

演奏表情生成システム COPER における 音楽グループ境界に基づく事例検索方式と ペダル情報生成の試みについて

野池賢二 片寄晴弘^{†, ††}

本稿では、事例を参照し演奏表情を生成するシステム COPER が用いている、音楽グループ境界に基づく事例の検索方式について述べる。また、ダンパーペダル情報の生成の試行とその結果について述べ、音符の属性値としてモデル化しづらいためにこれまで演奏表情パラメータとしてほとんど扱われてこなかった時系列情報の重要性にも触れる。

The Search Method based on Musical Group Boundary and A Trial of Pedal Information Generation in the Performance Rendering System “COPER”

Kenzi Noike[†] and Haruhiro Katayose^{††}

In this report, we describe that the search method based on musical group boundary information in the performance rendering system “COPER”, and describe a trial of generation of damper pedal action information. It had not been treated in the past research of performance rendering, because it is difficult to represent time series information as note’s attributes. We think this parameters is essential, try to generate it..

1. はじめに

現在の音楽情報処理研究の主要な研究テーマのひとつに“演奏の表情付け (Performance Rendering)”がある[1].「表情豊かな演奏の生成」などの言葉で言い表される、人間の演奏に含まれる速度変化、強弱付けなどの演奏表情を自動的に生成する研究は、それらをコンテスト形式で競う Rencon の定期的な開催によって、一層の盛り上がりを見せている[2].

これまでに提案されてきた演奏表情生成の手法は、大きく分けると、“ルール適用型”、“事例参照型”の二つに分けることができよう。事例参照型の演奏表情生成手法では、実際の演奏事例を大量に収録した事例データベースから表情付け対象曲に類似した事例を検索し、それらに含まれる演奏表情を参照することで演奏表情を生成する。

我々は、事例参照型の演奏表情生成手法による演奏表情付け研究に取り組んでおり、COPER と名づけたシステムを開発している[3][4]. COPER は、事例検索時に合致事例数に応じて検索条件を段階的に緩めるという特徴を持つ。

本稿では、COPER が用いている、音楽グループ境界に基づく事例検索方式について述べる。また、ダンパーペダル情報の生成の試行とその定性的な結果にも触れ、これまで演奏表情パラメータとしてほとんど扱われてこなかった、音符の属性値としてのモデル化がしづらい時系列情報を扱うことの重要性にも触れる。

2. 事例参照型演奏表情生成システム COPER

2.1 COPER の設計思想

事例参照型表情付けシステムが良質な結果を得るためには、次の 2 点に留意したシステムの構築が不可欠である。

- 事例データベースのデータスパースネス問題への対処
- 良質な表情を含んだ事例データベースの準備

これら 2 点の留意事項に対する我々の設計思想について述べる。

2.1.1 事例データベースのデータスパースネス問題への対処

事例参照型の表情付けシステムの先行例として、Kagurame[5][6][7], SaxEx[8],

[†] 関西学院大学 理工学部

School of Science and Technology, Kwansai Gakuin University

^{††} 科学技術振興機構 戦略的創造研究推進事業 CrestMuse プロジェクト

CrestMuse Project, CREST, JST

Ha-Hi-Hun[9], Itopul[10] が挙げられる。

データスパースネス問題への対処として、Kagurame は、2 のべき数的な単位で楽曲を階層的な旋律断片に分割し、検索単位を小さくすることと、その階層構造を利用した検索と表情の生成を行うことで対処している。SaxEx は、事例検索を抽象化データを用いて行うことで対処しており、その抽象化データとして IRM(Implication/Realization Model) に基づく構造[11] と、GTTM(A Generative Theory of Tonal Music)[12] の Metrical Structure 分析に基づく拍節的重要度の二つの規準を用いている。Ha-Hi-Hun も事例検索を抽象化データを用いることで対処しており、その抽象化データとして、GTTM の Time-Span Reduction 分析の木構造の deductive object-oriented database の包摂関係表現を用いている。Itopal も同様であり、抽象化データとして、線分近似による旋律概形を用いている。

