

演奏生成モデル PopE における演奏家個性記述能力に関する検討

A Study of Description Capability of Performance Characteristics on PopE

橋田光代*¹ 長田典子*² 片寄晴弘*^{3*2}
Mitsuyo Hashida Noriko Nagata Haruhiro Katayose

*¹和歌山大学システム工学研究科 *²関西学院大学理工学部 *²関西学院大学文学研究科
Wakayama University Kwansai Gakuin University Kwansai Gakuin University

This paper describes a musical interpretation model for Poly-phrase Ensemble: PopE. PopE, where “attention transition” is focused on in designing the model, generates expressive music performance as follows; 1) giving each voice part an independent performance expression, 2) introducing a parameter related to attention transition, and 3) synchronizing onset timing between voice parts by using prolongational reduction and the rule of attention transition. The paper also describes how the middle part of Chopin’s “Fantasie-Impromptu” awarded NIME-Rencon was rendered using PopE. The averages of ratio of tempo and velocity of the reconstruction of the performances of three pianists was 0.95 and 0.69 respectively. Ten subjects who experienced music identified the original performances from the reconstructed performances.

1. はじめに

演奏表情付けに関する研究 (Performance Rendering) は、音楽情報処理の中心的な研究領域の一つである [片寄 04]。主旋律のフレージングについては、音楽構造解析で部分的な人間の介入を前提とするものの、2000年の段階で人間の演奏者にひけをとらない演奏も始めている [橋田 02]。その一方で、ポリフォニー*¹に対する自然な表現が、これからの課題として認識されるに至っている。この課題については、拍内の微細な演奏表情を保存した状態での拍打型演奏生成インタフェースが提案されている [Raphael 03, 片寄 03]。しかし、これらはポリフォニーのための音楽解釈モデルに基づいたものではなく、楽曲毎に表情のついたテンプレートを用意しなければならないという制約があった。

我々は、ポリフォニーの自然な演奏表現を行うための音楽解釈モデル PopE のデザインを進めてきた [橋田 05]。本稿では、特に PopE の演奏記述能力に関する検討を中心課題とする。以下、第2章で、声部協調型音楽解釈モデル PopE の基本的な構成について述べる。第3章で、ショパン作曲「幻想即興曲」中間部を対象に (1) Rencon Award (第一位) の対象となった演奏の Rencon における主観的評価 (2) PopE の特徴的な処理であるアテンション移動時における音長を延ばすルールの意味の検討 (3) 3人のピアニストの演奏に対する記述再現能力に関する検討を実施する。

2. 声部協調型音楽解釈モデル PopE

PopE (Musical Interpretation Model for Polyphrase Ensemble) は、楽曲の各声部に対するアテンション (2.2 節) の移動と、GTTM [Lerdahl 83] のグループ構造と延長簡約を利用することにより、1) 各声部に独立した演奏表情を与える、2) アテンションと延長簡約を利用して声部間の発音タイミングを合わせる (同期処理, 2.3 節)、3) アテンションの移動に係る演奏ルールを導入する (2.4 節) ことを基本とする演奏生成モデルである。以下、モデルの概要について述べる。

連絡先: 669-1337 兵庫県三田市学園 2-1 関西学院大学理工学部情報科学科 片寄研究室, Tel&Fax: 079-565-7861, hashida@ksc.kwansai.ac.jp

*¹ 複数声部からなり、主旋律やそれ以外のパートが必ずしも同期的に進行しない音楽。

2.1 音楽構造と表情付けの基本ルール

以下に示す音楽構造に対して、演奏ルール (2.4 節) を適用することによって表情付けを行う。これらの内、演奏ルールに必要なグループ構造や頂点の抽出は、我々の別研究によって、ある程度の自動解析が可能である [橋田 04]。将来的にはすべての構造の自動解析を行うことが課題になるが、現時点ではユーザが与えるようにしている。

[音楽のパート] 楽曲を複数のパート (旋律) として取り扱う。

[グループ構造] パート毎に与える。各グループには、演奏ルールにおけるフレージングのための頂点を与えられる。

[アテンション] 時間軸上において最も優勢なパートとして考えられる音列に与える (2.2 節)。

[緊張 - 弛緩構造] 隣接する音または音列の音楽的な緊張 - 弛緩関係を階層的に表したもの。パート毎に延長簡約 [Lerdahl 83] を行うことで付与する。

