

事例に基づく演奏表情の生成

鈴木泰山[†] 徳永健伸[†] 田中穂積[†]

本論文では、事例に基づいて演奏表情を生成する手法を提案し、この手法を用いたシステムによる実験について報告する。人間による音楽演奏は、テンポや音の強さの揺らぎなどの演奏表情と呼ばれる要素を含んでいる。演奏表情の与え方は自由度が非常に高く、演奏者の経験や演奏スタイル、演奏時の気分などの演奏状況に応じて多様に変化する。しかしながら、従来の演奏表情の自動生成に関する研究では演奏状況を殆んど考慮していないため、演奏状況に応じた演奏表情の生成をすることができなかった。我々は演奏状況に応じた多様な演奏表情を生成するために、事例に基づく演奏表情の生成を提案する。この手法は過去の演奏データの中から、演奏する楽曲に類似した楽曲を探し、その楽曲に付与された演奏表情を利用して対象曲に演奏表情を付与する。我々はこの手法を利用した演奏表情生成システムを実装し、演奏状況に応じた演奏表情の生成を試みた。その結果、この手法が多様な演奏表情の生成に有効であることを確認した。

A Case-based Approach to the Generation of Musical Expression

TAIZAN SUZUKI,[†] TAKENOBU TOKUNAGA[†] and HOZUMI TANAKA[†]

This paper presents a case-based formulation of computational musical expression generation, and reports the results of experiments on the derived musical expression generation system. The majority of musical performance has musical expression, including such items as fluctuation in tempo and volume. Generally, musical expression is freely attributed by the performer, and is affected by various conditions surrounding performance (performance conditions), such as the performer's taste, performative style, mood, and so forth. However, past research on the automatic generation of musical expression has downplayed performance conditions. In this paper, we propose a case-based approach to the automatic generation of musical expression. This method easily enables the generation of musical expression reflecting the performance conditions. We have implemented the proposed method in a performance data generation system, and generated some expressive performance data from this system. According to our evaluation of generated performance data, our method was shown to be effective at generating musical expression customized to performance conditions.

1. はじめに

人間の音楽演奏は楽譜通りに厳密に行なわれるものではなく、テンポや音の強さなどの微妙な揺らぎや変化を与えられている。このようなテンポや音の強さの揺らぎや変化を演奏表情と呼ぶ。演奏表情の大部分は、演奏中に様々な情感や曲想を表現することを目的として与えられるものであり、演奏を自然で味わい深いものにする上で必要不可欠な要素である。しかしながら、楽譜には演奏表情に関する情報が殆んど記載されていないため、演奏表情は演奏者が直感や経験などに基づいて与えてやる必要がある。

近年、計算機を利用した音楽演奏が普及しつつあるが、こうしたシステムの大半は与えられた演奏データを単純に再生する能力しか持っておらず、演奏表情の生成は演奏データの作成者に委ねられている。演奏表情を含む演奏データの作成は多くの経験と労力を必要とするため、演奏表情を含む演奏データはあまり多くない。従って計算機による音楽演奏は演奏表情を伴わないことが多い。このような理由から計算機による音楽演奏はざこちなく面白みに欠けるという評価が一般的である。

この問題を解決するため、与えられた楽譜情報に対し、演奏表情を自動生成するための手法が数多く提案されている¹⁾。主な手法は、人間が与えた演奏表情の生成規則を利用して演奏表情を生成する手法や、様々な学習アルゴリズムを利用して人間の演奏データから演奏表情の生成規則を学習し、それに基づいて未知の

[†] 東京工業大学 情報理工学専攻
Department of Computer Science, Graduate School of
Information Science and Engineering, Tokyo Institute
of Technology

曲に対する表情付けを行なう手法^{2)~5)}などである。いずれも、なんらかの方法で一般的な演奏表情生成規則を作成し、演奏曲にその規則を適用することで演奏表情の生成を行なうので、規則に基づく手法と呼ぶことができる。

演奏表情は一般に自由度が高く、許容される範囲が広いという特徴がある。事実、演奏表情の与え方は、演奏者や時代、スタイル、演奏時の意図や気分などに応じて多様に変化する。ここではこれらの演奏表情に影響を与える楽曲外の要因をまとめて演奏状況と呼ぶ。同じ曲でも演奏状況に応じて全く異なる演奏表情を持った演奏になる。従って、利用者の趣向に応じた演奏表情の生成を行なうような自動演奏システムなどの構築を視野に入れた場合、楽曲の特徴や旋律情報だけでなく演奏状況も考慮できるような演奏表情生成手法が望まれる。しかしながら、演奏表情と演奏状況の関係を考慮した演奏表情の生成手法に関する研究は十分に行なわれているとは言い難い。

この一因として、殆どどの手法が規則に基づく手法を用いている点あげられる。規則に基づく手法では、演奏表情生成規則に基づいて演奏表情を生成するので、演奏状況に応じた演奏表情の生成を行なうためには演奏表情生成規則に演奏状況と演奏表情との関係を記述する必要がある。しかしながら演奏状況は多様で連続的に変化するため、演奏状況の変化が演奏表情に与える影響を演奏表情生成規則に取り込もうとすると、規則のパラメータや規則数が膨大な量になり現実的ではない。

こうした理由から、演奏状況を考慮した演奏表情の生成を行なうためには、演奏表情生成規則のみに頼った手法は適切ではない。そこで、我々はこれらの問題を解決する手法として事例に基づく演奏表情の生成手法⁶⁾を提案する。事例に基づく手法は問題解決の手法の一種で、問題解決に関する知識として同じ領域の問題の具体的な解決例を利用するというものである。事例に基づく手法の代表として事例に基づく推論(Case-based reasoning)⁷⁾が挙げられる。事例に基づく手法では、問題のドメインに関する知識をルールなどの形に抽象化することは行なわず、与えられた問題ごとに必要な知識を同様の解決例の中から見つけたし応用する。従ってルールに一般化してしまうと失われてしまうような希少な状況を考慮することが可能な点や、学習にかかるコストが比較的小さい点などの利点を持つ。

事例に基づく手法を用いた演奏表情生成は、Arcosらによって事例に基づく推論を利用した演奏表情の生

成⁸⁾が提案されている程度で、規則に基づく手法に比べて十分には研究が行なわれておらず、また演奏状況の導入についての検討も行なわれていない。

本論文で提案する事例に基づく演奏表情の生成手法は、演奏表情の生成過程において演奏状況などの要素を考慮することが容易であり、これにより演奏状況に応じて多様な演奏表情を生成することが可能である。本論文では我々の提案する事例に基づく手法について説明するとともに、この手法の要素技術についても述べる。また、提案手法を実装した演奏表情生成システムの概要と、このシステムを利用した演奏表情の生成実験についても述べる。