これに対し COPER は、“事例検索の基準そのものを段階的に緩める”という戦略によってデータスパースネス問題に対処する。これは、他の分野における事例ベース推論システムにおいて、事例検索時に適度な数の合致事例を見つけるために、類似度の計算レベルを段階的に緩めていくことに相当する。

実際には、多様な検索対象に対して、合致事例が常に適度に分布する事例データベースを用意することは困難であり、またそのような状況下で類似度計算レベルを適切に何段階も細かく用意することも難しい。類似度計算レベルを一段階緩めた途端に合致事例数が劇的に増加し、欲する適切な事例が多量の検索結果に埋もれてしまうことは容易に起こりうる。多数の合致事例が見つかった状況下で、それらの合致事例に含まれる特徴量を平均化して用いることは、必ずしもよい結果を得られるとは限らない。合致事例間に矛盾が存在する可能性が高く、また、矛盾が存在しなかったとしても、本研究のように演奏の表情付けを目的としている場合には、演奏表情情報を平均化することは避けたほうがよい可能性がある。

細かな計算レベルを用意することができない場合（あるいは意図的に用意しない場合）は、多数の合致事例が見つかった時点で統計的手法による選択を行うということが考えられる。例えば、Cope らの EMI[13] や Conklin らのコラルメロディ予測[14] などでは、統計的手法による選択（予測）がよく機能している。

以上から、我々は事例検索を、1) 適度な数の合致事例が見つかるまで、事例検索の類似度計算レベルを具体度の高いほうから低いほうに向かって制御し、2) 類似度計算レベルを一段階緩めたときに合致事例が劇的に増加してしまうときは、統計的処理による最尤事例選択を行う、という戦略で行う。

現時点では、統計的処理による最尤事例選択の部分は設計中であり、検索基準を段階的に緩める部分も、この事例検索戦略の効果を調べられる程度の初期的で簡易な実装となっている。

2.1.2 良質な表情を含んだ事例データベースの準備

事例参照型の演奏表情付けシステムが良質の演奏表情を生成するためには、事例データベース中の事例が良質の情報を持っていることが不可欠である。我々は、演奏事例の演奏表情記述形式と、事例データベースに収録する事例の選定に留意した。

演奏事例の演奏表情記述形式は、豊田らの演奏 deviation データベース[15] の設計方針に基づき、演奏表情を「拍以上のテンポ変動」と「拍以下の微細な“ずれ”」とに分離した記述形式を採用する。これにより、大局的な演奏表情と微細な演奏表情とを分けて扱うことも可能となる。

事例データベースに収録する事例集としては、現時点では、豊田らの演奏 deviation データベースのほかに、CrestMusePEDB[16] を用いている。CrestMusePEDB は、その設計方針が豊田らの演奏 deviation データベースとほぼ同じであり、収録データの元演奏の品質も明示されているからである。

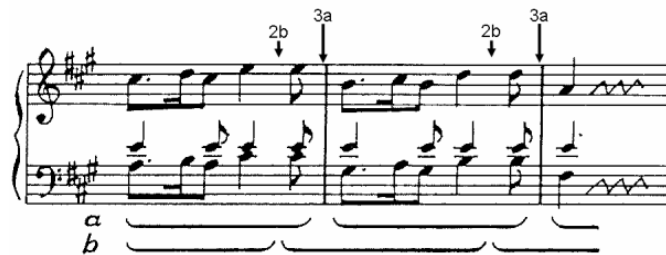
2.2 音楽グループ境界に基づく事例検索

図 1 は、ひとつの音符列に対して異なる二つの音楽グループ構造解釈が可能である場合 (a) のそれぞれのグループ構造を表現する演奏を、(b) と (c) にピアノロール状に図示したものである。図 1 (b) は、図 1 (a) の五線譜下部に示したグループ構造 a を表現した演奏、図 1 (c) は、同様にグループ構造 b を表現した演奏である。図 1 (b) では図 1 (a) の 3a のところで、図 1 (c) では図 1 (a) の 2b のところで“間”をあげ、さらに、直後の音を強めに演奏することで、この部分の音楽グループ境界が表現されている。

