複数旋律音楽に対する演奏表情付けモデルの構築

橋田光代[†] 長田典子[††]
河原英紀[†††] 片寄晴弘[††]

演奏表情付けシステムの音楽表現能力は近年おおいに向上してきたが，複数の旋律からなる楽曲の演奏表現についてはあまり考慮されていなかった．本論文では，ユーザの介在を前提とした演奏デザイン支援を行うための複数旋律の楽曲に対する基礎的な演奏表現モデル **Pop-E**（Polyphrase Ensemble）を提案し，その実装例について述べる．**Pop-E** では，各旋律の演奏表現を実施するにあたり，まず，声部別に演奏ルールを適用する．この結果，複数の声部間で演奏に必要な占有時間が異なり，声部間で発音のタイミングを揃えたい時刻にもずれが生じてしまう．この対策として，グループ構造と演奏上のアテンションパートを手がかりにして必要に応じて同期させる処理手法を導入している．本モデルに基づいて生成した「幻想即興曲」は，NIME-Rencon における聴き比べコンテストにおいて第 1 位を受賞した．3 人のピアニストの演奏を対象としたモデルの再構築実験では，音量よりも音長（テンポ）に対する高い再現性が確認され，**Pop-E** によって，少数のルール群で人間に近い演奏を生成しうる可能性が示唆された．

A Performance Rendering Model for Polyphrase Ensemble

Mitsuyo Hashida,[†] Noriko Nagata,[††] Hideki Kawahara[†††]
and Haruhiro Katayose[††]

While the expressive capability of music performance rendering systems greatly improved in the last few years, the rendition of multi-part ensembles still requires more detailed consideration. This paper presents the design and implementation of **Pop-E** (Polyphrase Ensemble), a performance rendering model for multi-part music that supports performance design with intervention from the user. **Pop-E** first generates expressive phrasing for individual parts, assigning independent timings for notes in each part. The notes among different parts are synchronized to generate a coherent ensemble, using a method based on grouping structure and "attentive part" markings. The performance of "Fantaisie-Impromptu" (Chopin, middle part) generated by **Pop-E** received first prize in the hearing contest at NIME-Rencon. The paper also presents evaluation of the descriptive capabilities of **Pop-E**, based on the resynthesis of performance by three pianists. The results of this experiment, as well as the results from hearing experiments, suggest that **Pop-E** is capable of generating performer-quality renditions from a small set of performance rules.

1. はじめに

演奏表情付け（Performance rendering）に関する研究は，音楽情報処理の中でも中心的な研究領域の 1 つである[1]．演奏の表情付けシステムのさきがけとしての取り組みは，1980 年代の Frydén ら[2]，Clynes ら[3] の研究にさかのぼる．1990 年以降は，GTTM[4] や IRM[5] などの認知的音楽理論の利用，学習システム[6),7)] や事例ベース推論によるアプローチ[8] も見られるようになったのに加え，2002 年度からはシステム生成演奏の聴き比べコンテスト（Rencon☆）が開催されている[9]．

2000 年にさしかかり，システムによって生成される音楽演奏の質が年々向上し，主旋律を主軸とした表情付けについては人間に近い表現力を持ったものが出始めている．近年は特に，複数旋律の自然な演奏表現に対する注目が高まり，Raphael による Orchestra in a Box[10] や奥平らの iFP[11] が提案されている．しかしながらこれらの研究は，人間による演奏の支援を目的として，あらかじめ表情の付いた伴奏を用意してお

† 関西学院大学理工学研究科/ヒューマンメディア研究センター
Research Center for Human Media, Kwansei Gakuin University
†† 関西学院大学理工学部
School of Science and Technology, Kwansei Gakuin University
††† 和歌山大学システム工学部
Faculty of Systems Engineering, Wakayama University

☆ Rencon（Performance Rendering Contest）は，演奏表情付け研究におけるシステムや生成演奏に対する評価基盤の構築を目的として 2002 年より始動したプロジェクトである．