2. 事例に基づく演奏表情生成

2.1 概要

我々の提案する事例に基づく演奏表情の生成手法の基本構成を図1に示す。我々の提案する手法では演奏表情の知識として演奏データ集を利用する。演奏データ集は人間が演奏したデータを集めたものである。この中の個々のデータは1)演奏曲情報、2)人間の演奏データ、3)演奏時の演奏状況からなる。演奏曲情報は主に演奏曲の楽譜情報からなる。演奏データは人間の演奏を記録したものである。演奏状況はその演奏データが演奏された際の演奏状況を示すものである。

演奏表情の生成の際には、1)対象曲情報と、2)演奏状況の設定、を入力として与える。対象曲情報は主に楽譜情報からなる。演奏状況は先に述べたように、生成される演奏表情の特徴や雰囲気や指定するパラメータである。演奏状況の詳細については3章で述べる。

入力が与えられると、まず最初に演奏データ集から対象曲に類似した楽曲を検索する。類似した楽曲の検索は、対象曲全体だけではなく対象曲を構成するフレーズや小節などの旋律断片に対しても行なう。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。これが参考事例集である。参考事例の検索については4章で述べる。

次に、参考事例集の中の全ての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その事例がどの程度参考になるかを表すスコアである。事例の重要度は「対象曲と事例との旋律断片の類似性」と「入力された演奏状況と事例の演奏状況の類似性」の2つの要素から決定する。いずれも入力と参考事例とが類似しているほど重要度は高くなる。旋律断片の事例の重要度は、対象となる断片同士の類似性だけではなくそれぞれの断片の前後に連続する断片同士の類似性なども考慮する。旋律断片の類似性の評

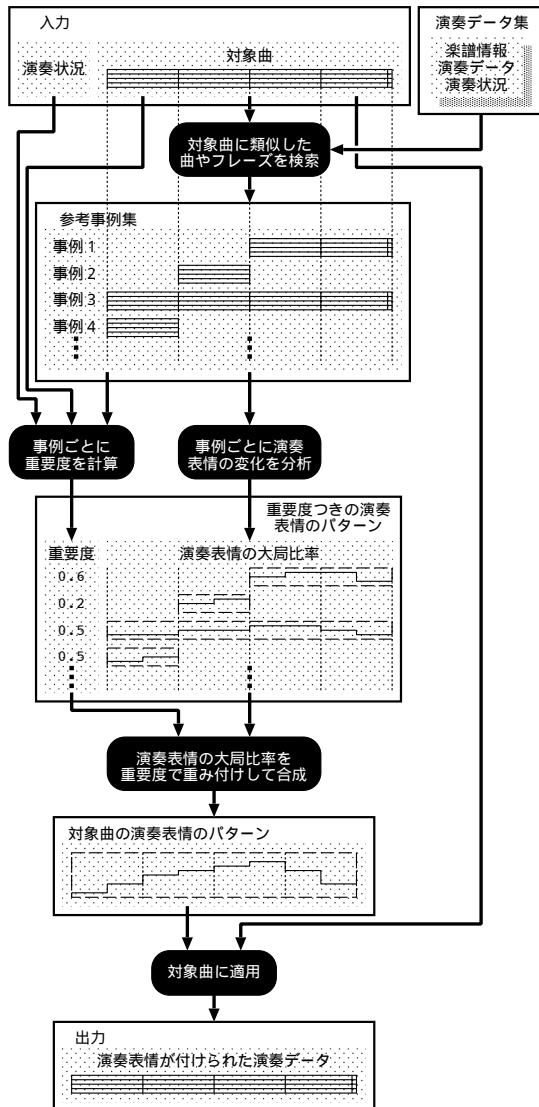


図 1 事例に基づく演奏表情の生成手法の基本構成

Fig. 1 An overview of our example-based musical expression generation methodology

価は参考事例の検索に用いた手法と同じものを用いると効率的である。演奏状況の類似性については 3.2 節で述べる。また、旋律断片の類似性の類似性については 4.3 節でその一例を示す。

また、重要度の評価と同時に全ての参考事例に対して演奏表情の分析を行なう。本手法では演奏表情の計算の際に、テンポや音の強さの絶対的な数値ではなく、これを相対的な変化量の比率に変換したものを利用している。この相対的な変化量を演奏表情の大局比率と呼ぶ。この段階では演奏表情の大局比率の計算を行なう。演奏表情の大局比率の分析は 5.1 節で述べる。

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大局比率の分析を行なうと、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。この集合の中の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。対象曲の演奏表情の大局比率の合成については 5.2 節で述べる。

最後に、この対象曲の演奏表情を対象曲の楽譜に適用し演奏表情の与えられた対象曲の演奏データを作成する。

2.2 特徴

本手法の最大の特徴は、演奏状況に応じた多様な演奏表情の生成が可能なことである。本手法では、重要度の評価の際に対象曲と参考事例との楽曲の類似性だけではなく入力された演奏状況と参考事例の演奏状況の類似性も考慮している。従って同じ楽曲が入力された場合でも演奏状況の入力によって事例ごとの重要度が変化する。演奏状況の近い事例により高い重要度が与えられることから、最終的に生成される演奏表情は、演奏状況の近い事例に見られる演奏表情の特徴をより強く反映したものになる。これによって、演奏状況の設定に応じた多様な演奏表情の生成が可能となる。

また、本手法は楽曲構造に関する認知的な知識を必要としないという特徴をもつ。演奏表情と楽曲の認知構造には密接な関係があり、類似した認知構造を持つ旋律の演奏表情は類似することが指摘されている。このことから、先行研究には、楽譜情報から直接的に演奏表情に変換するのではなく、楽譜情報から認知構造情報を取り出し、この認知構造情報を介して間接的に演奏表情の生成を行なっているものが多い。これらの手法では、旋律と認知構造との関係と、認知構造と演奏表情との関係の二種類の知識が必要となる。一方で、本手法では前提として、類似した認知構造を持つ旋律の演奏表情は類似するという点に加えて、類似した旋律の認知構造は類似するという点も暗に仮定している。つまり、これらを一つにまとめて、類似した旋律の演奏表情は類似するという前提としている。従って、生成の過程で事例集から獲得している知識は楽譜情報から演奏表情への直接的な関係である。これにより、本手法では旋律と認知構造の関係などの認知構造に関わる知識を必要としないという特長を持っている。

さらに、本手法では演奏データ集からルールなどの知識情報への変換を行なう必要がないため、新しい知識(演奏データ)の追加の際に生じるコストが非常に小さいという利点がある。

一方で、ルールなどの一般化した知識を利用する手法に比べ、本手法は演奏表情の生成のメカニズムが不透明なため、人間の演奏表情の生成の分析などを目的とした分析的研究には適さないなどの欠点もある。