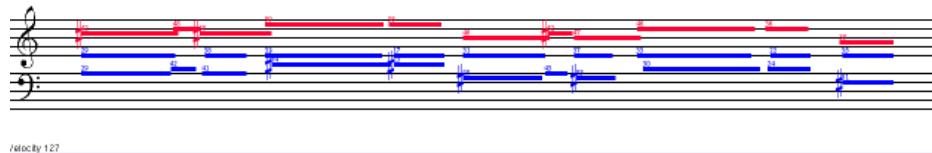
これは逆に、演奏情報から音楽グループ構造を得ることが可能であることも示しており、実際に、片寄らの演奏ルール抽出研究[17] では、ひとつの音符列に対する異なる複数の演奏情報から演奏ルールを抽出したときに、異なる二つの音楽グループ構造の抽出に成功している。

これらから、音楽グループ構造と演奏表情とは密接な関連があることがわかり、演奏表情を扱う上では音楽グループ構造への考慮が重要であることがわかる。

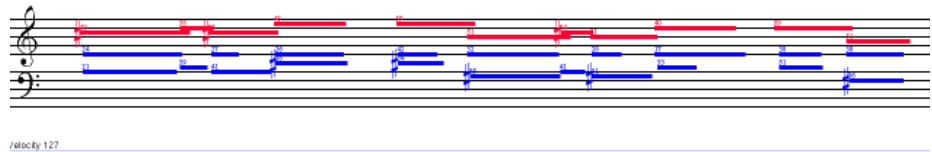
しかし、2.1.1 節で触れたように、事例データベースのデータスパースネス対策として、Kagurame は楽曲を機械的に分割している。三連符などへの考慮[6] はあるものの、基本的に小節線に基づいた 2 のべき数的な単位での機械的な分割であり、アウフタクトやフレーズ構造などの音楽グループ構造への考慮がないため、旋律断片への分割の過程で、音楽グループ構造を分断することが起こりうる。機械的な分割は、システム設計者の恣意性を排除でき、また、処理の高速化にも寄与するが、前述したように、音楽グループ構造と演奏表情とは密接な関連があると考えられるため、音楽グループ構造への考慮がない楽曲分割手法は、演奏表情情報を扱う上では適切ではないと我々は考える。



(a) K.331 の 2 とおりの音楽グループ構造



(b) 音楽グループ構造 a を表現する演奏



(c) 音楽グループ構造 b を表現する演奏

図 1 音楽グループ構造の違いによる演奏表情の違い

SaxEx, Ha-Hi-Hun, Itopul は、楽曲の音楽的な構造に基づく事例検索を行っているが、これらは楽譜情報に基づいて得られた楽曲構造を用いており、事例がもつ演奏表情情報は加味されていない。したがって、図 1 (b)、図 1 (c) の両方の演奏が事例データベースに収録されているとき、音楽グループ構造の表現が異なる演奏表情が同

等に混在して扱われることになり、そこから生成される演奏表情による音楽グループ構造表現が不適切になる可能性がある。

これらを踏まえて、我々は、演奏表情情報を加味した音楽グループ構造に基づく事例検索を行う。音楽グループ構造の分析手法として、GTTM の Grouping Structure 分析ルールのうち、GPR1, 2, 3, 4 を利用する。GPR2, 3 については、演奏表情情報を加味した分析を行えるように拡張した GPRp [18] を用いる。