[表情付けの基本ルール] 上記の各構造に対応する。機能面から整理すると (1) グループ構造を明確化するためのルールと (2) タメの表現に関するルールに大別される (2.4 節)。

2.2 アテンション

アテンション (注意) とは、感受する刺激の中から一部のみを選択し、活動の焦点を絞り込むことを指す。特に、視覚領域の認知心理学において研究が盛んであり、例えば、絵画を見た時、主観的には絵の全体を眺めたつもりでも、実際にはその情景のうち数箇所の小さい部分にしか注目していないということが知られている [Carter 03]。

音楽においては休符や長音を含め常時主旋律に対してアテンションが向けられていると考えられてきた。事実、演奏の表情付けシステムはその観点で構成されることがほとんどであった。しかし、我々は同期処理を考えるにあたり、アテンションが複数のパート間で頻繁に移動していると考えたことにした。基本的に時間的な空白区間を作らないという前提のもと、当該時刻においてプライマリに聞こえてくる音の列に対し、アテンションがあるという情報を付与することとした (図 1)。

2.3 同期処理

PopE では、それぞれの声部に対し個別に演奏表現を施す。この考え方は直感的にも受け入れやすいものであるが、このままでは全体的には全く統制が取れない演奏となってしまう。ここでは、(1) アテンションが付与されているグループのテン

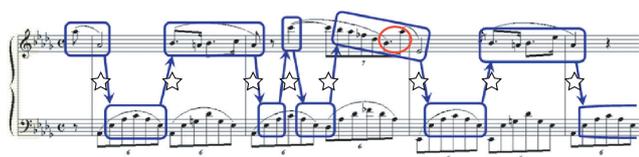


図 1: アテンションの移動

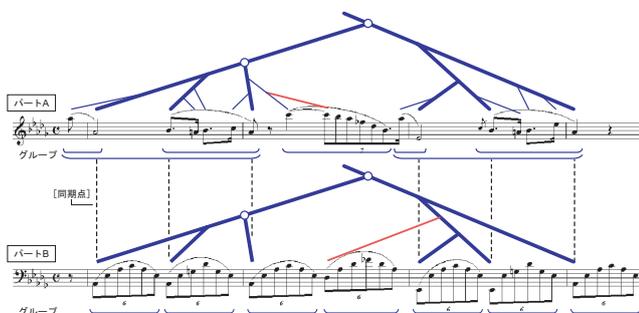


図 2: 同期点の算出: 木構造は GTTM の延長簡約による緊張 - 弛緩構造を表す。2 小節目第 2 拍は、グループ構造および緊張 - 弛緩構造の階層レベルとも不一致となることから同期点とならない。

が推移を楽曲全体のテンポ推移とする、(2) パート間のグループ構造および緊張 - 弛緩構造を比較し、緊張 - 弛緩構造の階層レベルが同じ、かつ、拍節以上で一致する場所を同期点とし、その部分においては、同時に発音させるものとする(図 2)、(3) 同期点に挟まれた区間において、アテンションが付与されたパートを基準として、演奏ルール適用後の IOI 比を保ったままスケールリングするという手続きでの同期処理を行う。

2.4 演奏ルール

表現力を豊かにするという目的を持ってルールベースによる表情付けシステムを構築する場合、通常、できるだけ多くの演奏ルールを用意する。PopE の構成においては、多くのルールを用意するのではなく、本質的な数少ないルールのみを取り扱うという考え方にに基づき、大きく分けて以下に示す二つのルールセットを用意している。

[グループ構造の表現ルール]演奏家は自身の理解した音楽的構造を明確化するという意図をもって演奏表現を行う[小澤 99]。その中でもグループ構造の表現は根幹的なものであるが、ここでは、代表的な表現手法として次の 2 つの演奏ルールを用意した。グループは階層構造として記述される。階層毎にグループ構造の表現ルールを適用するものとする。