本手法は事例に基づく手法であるという点で Arcos らによる事例に基づく推論を用いた演奏表情生成⁸⁾に比較的近い。といっても Arcos らの手法と本手法とで共通しているのは演奏表情生成の手続きの順序くらいで、相違点は多岐に渡る。顕著な相違点として、事例検索の単位、事例の応用方法、演奏状況の考慮などがあげられる。Arcos らの手法では、類似事例の検索などの処理を全て一音単位で行なっている。参考事例として、符長や音高、認知構造上の特徴などが類似している音符を検索し、その一音に与えられている演奏表情を利用するというものである。一方で、本手法では基本的に全ての処理は音符ではなくフレーズや小節などの旋律断片に対して行なっている。従って、本手法は Arcos らの手法に比べ、旋律の中での演奏表情の流れをよりの確に取得できると考えられる。本手法ではこの旋律単位での処理を実現するため、演奏表情生成のそれぞれの段階の手続きが Arcos らの手法と全く違ったものになっている。また、Arcos らの手法ではそれぞれの音符に対して最も類似している事例を一つだけ参考にしているが、本手法ではそれぞれの旋律断片について複数の参考事例を重要度で重み付けし組み合わせて利用している。

3. 演奏状況

演奏状況に応じた演奏表情の生成を行なうためには、演奏状況の類似性を評価する方法が必要である。ここでは本手法で利用している演奏状況の記述方法、および演奏状況の類似性の評価方法について述べる。

3.1 演奏状況の記述

先に述べたように演奏表情は、演奏者や演奏のスタイル、演奏者の演奏時の演奏意図や気分、演奏時の疲労度、前後に演奏した楽曲など、様々な要因によって大きく変化する。これらの演奏表情に影響を与える楽曲外の要因が演奏状況である。

演奏者に関する演奏状況は、基本的に演奏者が誰であるかによって表すことができる。演奏スタイルは「ロマン派風」や「バロック風」のようなジャンルの分類で表される。演奏時の演奏意図や気分は、主に「明るく」や「はずむように」といった曲想に関する要因の組合せである。こういった要因の中には、程度によって演奏表情に与える影響が変化するものもある。例えば「明るい」演奏にはただ明るい演奏だけ

ではなく「非常に明るい」演奏や「やや明るい」演奏などの程度の違いがあり、この程度の大小も演奏表情を決定する要因の一つである。

本手法では演奏状況を素性の集合で表す。演奏状況の素性とは上記のような演奏表情を決定する個々の要因である。個々の素性にはその要因の程度を表す値として -1 から 1 の間の数値を付与する。この数値を程度量と呼ぶ。例えば「A 氏による明るくはずむような演奏」の演奏データの演奏状況は、「A 氏が演奏」、「明るい」、「はずむような」の 3 つの素性を持つ。「明るい」、「はずむような」はいずれも際立っているわけではないので、程度量は正の値の中間的な数の 0.5 を与える。「A 氏が演奏」に関しては A 氏が演奏していることが明らかであれば程度量は 1 に近い値を与える。この演奏データが、ただ「明るい」ではなく「非常に明るい」演奏の場合は、「明るい」の程度量を 1 に近い値にし、逆に「若干明るい」程度であれば程度量を 0 に近づける。また「明るくない」といった否定的な素性には負の程度量を与える。

3.2 演奏状況の類似性の評価

我々の提案する演奏表情生成手法では参考事例の重要度を計算する際に入力された演奏状況と演奏データの演奏状況の類似性を評価する必要がある。そのために演奏状況ベクトルおよび演奏状況空間という概念を定義する。まず、前節で述べた演奏状況の素性と程度量を、それぞれ単位ベクトル、およびノルムと考える。そうすると、それぞれの演奏状況はベクトルの和で表すことができる。この演奏状況を表すベクトル和を演奏状況ベクトルと呼ぶ。式 (1) に演奏状況ベクトル v を示す。

$$v = \sum_{i \in V} a_i \cdot e_i \quad (1)$$

$$i = \{a_i, e_i\}$$

V は演奏状況に含まれる素性の集合、 a_i は素性 i の程度量、 e_i は素性 i の単位ベクトルを表す。

次に、ある演奏データ集に含まれるデータの演奏状況に含まれる素性の総数（異なり数）が n 個だったと仮定する。この n 個の素性の集合によって張られる n 次元空間を演奏状況空間と定義する。それぞれの演奏データの演奏状況の素性は演奏状況空間を張る素性の集合の部分集合であるから、すべての演奏状況ベクトルは演奏状況空間上のベクトルとなる。例えば「やや明るい演奏」、「明るくはずむような演奏」、「非常に重たく冷たい演奏」の 3 つの演奏データからなる演奏データ集の場合、演奏状況空間は「明るい」、「はずむような」、「重たい」、「冷たい」の 4 つの素性によって

張られる4次元空間である。

それぞれの素性の程度量はその素性が顕著であればあるほど絶対値が大きくなるので、特徴的な演奏は演奏状況空間上でのベクトルのノルムが大きくなり、また標準的な演奏は逆にノルムが小さくなる。

演奏状況が類似している場合、演奏状況に含まれる素性やそれぞれの素性が持つ程度量の比率も近くなるので、演奏状況空間での演奏状況ベクトルの向きや大きさが近くなる。したがって、比較する演奏状況ベクトルどうしの演奏状況空間上での向きやノルムの大きさの近さから演奏状況の類似性を評価することができる。式(2)に演奏状況ベクトル v と u の類似性 $R(v, u)$ の定義を示す。

$$R(v, u) = \frac{v \cdot u}{\max\{|v|^2, |u|^2\}} \quad (2)$$

この式の分子は演奏状況ベクトルの内積である。内積をそのまま使用すると演奏状況ベクトルの大きさによる違いが考慮できないので、大きい方の演奏状況ベクトルのノルムの2乗で割って正規化する。例えば、内積をそのまま利用すると、非常に明るい演奏 ($1 \cdot e_{bright}$) とやや明るい演奏 ($0.25 \cdot e_{bright}$) の間の類似性と、明るい演奏 ($0.5 \cdot e_{bright}$) とうしの類似性が共に0.25になるが、式(2)を用いるとそれぞれが0.25と1になり、ベクトルの大きさが同じである後の方が高い値になる。

この定義ですべての素性を互いに独立な単位ベクトルとみなしているが、厳密にはそれぞれの素性は一次独立ではなく依存関係にある。例えば、明るい演奏ははずむような演奏であることが多いし、逆に明るい演奏が同時に重たい演奏であることはまれである。しかしながら、こういった素性間の関連性を機械的に評価するのは困難なので、個々の素性は一次独立であると仮定している。

演奏表情生成では、入力した演奏状況の素性が演奏データ集の中に全く含まれていないと、演奏状況に応じた演奏表情が生成できない。よって、演奏表情生成では入力および演奏データで使用する素性をあらかじめ決める。

4. 類似旋律の検索

4.1 事例検索の効率化

一般に事例に基づく手法の問題として参考事例の不足がある。事例に基づく手法ではデータ集の中から対象に適用可能な事例を検索し、それらを分析して対象に適用する。したがって検索で得られる参考事例が少ないと有効な結果が得られない。

本手法において最も理想的な参考事例は対象曲そのものの演奏データである。しかしながら、演奏データ集の中にこのようなデータが含まれることは期待できないし、また曲全体が似ているような楽曲が含まれている可能性も稀である。つまり曲全体を比較して類似している曲を探すだけでは十分な量の参考事例が得られない可能性が高い。