GPR は、基本的に単旋律への適用を想定した分析ルールであるため、複数旋律に対してはそのままでは適用できない。GPR は主旋律に対して適用することが適切であると考えられるため、橋田らのモデル[19] のアテンションパートのような、“主旋律らしい声部”を自動的に抽出し、それに対して適用することが理想的である。しかし、アテンションパートを自動的に抽出することの実現は容易ではない。

そこで、“主旋律らしい声部”の自動抽出を暫定的に実現するために、井田らの試み[20] の“第一声部進行”と“ボイスリーディング”の考えを用いた。“第一声部進行”は、「音高が最も高い音符を主旋律の音符とみなす」という旋律評価方法であり、“ボイスリーディング”は、「隣り合う音符の連続性」についての評価である。我々は、“第一声部進行”に基づいて主旋律らしい声部を抽出し、“ボイスリーディング”によって連続性が低いと評価される音符への推移を避けるように実装した。

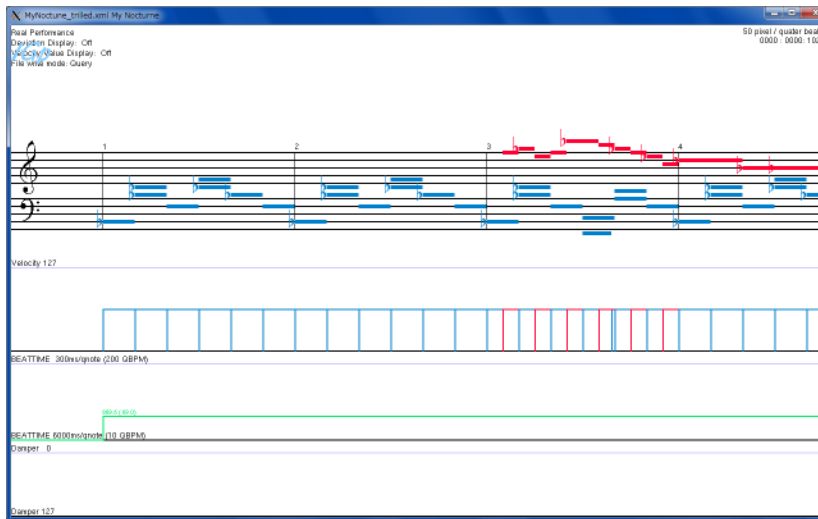
抽出した“主旋律らしい声部”に対して GPR(GPRp) の 1, 2, 3, 4 を適用することで、音楽グループ境界の候補を列挙することができる。GRP の各ルールによって“境界らしさ”の強さが算出されるので、その値と実験で求めた閾値[18] とにより列挙した境界候補を選別し、それに基づいた楽曲分割による旋律断片を利用して事例検索を行う。

2.3 ダンパーペダル情報生成の試み

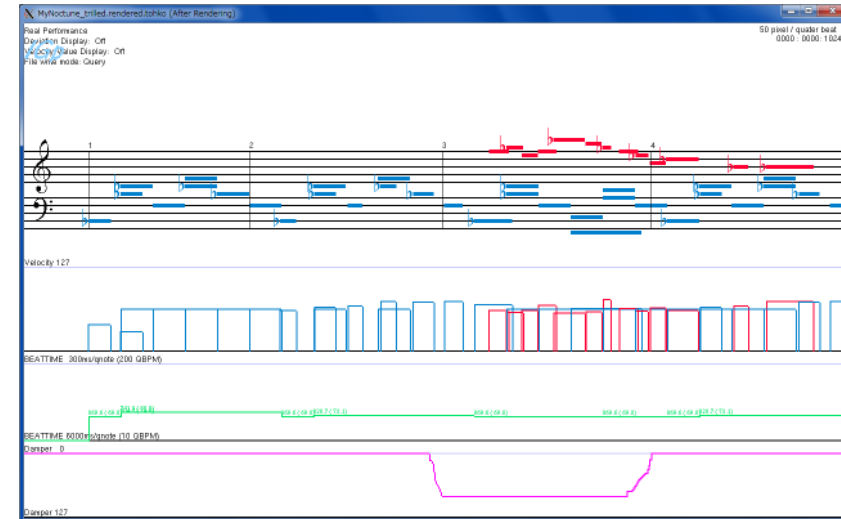
これまでに発表されてきた演奏表情付けシステムの中で、ペダル情報を演奏表情パラメータとして扱っているシステムは、我々の知る限りでは皆無である。これは、ペダル情報のような時々刻々と変化するパラメータは音符の属性値としてのモデル化が難しいため、機械学習やルール抽出の題材からは避けられてきたからではないかと思われる。