グループ開始音に対するアクセント付与:グループ開始音の音量および音長を、初期値に対する倍率を与える。

フレージング:グループ中の一音を頂点(延長簡約の二重弛緩点によって与える)とし、グループに含まれる音の音量および局所テンポを、グループ開始音から頂点にかけて直線的に増加させ、頂点からグループ終了音にかけて直線的に減少させる[保科 98]という定式化を行った。ここでの増加率(と減少率)が演奏ルールのパラメータである。

[タメの表現]このルールの設定の意味については 3.2 節にて検討を実施する。

装飾音、連符音:演奏表現において、装飾音や連符音に対しては、音長をのばして局所的にテンポを遅くしたり、音量を強めたりすることで、この部分の演奏を強調することが多い。対応

表 1: 演奏ルールのパラメータ

パラメータ	音量 (velocity)	音長 (length)
初期値 (演奏表情なし)	パートA: 64 パートB: 48	BPM: 112
グループ開始 アクセント	パートA: 130% パートB: 150%	100% (変化させず)
フレーズ表現 (頂点の高さ)	階層 1 (小): 125% 階層 2 (大): 150%	100% (変化させず)
タメ	装飾音 連符音	パートA: 200% パートB: 100%
	アテンション の移動	パートA→B: 200% パートB→A: 125%

する規則として、これらの音の音量および音長の初期値に対する倍率を与える。

アテンション移動:アテンションが別のパートに移動する際、アテンションが移動した音(図 1 中の印)に対してタメを与えるものとして音長比率を与える。パラメータについては、グループの階層レベルやつながりごとに設定する。

3. 演奏記述能力に関する検討

この章では(1)少数のパラメータ設定による演奏の主観的評価(2)PopEにおける特徴的な演奏ルール「タメ」に関する検討(3)3人のピアニストの演奏を題材とした、PopEの演奏記述再現能力に関する評価について述べる。

演奏生成対象としては、従来の表情付けシステムが取り扱ってこなかったショパンの「幻想即興曲」展開部から 12 小節とした。この部分は、2 つのパート A, B があり、全パートの発音時刻箇所 204 に対し、両パートが一致するのは 37 箇所である。3 拍子系(六連符)のパート B に対して、パート A は、2 拍子系の音型や、直前のタイからつながる複雑な七連符が含まれる。この部分の演奏においては、パート毎の適切な表現とパート間の連携が必須であり、モデルの機能を確認するには適切な題材である。

3.1 少数のパラメータ設定による演奏の主観的評価

PopE による演奏*2を NIME-Rencon (2004 年)[野池 05]に出品した。演奏ルールについては、できるだけシンプルなものにするという目標のもと、経験的に調整を行った結果、グループ開始アクセントにかかわる音量表現(2パート)、フレージングにかかわる音量表現(2階層レベル)、装飾音・連符音にかかわる音長表現(1パート)、アテンション移動時の音長表現(2レベル)の、計 7 つに限定した(表 1)。ただし、試聴の結果短く聞こえた二ヶ所(図 1 の 2 小節目に示した丸印)については、後処理として手作業による修正(Bb 音の音長を 150%、および Ab 音の音長を 250% 延長)を行った。

NIME-Rencon においては、試行的なチューリングテストとして人間の演奏を交えての聴き比べが行われた。我々のエントリーは「演奏の人間らしさ」「演奏の好み」の評点において、人間の演奏と比べての有意差は検出されず、NIME-Rencon での Rencon Award (第一位)に選ばれた。また、後処理の修正を行ったものを行わなかったものの両者を出品したが、聴き比べの結果では、両者に有意な差があるとは判断されなかった。ここでを行った後処理の修正の意味については次節で取り上げる。

3.2 「タメ」に関する検討

PopE では、もともとテンポを支配するパート設定を目的としてアテンションを導入した。アシュケナージ、中村紘子ら

*2 <http://www.m-use.net/research/PopE/>

表 2: 各演奏に対する PopE 演奏パラメータ: 値はいずれも基本値に対する割合を示す.