り，複数旋律音楽に対する演奏のモデル化を対象としていない．

本論文では，複数旋律音楽を対象とした演奏表情付けモデルとそのシステム化を取り扱う．以下，2 章では，従来研究における表情付けの現状と課題を整理する．3 章では，研究課題に対する本研究の対処について述べたうえで，複数旋律音楽のための演奏デザインについて述べる．4 章では，演奏生成処理の実装について述べ，5 章では，本モデルを用いて生成した演奏についての評価，考察について述べる．

2. 表情付けシステムの課題

情報処理の領域においては，楽譜から情緒豊かな演奏を実施する一連の過程を計算機で実現する**演奏表情付け**の取り組みがさかんに行われてきた[2),3),6)~8),12)]．2000 年以降は，システムによって生成される音楽演奏の質が年々向上し，主旋律を主軸とした表情付けについては人間に近い表現力を持ったものが出始めている．

表情付け研究における根本的な課題は演奏表現手法の定式化であるが，現時点での重要課題としては，以下の 3 つがあげられる．

(1) グループ構造解析の自動化

グループとは，1 つの声部内において，フレーズなどの音楽的なまとまりを構成する隣接した音列を指す．隣接するグループが連結し，階層的な構成になることも多い．演奏を生成するためにはグループ構造を一意に決定しておく必要があるが，現時点では，そのための決定的な理論は存在しない．GTTM[4)] をはじめとする認知的音楽理論の導入によって，ある程度の道筋は与えられたものの，個々の選好ルールの優先度制御が課題として残されている[13)]．

(2) 複数旋律音楽に対する自然な演奏の生成

これまでに多くの音楽家によって，経験的な音楽知識が演奏活動や音楽教育を通じて伝えられてきた．石川らによる研究[7)] や Widmer らのシステム[14)] のように，それらの経験的音楽知識を導入し，Rencon などで良好な評価を得てきたものもある[15)]．しかし，そのほとんどは，ショパン作品をはじめとする情緒性の高い楽曲の表現については苦手としていた．また，演奏表現の焦点が主旋律に絞られており，他の声部もあわせた表現に対する検討は不十分であった．

Orchestra in a Box[10)] や iFP[11)] などのように，ショパンの楽曲を題材としてとりあげ，高く評価されたものもある[16)]．しかし，前者は自動伴奏システム，後者は指揮システムとして実装されており，いずれも主体は人間による演奏であるうえに，各声部の微細な演奏表情はあらかじめ楽曲データに埋め込まれている．表情付けにおけるシステムの自律性を高めるためには，複数旋律音楽のための演奏モデルの構築が望まれる．

(3) 演奏デザインの効率的な支援

現在の表情付けシステム利用においては，楽譜データ以外にも，システムの使用者が，音楽構造情報（楽曲分析）と演奏情報（演奏ルール，演奏制御パラメータ）を入力する必要がある．前者は，上述のグループ構造解析の自動化に関連する事項であり，今後の重要な研究課題の 1 つである．後者については，パラメータ学習や事例利用によって自動化を目指す研究例もあるが，ユーザが自身の嗜好に沿った演奏表現のデザインを実施していくためには，ユーザがある程度システムに介入できる手段を確保しておくことが望まれる．ただし，そのために，過分に処理が煩雑になったり作業量が膨らんだりしてしまうと，使用しにくいシステムとなってしまう．ユーザの作業量を考慮しての演奏生成モデルやシステムのデザインが必要である．

3. 複数旋律音楽の演奏表現モデル Pop-E

Pop-E は，前章で述べた課題のうち，**複数旋律音楽に対する自然な演奏の生成**，**演奏デザインの効率的な支援**に焦点をあてている．本論文では，1 つの楽器で複数の声部を扱うピアノ曲を中心にとりあげるが，原理上，複数の楽器による合奏曲にも対応できる．以下，複数旋律音楽に対する自然な演奏の生成に対する対処を述べたうえで，Pop-E で利用する入力情報と処理の概要について述べる．