本手法ではこの問題を解決するために、対象曲や演奏データ集のデータを様々な長さの旋律断片に分割し、旋律断片を対象として参考事例の検索を行なう。対象曲や演奏データ集を旋律断片に分割すると、類似した断片が見つかる可能性と利用可能な参考事例の総量が共に増加するため、有効な参考事例が得られる可能性が向上する。

理想的には可能なすべての分割を行なうのが望ましいが、ひとつの曲に対する分割の境界の引き方は音符の数のべき乗になるため計算量の観点から現実的ではない。かといってランダムに分割したのでは有効な事例をとり逃がす可能性があるため、ある程度一貫性のある基準に従って分割する必要がある。そこで本手法では楽曲の分割の基準として楽曲の構成情報(楽曲構成)を利用する。ここで言う楽曲構成は、楽曲の中に見られる、小節、フレーズ、繰り返し、などの部品の階層的構造である。楽曲構造と呼ばれることもあるが、認知的楽曲構造とは違うものである。一般に楽曲構成は拍節構造を内包していることが多いため、楽曲構成に基づいて区切られた旋律断片どうしは拍節的にも類似した特徴を持ちやすく、類似した参考事例の検索に適している。

本手法では、事例の検索を行なう前に、対象曲を楽曲構成に基づいて半小節や小節、フレーズ、モチーフなど異なる数種類の大きさの単位の旋律断片に分割する。大きさの単位の異なる旋律断片を区別するために、大きさの単位が同じ分割の旋律断片の集合を階層と呼ぶ(図2)。最大の分割を第1階層、 i 番目に大きい単位で分割した旋律断片の集合を第 i 階層と呼ぶ。第1階層の旋律断片の大きさは曲全体で、最下層の旋律断片の大きさは音符である。 $i-1$ 階層の境界は i 階層の境界の部分集合であるものとする。また、 i 階層の旋律断片 s_i を包含している $i-1$ 階層の断片を s_i の親と呼ぶ。

4.2 大局的な演奏表情の取得

旋律断片に分割して事例を検索すれば下位階層の参考事例を増やすことができるが、上位階層の旋律断片については参考事例の不足という問題を解決できない。下位階層の断片に対する事例だけでは、連続する旋律

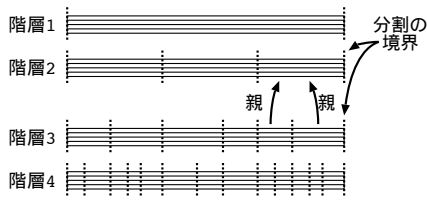


図 2 旋律断片の境界と階層

Fig. 2 Boundaries and hierarchy of phrase segments



図 3 旋律の類似性の評価例

Fig. 3 An example of phrase similarity evaluation

断片間の演奏表情のバランスなどの大局的な演奏表情を獲得することができないので、より上位の階層の旋律断片や曲全体に対する参考事例も必要である。

そこで、上位の階層の旋律断片に対する事例を得るために、参考事例を検索する際に旋律の局所的な部分に対する子細な比較は行なわず、旋律断片の全体的な流れに関する特徴のみから類似性の評価を行なう。ただし、このような方法で得られた事例は、局所的な部分の類似性が保証されていないので、これらの参考事例からは大局的な演奏表情のバランスや変化だけを獲得する。この大局的な演奏表情のみの獲得を可能にするのが演奏表情の大局比率による記述である。演奏表情の大局比率については 5.1 節で述べる。

4.3 旋律の類似性の評価

4.2 節で述べたように、本手法では参考事例を検索する際に、旋律断片の大局的な類似性を評価する手法が必要である。旋律の類似性の評価手法は楽曲検索^{9),10)}の分野などを中心に様々な提案がなされているが、大局的な類似性に重点を置いた比較方法はあまり提案されていない。本論文は演奏表情の類似性に関するものではないが、ここでは旋律の大局的な類似性の比較方法の一例として、実験システム(6章)で用いた演奏表情の評価手法を譜例(図3)を用いて説明する。

この手法では、旋律断片の大局的な特徴量の違いから旋律断片の類似性を評価する。類似性の評価に使用する特徴情報は、1) 主旋律の特徴、2) 音階(スケール)の特徴、3) リズムの特徴、の3種類である。数値

はいずれも小さいほど類似していることを表す。

主旋律の特徴は、旋律断片の前半および後半の音符の平均音高の変化量である。この数値は主に旋律が上昇系であるか下降系であるかを表す。式(3)に旋律断片 s の主旋律の特徴量 $C_m(s)$ の定義を示す。 N_f が旋律の前半の音符の集合、 N_l が旋律の後半の音符の集合、 $p(n)$ は音符 n の音高を表す。主旋律の特徴の類似性は、比較する二つの旋律断片の主旋律の特徴量の差の絶対値である(式(4))。

$$C_m(s) = \frac{\sum_{n \in N_l} p(n)}{|N_l|} - \frac{\sum_{n \in N_f} p(n)}{|N_f|} \quad (3)$$

$$D_m(s_1, s_2) = |C_m(s_1) - C_m(s_2)| \quad (4)$$

図3の譜例の場合、いずれも一小節からなるので小節の前半と後半の平均音高の差とする。旋律1の最初のC音の音高を60とすると、旋律1(s_1)の前半の音高は(76, 79, 77, 79, 76, 77, 74, 76)、後半の音高は(72, 84, 83, 84, 81, 83, 79, 81)となる。従って音高の特徴量は式(5)に示すように4.125である。同様に旋律2(s_2)は2.625である(式(6))。よって主旋律の特徴量の類似性は、式(7)に示す通り1.5となる。

$$C_m(s_1) = \frac{647}{8} - \frac{614}{8} = 4.125 \quad (5)$$

$$C_m(s_2) = \frac{610}{8} - \frac{589}{8} = 2.625 \quad (6)$$

$$D_m(s_1, s_2) = |4.125 - 2.625| = 1.5 \quad (7)$$

音階の特徴は旋律断片の中の音符を階名で表した時のそれぞれの音の分布である。ここでの階名とは、同じオクターブ内の主音との音高の差に基づくラベルである。なお、主音は楽曲の調に基づいて決定する。1オクターブは12音であるため、この特徴量は12の要素を持つベクトルである。式(8)と式(9)に音階の特徴量ベクトルの計算式を示す。 N は旋律断片 s に含まれる音符の集合である。 $q(n)$ は n の階名である。この特徴量はスケールなどのような音の分布の特徴を表す。音階の特徴の類似性は、特徴量ベクトルの音階ごとの数値の差の総計である(式(10))。

$$C_h = (c_{h,0} \ c_{h,1} \ \dots \ c_{h,11}) \quad (8)$$

$$c_{h,i}(s) = \frac{|\{n | n \in N, (q(n) = i)\}|}{|N|} \quad (9)$$

$$D_h(s_1, s_2) = \sum_{0 \leq i \leq 11} |c_{h,i}(s_1) - c_{h,i}(s_2)| \quad (10)$$