	F. Heming		H. Nakamura		V. Ashkenazy	
	音量 (velocity)	テンポ (length)	音量 (velocity)	テンポ (length)	音量 (velocity)	テンポ (length)
Part A						
default velocity	40	-	70	-	48	-
phrase (Lv1)	1.5	1.0	1.2	0.8	1.1	1.0
phrase (Lv2)	1.5	1.0	1.0	1.3	1.7	1.0
group start	1.2	0.8	1.2	1.1	1.25	1.3
triple	1.0	1.0	1.0	1.0	1.0	1.3
grace	1.1	0.7	0.9	1.3	1.1	0.7
attention (in)	1.8	1.1	1.4	1.0	1.2	1.1
attention (out)	1.3	1.5	1.3	1.6	1.2	1.3
leap (up)	1.0	2.7	1.1	2.7	1.2	3.0
leap (down)	1.0	3.4	1.2	5.5	1.2	4.0
Part B						
default velocity	38	-	45	-	35	-
phrase (Lv1)	1.0	0.8	1.0	1.0	1.2	1.3
phrase (Lv2)	1.3	1.0	2.0	1.4	1.8	1.7
group start	1.1	1.2	1.2	1.2	1.1	1.0
attention (in)	1.2	1.2	1.1	1.0	1.3	1.0
attention (out)	1.1	1.1	1.1	1.5	1.1	2.0
leap (up)	1.1	1.3	1.0	1.0	1.0	1.0
leap (down)	0.9	2.0	1.0	2.2	1.0	2.0

表 3: 実演奏と再構成演奏との相関係数

演奏者	パート	音量	テンポ
フジ子・ヘミング	A	0.76	0.94
	B	0.52	0.51
中村紘子	A	0.65	0.97
	B	0.36	0.83
V. アシュケナージ	A	0.67	0.94
	B	0.70	0.69
平均	A	0.69	0.95
	B	0.53	0.68
	A&B	0.59	0.77

の演奏を分析する過程において、我々は、装飾音符や連符に加え、アテンションパートの移動部分にもタメが存在していることに気づき、その結果、タメに関する演奏ルールを構成するに至った。

装飾音符や連符については解釈上の表現に加え、早いパッセージであるが故に演奏の際の運動負荷が演奏表現に影響を与えていると考えられる。卓越した演奏者にとってこのことが技術的な制約になることは考えにくい、一般的な演奏者にとっては負荷となるはずである。また、それを受容する鑑賞者もそのストレスを共有する可能性があり、それゆえ、楽譜上の音価より長めに演奏することが一般化したという仮説を持つことができる。

一方、アテンションが移動した部分のタメについてはどうであろうか。言葉や音の途中に白色ノイズ等が挿入された場合、音の長さは、もとの音の長さより、短く感じられるという[津崎 04]。津崎らは、対象の変化を知覚する際の知覚コストが時間経過感覚をマスクするという可能性を指摘している。パート間でのアテンションの移動にもなんらかの知覚コストが伴うと考えれば、そのコストを代償するためにタメが必要であるという解釈が成り立つ。

次に、後処理にて行った手作業による二ヶ所の修正(図 1 中赤印 Bb 音および Ab 音の音長の延長)の意味について考える。NIME-Rencon のエントリにおいては、この箇所はアテンションの移動が行われない部分として取り扱った。しかし、この部分では、7 度および 11 度の跳躍進行が見られる。大きな跳躍においては同一パート中でもアテンションの移動が起こると考えることにすれば、手作業による修正部分も、統一的なモデル(アテンションの移動によるタメの挿入)による説明が可能である。以上の検討事項は現段階では仮説の域を超えない。今後は、知覚コストの代償に関する考証を進めていきたい。

3.3 パラメータ再構成による演奏再現能力の評価

演奏表現モデルを評価する手法の一つとして、個々の演奏に合うようにモデル中のシステムパラメータをチューニングし、そのパラメータによって再現された演奏を評価するという方法がある[Clynes 84, 石川 02]。ここでは、フジ子・ヘミング、中村紘子、アシュケナージによる「幻想即興曲」の各演奏を対象として、PopE による再構成を実施した。各演奏パラメータ(表 2)については、音響信号による各演奏を手作業で MIDI 情報に変換したのに対し、最小二乗近似によるフィッティングによって算出した。