3.1 複数旋律音楽に対する自然な演奏の生成

Pop-E では，まず**演奏ルール**（3.3 節）の適用を通じて声部別の表情付けを行う．この処理により，複数の声部で独立の演奏表現ができ，音量やテンポに対する声部間の微細な発音タイミングのずれを生成できるようになる．ただし，この処理だけでは，そのずれが時間進行に沿って必要以上に蓄積し，演奏として成り立たなくなってしまう．そこで，蓄積したタイミングのずれを解消する手法として**声部間同期処理**（3.4 節）を導入し，声部間の微細なずれを保ちつつ演奏全体のテンポ統制を図る．以上の演奏生成の流れを**図 1** に示す．

3.2 入力情報

Pop-E はルールベース型の演奏モデルである．大規模なルール群を用意すれば表情付けの精度は向上する．しかし，演奏デザインの効率的な支援の観点からは，少数の絞り込まれた重要な演奏ルールが提供される方が望ましい．そこで，本研究では，グループの表

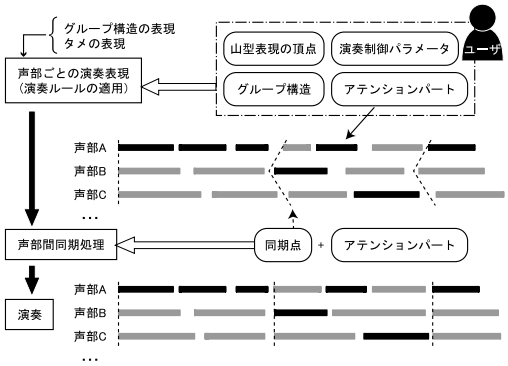

図 1 **Pop-E** による演奏表情生成の流れ
Fig. 1 Flow of performance rendering by **Pop-E**.

現とタメの表現に関するルールのみを扱うようにし，ユーザが制御すべきパラメータ数の抑制を図る．この演奏ルール適用に関して，ユーザがシステムに与える入力情報は以下の 4 項目である．

(1) グループ構造

グループ構造は，(a) 楽譜に記述された各音符の連桁，(b) スラー，(c) GTTM のグループ構造解析における局所的な階層レベルに対する判断基準[4] に基づいて決定される．具体的には，すべての音符を複数個のグループに分割し，各グループの両端となる音符に，開始音または終了音というタグを付与する．グループ構造は声部別に階層的に与えることができ，声部間では異なってもよい．同一声部内で隣接するグループに関しては，作曲者の意図として両グループの境界が重複する場合を考慮し，先行グループの開始音と後続グループの終了音を重複させてもよい．

(2) アテンションパート

複数旋律をともなう楽曲において，声部の役割にかかわらず断続的に目立って聞こえてくる音の並びをアテンションパートと定義する．**図 2** は，主旋律と他声部との掛け合いが行われている譜例である．ここでは，四角で囲まれた音符の並びのそれぞれがアテンションパートである．複数声部中のどの音が最も重要であるかを精密に判定するためには専門的な楽曲分析が必要となるが，基本的には，その楽曲を余すことなく口ずさむ際の音列として考えてよい．

(3) 山型表現の頂点

3.3.1 項で述べる．

(4) 演奏制御パラメータ

Pop-E における表情付けの処理は，入力情報が表情付けの条件に合致する楽譜上の各音符に対して，**音量**と**音長**を逐次的に強調するというプロセスを経る．音量と音長は，同一声部内のすべての音符に共通の

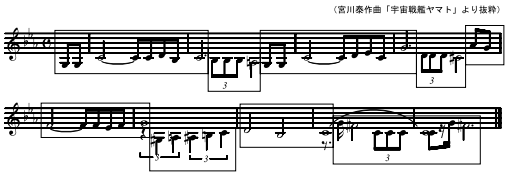

図 2 アテンションパートの例
Fig. 2 Example of an attentive part.