図3の旋律1は八長調なので主音の音高は72, 84などである。これに基づいて各音の音高を階名で表記すると(III, V, IV, V, III, IV, II, III, I, I, VII, I, VI, VII, V, VI)となる。旋律2の場合、変口長調なので主音は70, 82などとなる。それぞれの旋律の音階の分布

および特徴量は表1に示す通りである。いずれの旋律も音符の総数が16なので、特徴量は分布の数値を16で割ったものである。音階の類似性は音階ごとの差の総和なので0.375である。

リズムの特徴は旋律断片の長さを拍単位で表したものである。あまりにも長さの異なる旋律断片は、主旋律や音階の特徴が類似していたとしても有効な事例とは言いがたいことから、旋律断片の類似性に旋律断片の長さを考慮する。リズムの特徴の類似性は、長い方の断片の拍数を短い方の断片の拍数で割ったものである(式(11))。\$l_s\$は旋律断片\$s\$の拍数である。

$$D_r(s_1, s_2) = \begin{cases} l(s_1)/l(s_2) & l(s_1) \geq l(s_2) \\ l(s_2)/l(s_1) & l(s_2) > l(s_1) \end{cases} \quad (11)$$

図3の旋律はいずれも四分の四拍子で長さが一小節なので拍数は4である。よって、類似性は1となる。実験システムではこれらの特徴量の類似性を式(12)に示す式で正規化し合計したものを旋律断片の類似性としている。図3の2旋律の場合、類似性は3.75となる(式(13))。この数値は類似性が高いほど値が小さくなる。事例の重要度は類似性が高いものの数値が高くなる必要があるため、符合を反転し指数関数に適用したものを\$e^{-D(s_1, s_2)}\$を重要度として利用している。

$$D = D_m + 6 \cdot D_h + (D_r - 1) \quad (12)$$

$$D(s_1, s_2) = 1.5 + 6 \cdot 0.375 + 0 = 3.75 \quad (13)$$

本論文は演奏表情の類似性について深く言及するものではないため、この評価手法はあくまで一例として単純な手法を示したに過ぎない。この手法を改善することにより、演奏表情生成システムの能力が向上する可能性は大きいと考えられる。

表1 評価例の旋律の音階の分布と特徴量ベクトル
Table 1 Scale distribution and characteristic vector of each example

	分布		特徴量		
	\$s_1\$	\$s_2\$	\$s_1\$	\$s_2\$	差
I	3	3	0.188	0.188	0.000
#I	0	0	0.000	0.000	0.000
II	1	3	0.063	0.188	0.125
#II	0	0	0.000	0.000	0.000
III	3	3	0.188	0.188	0.000
IV	2	3	0.125	0.188	0.063
#IV	0	0	0.000	0.000	0.000
V	3	2	0.188	0.125	0.063
#V	0	0	0.000	0.000	0.000
VI	2	1	0.125	0.063	0.063
#VI	0	0	0.000	0.000	0.000
VII	2	1	0.125	0.063	0.063
合計	16	16	1.000	1.000	0.375

5. 演奏表情の大局比率

5.1 演奏表情の大局比率の分析

演奏事例集から演奏表情を獲得し対象曲に適用するためにはテンポや音の強さなどの演奏表情の数値(演奏表情値)の変化量を表現し、記録しなければならない。従来、演奏表情の記述には、変化量を曲線近似して記述する方法¹¹⁾や、音符ごとの演奏表情の数値を記録する手法などが用いられている。前者は、旋律が長くなればある程度の種類の曲線をあらかじめ用意する必要があることや、近似によって微妙な演奏表情が失われるなどの問題がある。また、後者は4.2節で述べた大局的な演奏表情の取得や記述が困難である。

そこで、本手法では演奏表情値の変化を表すために演奏表情の大局比率を用いた表記法を提案する。演奏表情の大局比率は、注目する旋律断片での局所的な演奏表情値を、その親の断片の演奏表情値の平均値との比率で表したものである。旋律断片\$s_{i,j}\$の演奏表情の大局比率\$P_{exp}\$を式(14)に示す。

$$P_{exp}(s_{i,j}) = \frac{\exp(s_{i,j})}{\exp(s_{i-1,k})} \quad (14)$$

$$\{s_{i,j} \subset s_{i-1,k}\}$$

\$s_{i,j}\$は\$i\$階層の\$j\$番目の旋律断片である。\$s_{i-1,k}\$は\$s_{i,j}\$の親である。\$\exp(s)\$は断片\$s\$の演奏表情(テンポや音の強さなど)の平均値である。

例として図4に4小節からなるフレーズの小節ごとのテンポについての演奏表情の大局比率の計算例を示す。このフレーズは全体の平均テンポが120(0.5秒/四分音符)で、それぞれの小節の平均のテンポが順番に115(0.52秒/四分音符)、133(0.45秒/四分音符)、150(0.4秒/四分音符)、95(0.63秒/四分音符)で演奏されたとする。また、各小節と4小節からなるフレーズ全体が旋律断片であったとする。

テンポは演奏時間の逆数であるため、平均などの計算を行なうと実際の値と一致なくなってしまう(上記の例でも115, 133, 150, 95の平均は120にはならない)。よって、テンポに限ってはテンポの逆数である「四分音符の演奏長(秒/四分音符)」を演奏表情値として利用する。

演奏表情の大局比率は注目する旋律断片の演奏表情値とその親の演奏表情値の平均との比率である。この例では各小節の演奏表情の大局比率は個々の小節での四分音符の演奏長と全体の四分音符の演奏長との比率になる。よって、この例の各小節のテンポの演奏表情の大局比率は順に1.04, 0.9, 0.8, 1.26, となる。

演奏表情の大局比率から個々の旋律断片\$s_{n,j}\$の演

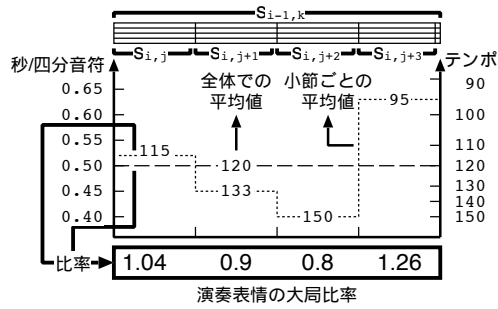


図 4 テンポの演奏表情の大局比率の計算例

Fig. 4 An example of musical expression ratio calculation for tempo

奏表情値の平均値 $exp(s_{n,j})$ を計算する際には、式 (15) に示すように曲全体での平均値 $exp(s_{1,1})$ に旋律断片 $s_{n,j}$ 自身と旋律断片 $s_{n,j}$ の全ての祖先 $\{s_{i,j_i} | s_{n,j} \subset s_{i,j_i}, 1 < i \leq n\}$ の演奏表情の大局比率 P_{exp} を掛け合わせればよい。

$$exp(s_{n,j}) = exp(s_{1,1}) \cdot \prod_{1 < i \leq n} P_{exp}(s_{i,j_i}) \quad (15)$$