図 3, 図 4 に PopE による演奏の再構成結果を示す。音量(velocity)とテンポ(length)について、元演奏と再構成演奏との相関係数(再現率)を求めたところ、全体的には音量よりテンポ、パート B よりパート A の方が高い数値が得られた(表 3)。中村の演奏において、テンポはパート A が 0.97、パート B が 0.83 となり、ヘミングおよびアシュケナージの演奏においては、ともにパート A のテンポが 0.94 となった。石川らの研究[石川 02]においては、21 個の演奏ルールを用いた最小二乗近似によるフィッティングによって、ここでの音量、テンポに対応する相関係数として、それぞれ、0.655, 0.273 を得たと報告されている。テンポに関しては、石川らの報告に比べ、PopE の相関係数が圧倒的に高い。対象曲が異なるため単純に比較することはできないが、PopE が演奏解釈の本質を捉えていることを示す一つの判断材料となろう。

聴取による比較評価においては、まず元の演奏に含まれる微細な演奏表情の違いの聴き分けができることが前提となる。一般対象者とした場合、差の判別ができないこともあるが、音楽経験者およびクラシック愛好者 10 名の被験者による評価において、全員が PopE で再構成した各演奏の元の演奏(演奏者名)を言い当て、さらに、その根拠とした部分を指摘した。テンポに比べ、音量の相関係数はさほど高くない。識別が可能であったことからすると、被験者はテンポ表現を重視して聴取を行った可能性が高い。実際、複数の被験者が「2 小節目におけるパート A の七連符から 3 小節目冒頭(図 3 および図 4 の四角で囲まれた部分)のテンポ表現に 3 人の特徴がよく表れていた」と指摘している。この部分は音楽的に最も重要とされる箇所であり、被験者も他の部分より注意して聴いたと思われる。この部分のテンポ表現は、値としては若干のばらつきがあるものの、形状は概ね再現されていると言えるだろう。

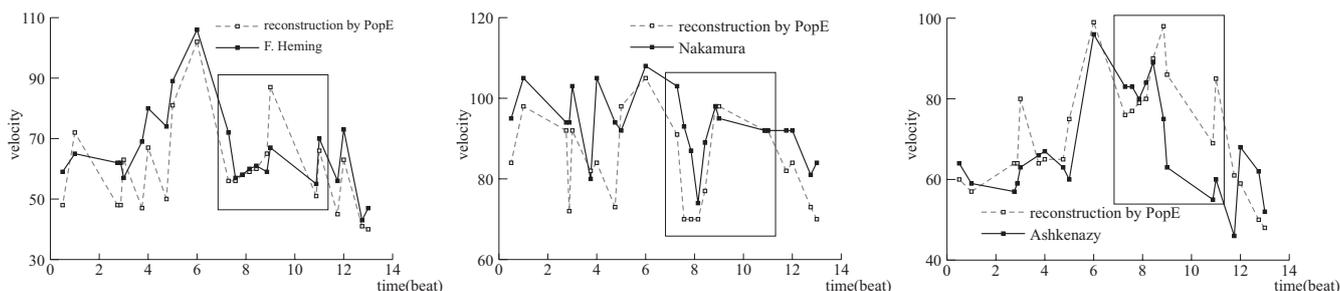


図 3: PopE による再構成 : 音量 (velocity)