しかし、人間がピアノを演奏するときに、「ペダルを使わずに」表情豊かに演奏してください」という指示がなされると、通常通りに表情豊かに演奏することは難しい。ペダルを使わずに弾くことができたとしても、そのときの演奏の音の響きは、聴き慣れた表情豊かな演奏とは異なるであろう。

演奏表情付けシステムがペダル情報を扱わない状況は、人間が“ペダルを使わずに弾かされている”状況と等価であり、いくら演奏速度や打鍵強度などを人間の演奏に近づけても、そのデータを演奏して聴こえてくる音は人間の演奏とは異なる音となる可能性がある。

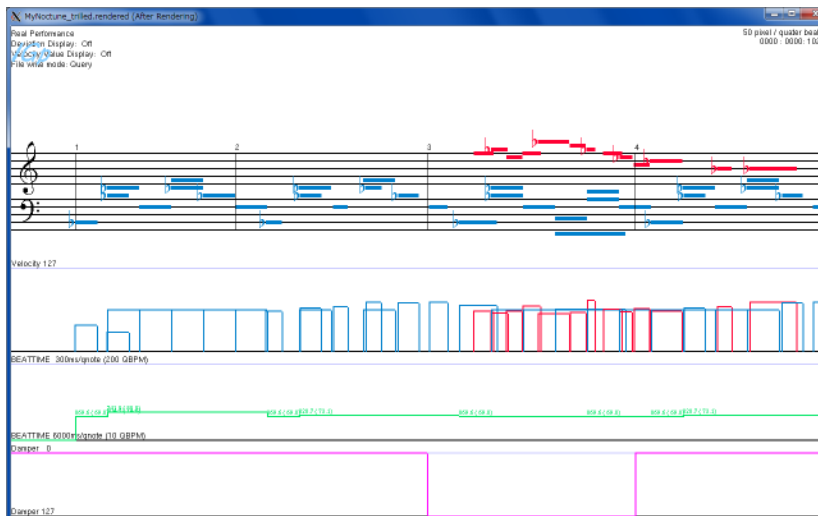


(a) 演奏表情生成前



(c) ダンパーペダル踏み/離しテンプレートを適用

図 2 ダンパーペダル情報の生成の様子



(b) 演奏表情生成後

そこで我々は、「人間のように表情豊かなピアノ演奏」を自動生成する上でペダル情報も扱うことは必要不可欠であると考え、事例参照型の演奏表情付けシステムにおいてペダル情報も扱うことにした。手始めに、ピアノの3種のペダルのうち、演奏音への影響が最も大きいダンパーペダルの情報を演奏表情として生成することを試みた。

COPER は、音符の属性値として表現できる演奏表情パラメータだけでなく、時々刻々と変化する時系列情報をも扱えるように設計してある。そこで、ダンパーペダル情報も生成する初期的な試みとして、事例に含まれているダンパーペダル情報をそのまま表情付け対象曲に転写するようにした。これは、非常に稚拙な実現方法であり、生成の根拠となる因果律も希薄であるが、生成される演奏表情にダンパーペダル情報が含まれることの効果を知るという観点では十分な方法である。

図 2 に、ダンパーペダル情報の生成の様子をピアノロール表示で示す。図 2 (a) が演奏表情生成前の演奏データであり、(b) が演奏表情生成後の演奏データである。図 2 (b) の下部のピンク色の線がダンパーペダルの踏度を表しており、(a) にはなかったダンパーペダル情報が、(b) にはあることがわかる。