$\{s_{n,j} \subset s_{i,j_i} \text{ for each } i\}$

図 5 にこの計算の一例を示す。半小節の演奏表情の大局比率は小節の演奏表情値と半小節の演奏表情値の比率である。また、小節の演奏表情の大局比率はフレーズの演奏表情値と小節の演奏表情値の比率である。従って、半小節の演奏表情の大局比率と小節の演奏表情の大局比率を掛け合わせることで、フレーズの演奏表情値と半小節の演奏表情値の比率が求まる。これにフレーズ全体の演奏表情値を掛けると、半小節の演奏表情値が得られる。

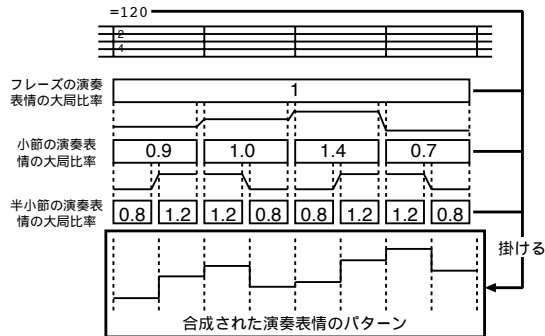


図 5 演奏表情の大局比率からの演奏表情値の計算の例

Fig. 5 An example of musical expression generation from the musical expression ratio

5.2 演奏表情の大局比率の合成

演奏表情生成の最終段階では、個々の参考事例の演奏表情の大局比率を合成し、対象曲の演奏表情を作成

する。本節では演奏表情の大局比率の合成について述べる。

対象曲の演奏表情の大局比率の合成は次の手順で行なう。

- (1) 対象曲の旋律断片ごとに、参考事例の演奏表情の大局比率を合成し、旋律断片の演奏表情の大局比率を作成
- (2) 各旋律断片の演奏表情の大局比率を組み合わせ、対象曲の音符ごとの演奏表情値を計算

旋律断片ごとの演奏表情の大局比率は、参考事例の演奏表情の大局比率の重要度による加重平均である。対象曲の旋律断片 $s_{i,j}$ の演奏表情の大局比率 $P_{exp}(s_{i,j})$ の計算を式 (16) に示す。 $E_{i,j}$ は $s_{i,j}$ に対する参考事例の集合、 $W(s)$ は参考事例 s の重要度を表す。

$$P_{exp}(s_{i,j}) = \frac{\sum_{s \in E_{i,j}} W(s) \cdot P_{exp}(s)}{\sum_{s \in E_{i,j}} W(s)} \quad (16)$$

曲全体の演奏表情の合成は、個々の旋律断片の演奏表情の大局比率を掛け合わせることで算出する。5.1 節で述べたように、個々の旋律断片の演奏表情の数値は式 (15) によって演奏表情の大局比率から計算することができる。音符も旋律断片なので、対象曲の個々の音符の演奏表情は音符自身及び音符の全ての先祖の旋律断片の演奏表情の大局比率の積に曲全体の演奏表情を掛けることで得られる。旋律断片の演奏表情の大局比率は式 (16) で求めたものを使用する。曲全体の演奏表情 $exp(s_{1,1})$ は曲全体に対する参考事例から獲得する。

6. 演奏表情生成システム

我々は本論文で提案する手法の有効性を検証するために、演奏表情生成実験システムを作成した。このシステムの構成を図 6 に示す。

実験システムで利用する演奏データ集の演奏データは 1) 楽譜情報、2) 楽曲構成情報、3) 演奏データ、4) 演奏状況、からなる。楽譜情報と楽曲構成情報は人手で作成した独自の書式のテキストファイルで与える。演奏データは標準 MIDI フォーマット形式 (SMF) のシーケンスデータで与える。SMF 形式のデータはコンピュータとシンセサイザを用いて容易に記録、作成することが可能である。演奏状況の設定は素性と程度量の組を列挙したテキストで与える。

システムの入力は、1) 楽譜情報、2) 楽曲構成情報、3) 演奏状況設定、からなる。これらはいずれも演奏データと同じ書式のテキストファイルである。楽曲構成情報は対象曲を旋律断片に分割する際に使用する。

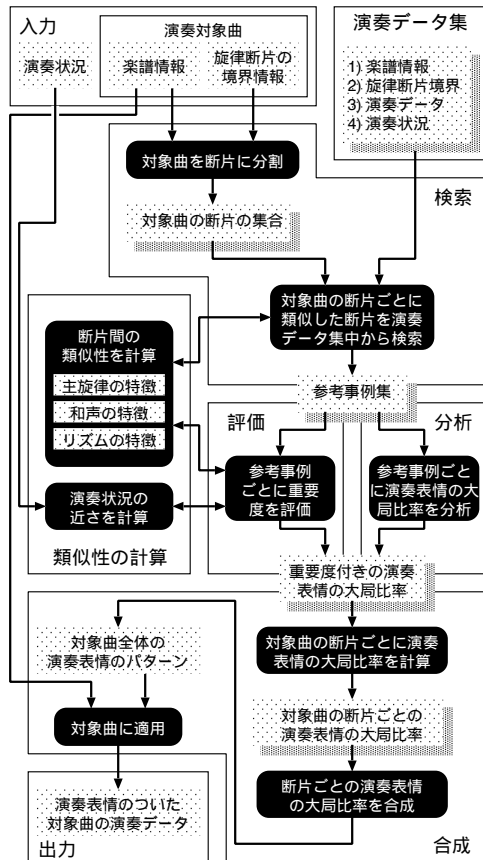


図 6 実験システムの構成

Fig. 6 The architecture of our experimental system

参考事例の検索の際の旋律の類似性の計算は 4.3 節で述べた方法を用いる。

旋律の検索では 4.3 節で述べた手法を用いて旋律断片の類似性を比較し、結果が閾値を越えたものを参考事例とする。事例の重要度は 1) 対象曲の断片との類似性, 2) 前後に隣接する断片どうしの類似性, 3) 親となる断片どうしの類似性, 4) 演奏状況の類似性, から求める。旋律断片の類似性は参考事例の検索と同じ方法で評価する (4.3 節)。演奏状況の類似性は 3.2 節で述べたように演奏状況ベクトルの内積から計算する。旋律の類似性と演奏状況の類似性の最大値が同じになるように正規化して合計し、指数関数に適用したものを重要度とした。生成された演奏データは SMF 形式のシーケンスデータで保存する。

このシステムでは、鍵盤楽器による単旋律の楽曲の演奏を対象としており、1) 局所テンポ, 2) 音符ごとの演奏長, 3) 音の強さ, の 3 種類の演奏表情を取り扱う。局所テンポは一音ごとのテンポである。音符ごとの演奏長は、楽譜上に記載されている音符の長さ、と

演奏長 (打鍵から離鍵までの時間) との比率である。通常は 1 よりやや小さい値になる。この数値が 0 に近ければスタッカートになり、1 より大きければレガートになる。音の強さは演奏音の打鍵の強さである。