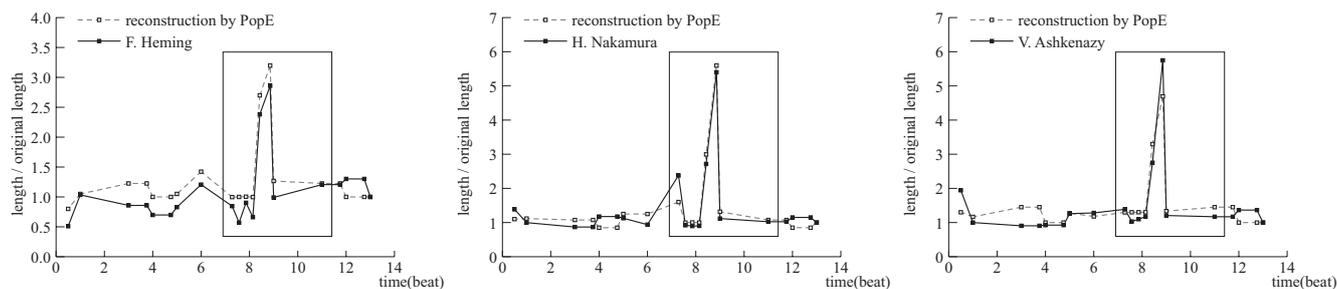


図 4: PopE による再構成 : テンポ (length)

4. まとめ

本稿では、声部協調型音楽解釈モデル PopE の演奏記述能力に関する検討を中心に紹介した。4 種類、7 パラメータの簡単な演奏ルールのみで生成した演奏が、他システムの生成演奏に勝ることが聞き比べによって実証された。また、PopE による演奏の再構成によって、3 人の演奏家による表現の特徴が反映されていることを確認した。演奏ルールの数が少ないということは、このモデルが音楽表現の本質を捉えている可能性を示唆するものである。

現段階では、楽曲構造、アテンションを人間が与えるという形での実装を行っているが、PopE ではその自動化を視野に入れ、システムティックなデザインを行っている。今後は、自動化処理を実装し、システムがユーザに候補を提示するという形のインタフェースを構築していきたい。アテンションの抽出に対してはマルチエージェントモデルの導入を検討している。その上で、様々な楽曲に対するレンダリング事例を増やしていく予定である。

参考文献

[Carter 03] Carter, R.: 脳と意識の地形図, 原書房 (2003)

[Clynes 84] Clynes, M.: A Composing program incorporating Microstructure, in *International Computer Music Conference (ICMC) Proc.*, pp. 225–232, ICMA (1984)

[橋田 02] 橋田, 野池, 平賀, 平田, 片寄: FIT 2002 RENCON Workshop — 報告と課題 —, 情処研報 2002-MUS-48, pp. 35–40 (2002)

[石川 02] 石川修, 片寄晴弘, 井口征士: 重回帰分析のイタレーションによる演奏ルールの抽出と解析, 情報処理学会論文誌, Vol. 43, No. 2, pp. 68–76 (2002)

[片寄 03] 片寄 晴弘, 奥平 啓太, 橋田 光代: 演奏表情テンプレートを利用したピアノ演奏システム: sfp, 情報処理学会論文誌, pp. 2728–2736 (2003)

[Lerdahl 83] Lerdahl, and Jackendoff, : *A Generative Theory of Tonal Music*, MIT Press (1983)

[野池 05] 野池, 橋田, 平田, 片寄, 平賀: NIME04 RENCON 開催報告と次回への課題, 情処研報 2005-MUS-59, pp. 71–76 (2005)

[Raphael 03] Raphael, C.: Orchestra in a Box: A System for Real-Time Musical Accompaniment, *Proceedings of International Joint Conference of Artificial Intelligent (IJCAI)* (2003)

[津崎 04] 津崎実, 加藤宏明: 持続時間収縮錯覚と時間間隔による文脈の関係について, 情処研報 2002-MUS-57-28, pp. 161–166 (2004)

[橋田 04] 橋田 光代, 片寄 晴弘, 野池 賢二, 保科 洋, 河原 英紀: 音楽聴取に関する一検討: グループと頂点の推定, FIT2004 情報科学技術レターズ, Vol. 3, No. 3, pp. 145–148 (2004)

[橋田 05] 橋田光代, 野池賢二, 片寄晴弘: 声部協調型音楽解釈モデルに基づく演奏の表情付け, 情処研報 2005-MUS-59, pp. 61–66 (2005)

[小澤 99] 小澤 征爾, 堤 剛, 前橋 汀子, 安田 謙一郎, 山崎 伸子 (編): 斉藤秀雄講義録, 白水社 (1999)

[片寄 04] 片寄 晴弘: 音楽生成と AI, 人工知能学会誌, Vol. 19, No. 1, pp. 21–28 (2004)

[保科 98] 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)