図 2 (b) から、確かにダンパーペダル情報が生成されていることは確認できるが、ピンク色の線が階段状になっていることからわかるように、このダンパーペダル情報では、“ペダルを一瞬にして最深まで踏み込み、一瞬にして完全に離す”ことになる。

このようなデータが生成されるのは、事例データベースに含まれている事例に、ダンパーペダル情報を“踏む”、“離す”の2値だけで構成している事例が存在しているからである。この演奏データを、MIDI 制御できるピアノ（YAMAHA サイレント・ピアノ、CEUS グランド・ピアノ[21]）を使って演奏したところ、バタバタと激しくペダルが動き、聴衆にとっての印象が悪く、音の響きへの貢献具合を評価する以前に改善すべき問題であることがわかった。

そこで、簡易的な改善策として、既存の人間の演奏からペダル操作情報をあらかじめ抽出し、それをテンプレート化しておき、演奏表情生成後にそのテンプレートを適用することで、人間が踏む／離すような滑らかな動作へのスムーズングを行った。その結果を図 2 (c) に示す。ピンク色の線で表されたダンパーペダル踏度をみると、図 2 (b) と違い、(c) では滑らかに変化していることがわかる。

3. 評価

3.1 生成される演奏表情の評価

生成される演奏表情の質の評価は、Rencon での投票結果をそのまま利用した、定性的で主観に基づく評価しかまだ行っていない。最新の Rencon である ICMPC10 Rencon での投票結果を図 3 に示す。ICMPC10 Rencon では、2 曲を生成し、それぞれの“自然さ”と“表情豊かさ”の2項目を投票によって評価した。COPER は、2 曲の2評価項目の結果を総合した順位では参加 4 システム中の最下位ではあったが、“表情豊かさ”の総合評価では3位であった。投票数の差の有意性は未検証であるが、図 3 のグラフを概観すると YQX 以外のシステムの間には、顕著な差はないように見える。

投票後に、コンテスト会場にて何人かの方に COPER が生成した演奏表情のうち、特にダンパーペダルの効果についての感想を求めたところ、

- うまく効いていたところがあった。
その部分は他のシステムと音の響きが違って、よく聴こえた。
- 長い時間、踏みっぱなしになっていたところがあった。
あれがあるくらいならば、ペダル動作がないほうがいいのかもかもしれない。

というよい点と悪い点のご意見をいただいた。長い時間、ダンパーペダルを踏みっぱなしにしてしまう演奏データが生成されてしまうのは、“踏む”と“離す”の事例を対で管理していないことが原因である。この結果から、因果律に基づいてペダル情報を生成することが必須であることを、あらためて認識できた。

しかしながら、これらから「適切にダンパーペダル情報を生成すれば、生成する演奏表情の質が上がる」ことが確認でき、これまで演奏表情付けシステムが扱ってこなかったペダル情報の生成が、生成演奏の質を上げるために重要であることが確認できた。

今後は、主観に基づく定性的な評価だけではなく、客観的で定量的な評価を実施することを視野に入れていきたい。その際の評価方法としては、いま現在のアクティビティの高い研究グループのシステムが採用している演奏表情パラメータだけではなく、少し先を視野に入れ、多様な演奏表情パラメータを評価できるようにしたい。