7. 評価実験

7.1 実験

6 章で述べたシステムを使用して演奏データを生成した。実験曲として、ツェルニーの「八小節の練習曲」と「125 のパッセージ練習曲」から 4 分の 4 拍子の曲を 21 曲選択した。それぞれの曲について、音楽演奏の教育を受けた演奏者に「ロマン派風」と「古典派風」の二種類の演奏を行なってもらった。それぞれの演奏データの演奏状況は、素性を「ロマン派風」もしくは「古典派風」のいずれかで程度量を 1 とした。

この 21 曲の中から対象曲を 1 曲を選択し、残りの 20 曲の演奏データを演奏データ集として、対象曲の演奏表情の生成を行なった。従って、対象曲の演奏データは演奏データ集には含まれていない。生成データの演奏状況設定は、素性として「ロマン派風」と「古典派風」を与えた。それぞれの程度量の組合せを“1:0”, “0.7:0.3”, “0.5:0.5”, “0.3:0.7”, “0:1” とした 5 種類の演奏データを生成した。後述の評価では、この中から主に演奏状況設定を「ロマン派風:1, 古典派風:0」としたものと「ロマン派風:0, 古典派風:1」としたものを利用した。対象曲と演奏データ集の組合せを順次変更し、21 曲全てに対して演奏表情の生成を行なった。生成された SMF データは YAMAHA の Clavinova CLP-760 を用いて再生し、聴取実験用に録音した。

7.2 結果

生成された演奏表情の質と 2 種類の演奏スタイルの弾き分けの 2 点を中心に、人間による聴取と数値的な比較から演奏を評価した。聴取による評価では、生成された演奏データを音楽大学でピアノ演奏について学んだ演奏経験者および音楽教育者 2 名に聴かせ、演奏技術や演奏法などの妥当性の観点から演奏について評論してもらった。聴取者の内 1 名は演奏データ集のデータの演奏者である。数値による比較では、主に生成された演奏データと人間の演奏データとの差異や、生成された二種類の演奏状況のデータの間の差異を比較した。

まず、生成された演奏表情の質の評価について述べる。生成された演奏データの演奏表情は概ね人間の演奏のように聴こえる自然なものであった。演奏データ

集の中には弾き損じにより非常に弱くなってしまった音が見られたが、生成されたデータでは極端に弱い音は見られなかった。複数の事例を参考にしているため、単独の事例にしか見られないこういったエラーがある程度吸収されたためであると考えられる。

演奏表情の変化が顕著な部分で人間の演奏に良く似た傾向の変化を示すことが多かった。このような部分の多くが一般に演奏表情の与え方の特徴として良く知られているものであった。そのため、同じような演奏表情を持った事例が数多く得られたのではないかと思われる。人間の演奏とシステムが生成した演奏のテンポ変化の比較を図7に、音の強さの比較を図8にそれぞれ示す。これらはツェルニーの八小節の練習曲の1番の「ロマン派風」の演奏データである。図7の縦軸は演奏データ全体のテンポの平均値を音符単位の局所テンポで割った数値である。数値が大きいほどテンポが遅いことを表す。図8の縦軸は音符ごとの音の強さの値を演奏データ全体の音の強さの平均値で割ったもので、数値が大きいほど音が大きいことを表す。横軸の単位はいずれも小節である。テンポのグラフからは4小節目の末尾や6小節目の末尾、終末部分など、テンポが極大値や極小値をとっている場所が一致していることがわかる。また、音の強さのグラフからも前半部分や終末部分などの変化の傾向が類似していることが見てとれる。

一方で、音の強さの後半部分など演奏表情の変化の傾向が明らかに異なる部分も見られる。これらの部分の多くについては間違いと言うほどのものではないという意見が得られた。今回は演奏状況を「ロマン派風」と「古典派風」の2つで記述したが、同じロマン派風に分類される演奏表情の中にも様々な演奏の違いがあり、異なる部分の多くはこういった演奏状況に記述されていない細かな演奏状況の違いによって生じているのではないかと考えられる。

次に演奏状況に応じた弾き分けに関する評価について述べる。図9にシステムが生成した二種類の演奏状況のデータのテンポの比較を示す。演奏曲は図7と同じである。このグラフからも明らかなように、生成された二種類の演奏表情は明らかに異なるものであった。ロマン派風の演奏は古典派風の演奏に比べてテンポの変化の範囲が非常に広がっており、この点は一般的に知られているロマン派風の演奏の特徴とも一致している。また、生成された二種類の演奏状況の演奏データを聴取者によって「ロマン派風」か「古典派風」かの判定をさせたところ、判定に使用した全ての楽曲について正解が得られた。これらから、少なくとも今

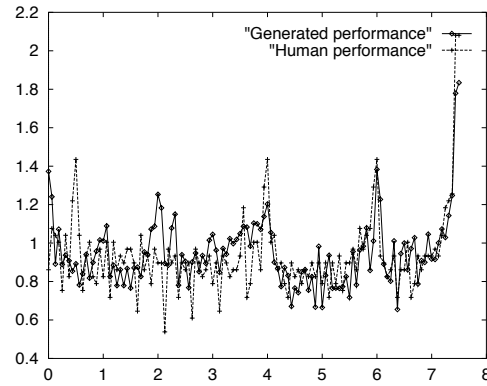


図7 システムの生成した演奏と人間の演奏のテンポの比較（八小節の練習曲，1番，「ロマン派風」）

Fig. 7 The tempo curve of the system v.s. human performance (No. 1, 160 Kurze Übungen, “Romantic”)

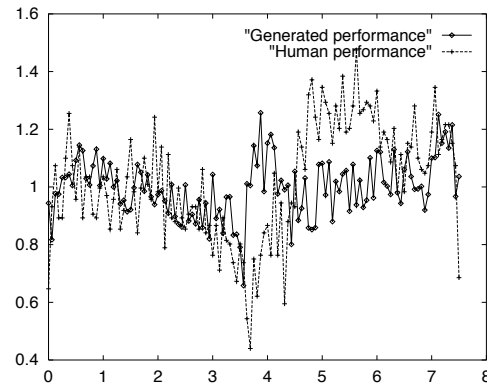


図8 システムの生成した演奏と人間の演奏の音の強さの比較（八小節の練習曲，1番，「ロマン派風」）

Fig. 8 The volume curve of the system v.s. human performance (No. 1, 160 Kurze Übungen, “Romantic”)

回与えた二種類の演奏状況の弾き分けは適切に行なわれていると評価することができる。

8. ま と め

本論文では演奏状況に応じた多様な演奏表情の生成が可能手法として、事例に基づく演奏表情の生成手法を提案した。この手法を利用した演奏表情生成システムを構築し、我々の提案する手法が演奏状況の設定に柔軟に対応し多様な演奏表情の生成を行なうことが可能であることを確認した。

今後、検討すべき課題としては、旋律の類似性の評価方法の検討、参考事例の検索の高速化、演奏状況についての考察、旋律の境界情報の演奏表情への影響、和音を含む楽曲への適用などがあげられる。

