからテンポ記号の類似性評価の導入は、テンポのゆらぎを安定させ、対象曲に適切な曲全体のテンポにさせる効果があると考えられる。また、提案手法の導入によって豊かさを高めることができた。一方で、アンケート結果から、テンポや強弱のゆらぎを中心に豊かさを評価することが示唆された。このことから、提案手法の導入はテンポや強弱のゆらぎを増幅させる効果があると考えられる。

5. おわりに

本稿では、これまでに Kagurame Phase-III で用いていた旋律類似性評価手法を改善させるため、音高方向グラデーションにより音高が異なる音符を認識するピアノロール画像の提案とその比較手法の提案を行った。特に、複数の音符と相関をとった場合の対処手法の提案を行った。そして、その有用性を受聴実験により検証した。

その結果、本稿で提案した手法は、テンポ記号の類似性評価手法と組み合わせることにより、従来手法よりも自然で豊かな人間らしい演奏の生成を可能とすることが明らかとなった。しかし、楽曲によっては演奏の自然さを低めてしまう場合があることも明らかとなった。

今後は、テンポ記号の類似性評価と提案手法に重みを付加し、楽曲ごとに適切な重みを算出する手法の提案が必要である。

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) 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) 寺村佳子, 前田新一: 統計的学習によるテンポの変動を考慮したピアノ演奏模写, 情報処理学会報告, Vol.2010-MUS-84, No.12, pp.1-6, (2010)
- 5) 橋田光代, 北原哲朗, 鈴木健嗣, 片寄晴弘, 平田圭二: 演奏表情付けコンテスト SMC-Rencon 開催報告, 情報処理学会研究報告, Vol.2011-MUS-92, No.4, pp.1-6, (2011)
- 6) 日野達也, 鈴木泰山, 野池賢二, 徳永幸生, 杉山精: 聴感による旋律の演奏表情の類似性評価-ピアノロール画像を用いた旋律類似性評価結果との比較, 情報処理学会研究報告, 2011-MUS-89, (2011)
- 7) 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)
- 8) 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試聴評価, 情報処理学会研究報告, 2005-MUS-59, Vol.2005, No.14, pp.43-48, (2005)
- 9) 日野達也, 鈴木泰山, 野池賢二, 徳永幸生, 杉山精: ピアノロール画像の比較による旋律類似性評価手法の検討-音高の違いを考慮した旋律概形の比較-, 情報処理学会研究報告, Vol.2010-MUS-86, No.17, pp.1-6, (2010)
- 10) 柴崎正浩, 鈴木泰山, 米村俊一: ピアノロール画像の比較による旋律類似性評価手法の検討-若干の音高の違いを許容する旋律概形の比較評価-, 情報処理学会研究報告, Vol.2013-MUS-100, No.12, pp.1-6, (2013)
- 11) 柴崎正浩, 鈴木泰山, 徳永幸生, 杉山精: 聴感による演奏表情の評価特性の分析, 第 74 回全国大会講演論文集, 2012(1), pp.363-365, (2012)
- 12) 柴崎正浩, 鈴木泰山, 徳永幸生, 杉山精: テンポの類似性を取り入れた事例に基づく演奏表情の生成(II)-試聴実験によるその効果の検証-, 情報処理学会研究報告, Vol.2012-MUS-97, No.8, pp.1-6, (2012)
- 13) 大蔵康義著: 人は音・音楽をどのように聴いているのか-統計による実証と楽曲リスト-, 国書刊行会出版, (2010)