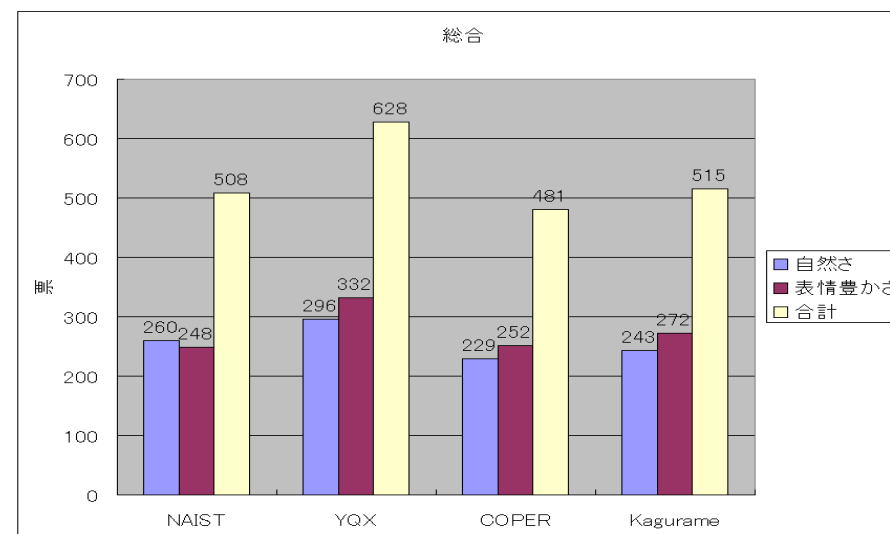


図 3 ICMPC10 Rencon の結果

3.2 事例検索速度の評価

事例参照型の演奏表情付けシステムの一般的な特徴として、事例データベースの規模を大きくすればするほど生成される表情の質は向上するが、結果を出力し終えるまでの時間が増大する、ということがある。これは、Rencon のような「制限時間内で生成する」という条件が課せられることを想定すると、考慮すべき事項である。

そこで、現在の COPER の事例検索にかかる速度を概算した。

現在の COPER の事例データベースは、728 旋律断片 (17 曲) を収録している。この事例データベースを用い、ICMPC10 Rencon での課題曲「My Mozart in Sentiment」

の表情付けを行ったところ、課題曲は 34 旋律断片に分割され、4 秒程度で処理し終えることができた。Rencon では、表情付け処理に 60 分の持ち時間を与えられるので、これらから単純に逆算すると、事例データベースに収録する旋律断片数を 655200 に、曲数で言い換えると 15300 曲にまで大きくすることが可能である。

このことから、現在の COPER の演奏表情付けアルゴリズムは十分に速いため、制限時間にはあまり留意する必要がなく、事例データベースの規模を大きくできることがわかった。今後は、生成される演奏表情の質を向上させるために、事例データベースの事例収録数を増大させ、内容の拡充に努めていきたい。

4. おわりに

本稿では、事例参照型の演奏表情付けシステム COPER が行っている、GPR (GPRp) を利用した音楽グループ境界に基づく事例検索方式について述べた。また、これまでの演奏表情付けシステムではほとんど扱われてこなかったダンパーペダル情報の生成に着目し、その生成を試み、重要性について触れた。

現在の COPER は、掲げている演奏表情生成戦略と比べて、非常に初期的な実装に留まっており、また事例の統計的な選択については未実装のままである。ダンパーペダル情報生成の試行についても、非常に稚拙な実現方法であり、生成の根拠となる因果律も希薄である。

しかし、これまでの演奏表情システムが注意を払ってこなかった「事例の演奏表情も加味した検索が必要であること」、「ペダル情報を演奏表情として扱うことの重要性」に触れ、問題提起の役割は果たしているのではないかと思われる。

解決すべき課題は山積しているが、ひとつひとつ着実に取り組んでいきたい。

なお、COPER は、<http://noike.info/~kenzi/cgi-bin/cow/> にて公開しており、いつでもだれでも試用し、評価することができるので、興味を持たれた方はぜひ試していただきたい。

謝辞 ダンパーペダル情報の生成試行にあたって、NICT ユニバーサルメディア研究センターの大島千佳さんに演奏情報のご提供をいただきました。謹んで感謝の意を表します。

参考文献

- 1) 平賀瑠美, 平田圭二, 片寄晴弘: 蓮根, 目指せ世界一のピアニスト, 情報処理, Vol.43, No.2, pp.136-141 (2002)
- 2) 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, pp.67-72 (2008)