MIDI velocity 値，および楽譜上の音価が初期値として設定されている．演奏制御パラメータは，演奏ルールに応じて，対象となる音のそれぞれの初期値を拡大する（条件に合った音を強調する）比率を表す．

音長は，通常，その音の発音時刻を起点として，消音時刻までの時間長，あるいは同一声部上における次音の発音時刻までの時間長（Inter-Onset Interval，以下 IOI と呼ぶ）の 2 通りの扱い方がある．これらのうち，消音時刻は，レガートやスタッカートなど，音符固有の微細な演奏表現を決定するのに欠かせない情報である．しかしながら，それらの演奏表現は，ピアノ楽曲においてはペダリングのモデルをともなって初めて意味を持つと考える☆．一方，IOI は，ペダリングに関係なく，各音の発音時刻が分かれば自動的に得られる．つまり，IOI の制御は各音の発音時刻の制御と同義であり，楽曲全体に対する局所的なテンポだけでなく，2 音間の微細な ‘間’ についてもまとめて表現できるという利点がある．したがって，本論文では，音長を IOI から得るものとし．消音時刻については検討の対象から外す．

3.3 演奏ルール

3.3.1 グループの表現ルール

グループ構造に対する演奏表現は音楽演奏表現において根幹的なものである．**Pop-E** では，これまでの演奏システムが採用してきた代表的な表現手法として，以下の 2 つのグループ表現を取り扱う．

(a) グループ先頭音へのアクセント付け グループの開始音に対し音量を強め，それにあわせて，音長を伸ばす．このような発音時のアクセント付けは演奏表情付けの基本的な手法の 1 つである[17),18)]．本研究では，当該音符を演奏制御パラメータの対象とする．

(b) 山型表現 グループ中のある 1 音を**頂点**として，その音が最も強調されるように，グループ開始音から頂点までに音量と演奏速度を増加させ，頂点を過ぎればそれらを減少させるという手法（**図 3**）であり，演奏

☆ ペダリングは，それ自体が独立した研究テーマとして扱うべき重要な研究課題である．

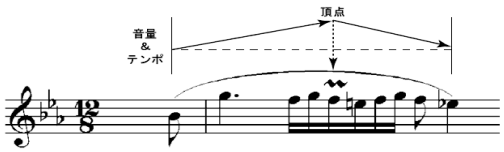

図 3 山型表現に留意した演奏表現
Fig. 3 Example of an arched phrasing expression.

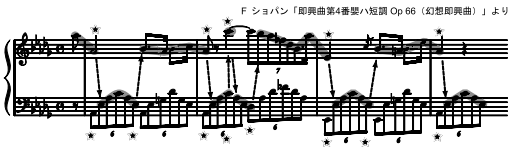

図 4 アテンションの移動
Fig. 4 Attention transition.

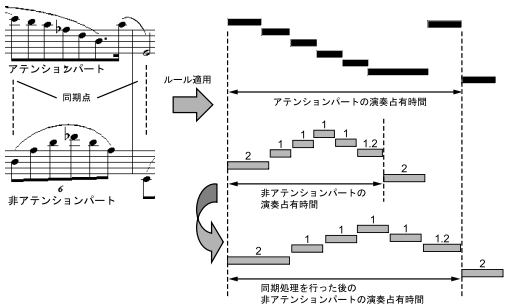

図 5 非アテンションパートのスケーリング
Fig. 5 Scaling of non-attentive part.