本手法では旋律の類似性の評価方法が生成される演奏表情の質に大きな影響を与える。実験システムでは

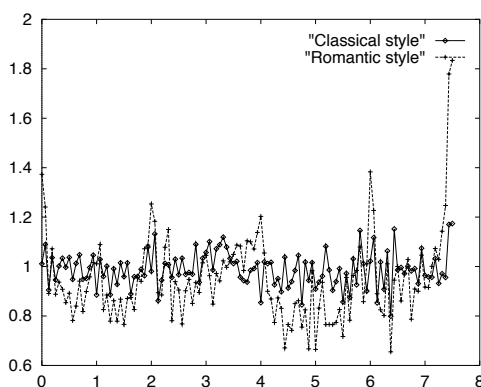


図9 「ロマン派風」と「古典派風」の演奏のテンポ変化（八小節の練習曲，1番）

Fig. 9 The tempo curve of the "Romantic" and "Classical" style performances (No. 1, 160 Kurze Übungen)

旋律の類似性を単純な方法で評価したが，この部分についてはまだ検討する余地がある。

実験システムでは演奏データ集の全ての楽曲と比較し，類似性の高いものを検索しているが，演奏データ集の規模や対象曲の旋律断片の数が大きくなると検索に要する時間が膨大になってしまう。そこで，実用的なシステムを構築するためにはインデックスなどを利用して全検索を避けるような旋律検索方法を実装する必要がある。

本論文の例示や実験などで用いた演奏状況の素性にはあまり妥当な根拠がなく，ある意味で一般的なものを適当に選択したに過ぎない。演奏ごとの特徴を演奏状況で適切に表現するためには，演奏状況にどういった素性の集合を用いるべきかについて音楽心理学¹²⁾や音楽認知学などの観点から検討する必要がある。また，楽曲外の要因だけではなく単旋律の楽曲における和声など楽譜のみからでは一意に決定できない楽曲情報についても演奏状況の一部として考慮する必要がある。

4.1 節の冒頭で述べた通り，本手法では計算コストの観点から楽曲の境界情報として楽曲構成を利用しているが，類似旋律の検索が有効に行なわれるのであれば，必ずしも楽曲構成に基づいたものである必要はない。そこで，4.1 節に示した理想的な分割手法や他の知識に基づく分割手法を用いて生成した演奏表情と，本論文で示した分割手法を用いた場合との比較を行なうと共に，計算コストと演奏表情の双方の観点から良好な結果が得られる分割手法を追求する必要がある。

本論文で用いた実験システムは，システムの複雑化を避けるため単旋律の楽曲にしか対応できない。しかし，単旋律の楽曲では演奏表情の適切性の評価が難し

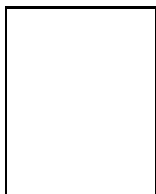
いという意見があったので，和音を含む楽曲に対応できるシステムを構築し，一般の楽曲を用いた評価を行なうことも重要である。

参考文献

- 1) 平賀瑠美: 演奏の表情付け, bit 別冊, コンピュータと音楽の世界, 共立出版, pp. 270 - 282 (1998).
- 2) Gerhard Widmer: Modeling the rational basis of musical expression, Computer Music Journal, Vol. 19, No. 2, pp. 76-96 (1995).
- 3) 白川健一, 小田安彦, 熊谷俊行, 梶川嘉延, 野村康雄: 演奏者情報を加味したピアノの自動演奏システムの構築: ~ 音符属性に着目した個人的解釈ルールの生成 ~, 情報処理学会研究報告, Vol. 95, No. 46, 95-MUS-10, pp. 1-6 (1995).
- 4) 山内敏彰, 野池賢二, 野瀬隆, 乾伸雄, 小谷善行, 西村怒彦: Ga を用いたピアノ演奏の局所テンポ解析, 情報処理学会研究報告, Vol. 96, No. 19, 96-MUS-14, pp. 1-6 (1996).
- 5) 青野裕司, 片寄晴弘, 井口征士, 重回帰分析を用いた演奏表現法の抽出, 情報処理学会論文誌, vol. 38, No. 7, pp. 1473-1481 (1997).
- 6) Taizan Suzuki and Takenobu Tokunaga and Hozumi Tanaka: A case based approach to the generation of musical expression, Proc. of IJ-CAI '99, pp. 642-648 (1999).
- 7) Janet Kolodner: Case-based reasoning, Morgan Kaufmann (1993).
- 8) Josep Lluís Arcos, Ramon Lopez de Mantaras, and Xavier Serra: SaxEx: a case-based reasoning system for generating expressive musical performances, Proc. of the 1997 ICMC, pp. 329-336 (1997).
- 9) 園田智也, 後藤真孝, 村岡洋一: WWW 上での歌声による曲検索システム, 電子情報通信学会技術報告, Vol. 97, No. 560, pp. 25-32 (1998).
- 10) 西原祐一, 小杉尚子, 紺谷精一, 山室雅司: ハミング検索用音楽データベースの分析~部分曲分割方法の改善に向けて~, 情報処理学会研究報告, Vol. 99, No. 84, 99-MUS-32, pp. 7-12 (1999).
- 11) 田口友康: 自動演奏と演奏モデル: 演奏の芸術性を目指して, bit 別冊, コンピュータと音楽, 共立出版, pp. 42-53 (1987).
- 12) 寺西立年, 大串健吾, 宮崎謙一: 音楽の心理学 (下), 西村書店 (1987).

(平成 0 年 0 月 0 日受付)

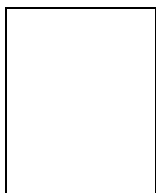
(平成 0 年 0 月 0 日採録)



鈴木 泰山 (学生会員)

1973年生。1995年東京工業大学工学部情報工学科卒業。1997年東京工業大学大学院情報理工学研究科修士課程修了。同年東京工業大学大学院情報理工学研究科博士課程入学。

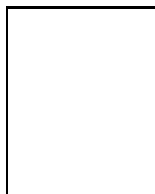
現在、同大学院博士課程に在学。音楽情報処理に関する研究に従事。



徳永 健伸 (正会員)

1983年東京工業大学工学部情報工学科卒業。1985年同大学院理工学研究科修士課程修了。同年(株)三菱総合研究所入社。1986年東京工業大学大学院博士課程入学。現在、同大

学院情報理工学研究科助教授。自然言語処理、計算言語学に関する研究に従事。工学博士。認知科学会、人工知能学会、言語処理学会、計量国語学会、Association for Computational Linguistics、各会員。



田中 穂積 (正会員)

1964年東京工業大学理工学部制御工学科卒業。1966年同大学院修士課程修了。同年電気試験所(現、電子技術総合研究所)入所。1983年より東京工業大学工学部助教授。現在、同

大学院情報理工学研究科教授。自然言語処理、人工知能に関する研究に従事。工学博士。電子情報通信学会、認知科学会、人工知能学会、計量国語学会、言語処理学会、Association for Computational Linguistics、各会員。