- 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, pp.67-70 (2008)
- 4) Kenzi NOIKE, Haruhiro KATAYOSE: Corpus-based performance rendering system with Web UI: COPER on the Web, Rencon Workshop at ICMPC 10, (2008)
- 5) Taizan Suzuki: The Second Phase Development of Case Based Performance Rendering System "Kagurame", In Working Notes of IJCAI-03 Workshop on methods for automatic music performance and their applications in a public rendering contest, pp.17-25 (2003)
- 6) 清水厚志, 鈴木泰山, 野池賢二, 金子雄介, 徳永幸生, 杉山精: 事例に基づく演奏表情生成システムにおける旋律断片自動生成アルゴリズムの改良と評価, 情報処理学会研究報告 2007-MUS-72, pp.7-12 (2007)
- 7) 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における旋律類似性の評価内容の視覚化, 情報処理学会研究報告 Vol.2009-MUS-80, No.8 (2009)
- 8) Arcos,J.L., de Mantaras,R.L., Serra,X: SaxEx: A case-based reasoning system for generating expressive musical performances, In Proceedings of the 1997 ICMC, pp.329-336 (1997)
- 9) Hirata, K., Hiraga, R.: Ha-Hi-Hun plays Chopin's Etude, In Working Notes of IJCAI-03 Workshop on methods for automatic music performance and their applications in a public rendering contest, pp.72-73 (2003)
- 10) 伊藤洋介, 橋田光代, 片寄晴弘: 複数の生成プロセスが制御可能な演奏生成システム「Itopul」, 情報処理学会研究報告 2007-MUS-73, pp.45-50 (2007)
- 11) Narmour, E.: The Analysis And Cognition Of Basic Melodic Structures, the University of Chicago Press, (1977)
- 12) Fred Lerdahl, Ray Jackendoff: A Generative Theory of Tonal Music, MIT Press, (1983)
- 13) Cope, D.: Computers and Music Style, A-R EDITIONS (1991)
- 14) Conklin, D., Witten, I.: Multiple Viewpoint Systems for Music Prediction, Journal of New Music Research, Vol.1, pp.51-73 (1995)
- 15) 豊田健一, 片寄晴弘, 野池賢二: 音楽解釈研究のための演奏 deviation データベースの作成, 情報処理学会研究報告 2003-MUS-51, pp.65-70 (2003)
- 16) 橋田光代, 松井淑恵, 北原鉄朗, 酒造祐介, 片寄晴弘: 音楽演奏表情データベース CrestMusePEDB ver.1.0 の公開について, 情報処理学会研究報告 2007-MUS-72, pp.1-6 (2007)
- 17) 片寄晴弘, 福岡俊之, 井口征士: 音楽解釈システム MIS における演奏ルールの抽出について, 情報処理学会第 40 回全国大会講演論文集, 6R-6, pp.1587-1588 (1990)
- 18) 野池賢二, 橋田光代, 竹内好宏, 片寄晴弘: 聴取者傾向を加味した GTTM グルーピング規則適用の演奏表情パラメータへの拡張, 情報処理学会研究報告 2004-MUS-57, pp.11-16 (2004)
- 19) 橋田光代, 長田典子, 河原英紀, 片寄晴弘: 複数旋律音楽に対する演奏表情付けモデルの構築, 情報処理学会論文誌, Vol.48, No.1, pp.248-257 (2007)
- 20) 井田健太郎, 平田圭二, 東条敏: GTTM に基づくグルーピング構造及び拍節構造の自動分析の試み, 情報処理学会研究報告 2001-MUS-42, pp.49-54 (2001)
- 21) Kenzi NOIKE, Yosuke ITO, Haruhiro KATAYOSE: Corpus based performance rendering system: COPER, ISMIR2006 Rencon Performance Rendering Contest (2006)