上のまとまり感を身体的な一息の範囲で表現するものとして，演奏家の間では特に重要視されている[19)~21)]．前処理として，グループ内の頂点音は入力情報として与えられる．演奏制御パラメータは頂点音を対象とする．頂点音以外の音符それぞれの音量・音長を拡大する比率については，グループ両端の音符（初期値）から頂点音に至るまでの直線補間によって与える．この手法は，これまでに文献 7) などにおいて採用された実績があり，おおむね良好な演奏評価が得られていると報告されている．山型表現のルールはグループ構造の階層ごとに適用する．

3.3.2 タメの表現ルール

演奏に対するタメの表現は，グループ表現と同様に，情緒豊かな演奏表現のために欠かせない重要な要素である．演奏家がタメという言葉を用いる場合，当該の音符の音長を延ばすのと，当該の音符の発音直前に時間的な間隙（‘間’）を挿入する 2 つの場合がある．本研究では，前者をタメとして取り扱う．

タメの表現の対象となる事例については，2 章で触れたように，音楽家らの経験知識が個別に語られているものの，システムへの実装を意識した定式化は進んでいない．本研究においては，定式化に向けた経験的知識を整理し，次の 3 項目を表現の対象とする．

(a) 装飾音・連符音 装飾音とは，旋律上のある音符に付け加えられた音価の短い音符☆であり，連符音とは，ある音符の時価を分割するのに特殊な方法をとった一連の音符☆☆を指す．演奏制御パラメータは，該当するすべての音符を対象とする．ただし，連符音については，旋律中単発的に現れるものを対象とする．たとえば，**図 4** において，上段の声部に現れる七連符は本ルールが適用され，下段の声部においてリズムパターンとして繰り返される六連符は適用されない．

(b) 跳躍音程 同一声部内において，隣接する 2 音の音程が大きく離れている場合，跳躍先行音を演奏制御パラメータの対象音とする．適用する音程の閾値については今後の議論が待たれるところであるが，ここでは減 5 度（半音 6 つ分）以上の音程を対象とする．

(c) アテンションの声部移動音 アテンションパートが別の声部に移動する際，移動先の音符の直前に発音された音符を演奏制御パラメータの対象音とする．この例を図 4 中に示す．矢印で結ばれた音符がアテンションの声部移動音，星印の音符がルール適用の対象音である．パラメータは，声部の組合せやグループの階層レベル別に設定してもよい．

なお，(a) と (b) については楽譜から，(c) については楽譜の声部とアテンションパートにより算出する．

3.4 声部間同期処理

3.1 節で述べたように，**Pop-E** では，メロディや伴奏といった声部の役割にかかわらず，すべての構成要素に対して声部ごとに演奏ルールをまず適用する．この処理によって，異なる声部間における発音タイミングのずれが，時間進行にともなって必要以上に蓄積されるため，楽曲全体の統制をとるには，適宜声部間のタイミングをあわせる**同期時刻**を算定する必要がある．そこで，アテンションパートの占有時間を基準として，非アテンションパートの占有時間をスケーリングすることによって，全体の演奏時間推移を調整する．この様子を**図 5** に示す．図の左側上部に示された声部がアテンションパートであった場合，スケーリングの対象は下側の声部となる．図の右側上段・中段は，アテンションパート・非アテンションパートそれぞれに対して演奏ルールを適用した結果，下段は，声部間同期処理を施した後の非アテンションパートである．

なお，同期時刻については，複数の声部において，

☆ 通常，楽譜には被装飾音符の直前に小さく描画される．
☆☆ たとえば，四分音符を三等分する場合，三連符という．

グループの開始音もしくは終了音のいずれかが楽譜の時間軸上で一致する音符の発音時刻によって与える．同期時刻上にある音符は，同時に発音させる．

4. Pop-E の実装

本モデルに基づく表情付けシステムについて述べる．本システムにおける演奏生成処理は，前向き推論エンジン OPS/R2 上で実装され，演奏ルールおよび制御モジュールから構成される．この処理に必要な入力情報は，楽譜（各音の音高と音価，楽譜上での基本発音時刻）と構造情報（グループ構造，山型表現の頂点，アテンションパート）である（**表 1**）．外部情報として，MusicXML 形式の楽譜・構造情報と演奏ルールの制御パラメータとを読み込み，それらをワーキングメモリ上に展開し，演奏ルール適用処理，声部間同期処理を経て，SMF 形式の表情付き演奏データを生成する．これらの表情付き演奏データは，市販のシンセサイザを通じて音響信号へと変換される．

4.1 演奏ルール適用処理

Pop-E における演奏表情を生成する際の中心的な制御変数は，各声部中の音符 n に対する音量 v_n と音長 l_n である．ここでの音符には休符も含む．それぞれの初期値は，同一声部内のすべての音符に共通の velocity 値と楽譜上の音価である．v_n，l_n は，当該音符に関連づけられた構造情報に適合する演奏ルール $R_{1,2,\ldots i}$ をすべて探索し，v_n，l_n のそれぞれに対して R_i が規定する音量・音長に関する拡大比率 $P_{v,n}(R_i)$，$P_{l,n}(R_i)$ を逐次乗算することで得られる．

$$v_n \text{または} \, l_n = \text{初期値} \times \prod_{i=0}^{r} P_{*,n(R_i)}$$

（ここで P_* は P_v または P_l とし，$P_{*,n}(R_i) \geq 1$）

$P_{v,n}(R_i)$，$P_{l,n}(R_i)$ は，本システムにおける演奏制御パラメータとして，外部ファイルに記述される．

山形表現のルール R_m における各音の拡大比率 $P_{v,n}(R_m)$，$P_{l,n}(R_m)$ は，図 3 に示すように線形補間値を用いる．頂点となる音符を a，各音の発音時刻 t_n，当該グループの頂点時刻 t_a，当該グループ裾野の時刻 t_f（$t_n \leq t_a$ のときはグループ開始音の発音時刻，$t_n > t_a$ のときはグループ終了音の発音時刻を表す．時刻はいずれも楽譜上でのもの）として，以下の式で計算される．

$$P_{*,n}(R_m) = \frac{(t_n - t_f) \times P_{*,a} + (t_a - t_n)}{t_a - t_f}$$

（ここで P_* は P_v または P_l）

表 1 **Pop-E** による演奏生成で必要な楽曲情報
Table 1 Score information for **Pop-E**.

楽譜	構造情報
音高	グループ構造
音価	山型表現の頂点
発音時刻	アテンションパート

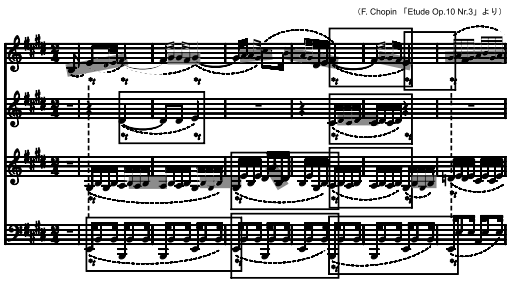

図 6 声部間同期処理の対象例
Fig. 6 Example of attentive part and synchronizing point.

4.2 声部間同期処理

上記の処理により，各音符の音長 l_n が計算されている．各音符の演奏時刻は，各声部においてそれぞれの l_n を順次加算していくことによって計算されるが，適用されるルールが異なるため，声部間の同期時刻がずれてしまっている．そこで，非アテンションパートの占有時間のスケーリングによって同期時刻を揃える．

図 6 を例に示す．図中，薄く色づけされた音符列がアテンションパート，記号（*t）で記された音符の楽譜上の発音時刻が同期時刻，四角で囲まれた範囲がスケーリングの対象を表す．各対象範囲に対し，アテンションパートの各音の音長を加算しておくことにより，その区間の占有時間 S_a を求める．続いて，非アテンションパートにおける音長を加算し，その占有時間 S_v を求める．そのうえで，S_a/S_v を非アテンションパートの各音の音長に乗ずる．以上の処理を，すべての同期時刻の発音時刻が一致するまで実行する．

5. 評　　価

5.1 生成演奏に対する聴取評価

これまでに，**Pop-E** を用いて**表 2** にあげた楽曲の演奏生成を実施した．このうち， F. ショパン「即興曲第 4 番嬰ハ短調 “幻想即興曲” Op.66」を NIME-Rencon [22] に出品し，聴き比べの評価を行った．以下，この NIME-Rencon における評価状況について述べる．

5.1.1 コンテスト実施内容と結果

NIME-Rencon コンテストのうち，筆者らが参加した規定部門では，5 つのシステムの演奏と，人間が弾

表 2 生成楽曲
Table 2 Pieces rendered by **Pop-E**.

作曲者名	楽曲名（いずれも抜粋）
モーツァルト	メヌエット K.1(1e)
ショパン	即興曲第 4 番嬰ハ短調 Op.66 「幻想即興曲」
チャイコフスキー	「白鳥の湖」より第 2 幕のアダージョ
ラフマニノフ	パガニーニの主題による狂詩曲

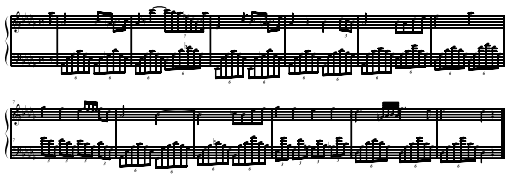

図 7 F. ショパン「即興曲第 4 番嬰ハ短調 “幻想即興曲” Op.66」より展開部
Fig. 7 Middle part of Chopin's “Fantaisie-Impromptu”.

いた 2 曲分の演奏とをあわせた 7 演奏に対し，演奏とシステム（または人間）との対応関係を伏せたブラインド方式での聴き比べが行われた．評価方法は，(1) 演奏の人間らしさ，(2) 演奏の好みの 2 項目に対する 5 段階評価の投票であり，各項目の平均点が評価値とされた．各演奏は (1) に対する評価値（同点の場合は (2) に対する評価値）によって順位付けされた．聴取者による最終的な有効投票数は 51 であった．

投票の結果，人間による 2 演奏，**Pop-E** による演奏が高い評価値となった．**Pop-E** による演奏は，システムによる生成演奏の中で最高評価を得て，Rencon Award を受賞した．

5.1.2 エントリ曲

筆者らがエントリした「幻想即興曲」（**図 7**）は 2 つの声部 A（右手），B（左手）からなり，すべての音符（204 音）の各発音時刻に対し，声部間において発音時刻が一致するのは 37 カ所である．1 拍（四分音符）を三連符で表す声部 B に対し，声部 A は，1 拍を八分音符 2 つで表す音型を基本とし，単発的な三連符や，タイからつながる複雑な七連符が含まれる．演奏においては，2 小節目におけるテンポの揺らぎが大きいことが知られており，演奏表現モデル **Pop-E** の能力を確認するには適切な題材といえる．

この演奏生成においては，なるべく単純なパラメータセットを用いることを念頭に置いた．演奏ルールの適用☆に際しては，音量表現はグループ表現のルールのみ，音長表現はタメの表現ルールのみで行った．グループの階層は 2 レベルとし，山型表現に対するパラメータは両声部に共通とした．アテンションの声部移動音に対するパラメータはグループの階層レベルにかかわらず共通とした．装飾音・連符音のルールについてもそれぞれ共通のパラメータを用いた．

☆ 3.3 節で述べたルールのほか，ペダリングに関する簡易な演奏ルールを用意した．声部 B の各グループの開始音において，ペダル値（MIDI コントロール・チェンジ 64）を，発音時刻で 0，消音時刻で 127 になるように 5 段階で線形補間した．このルールに関しては，演奏制御パラメータは存在しない．

表 3 「幻想即興曲」に付与した演奏制御パラメータ
Table 3 Control parameters for “Fantaisie-Impromptu”.

	声部 A（音量/音長）	声部 B（音量/音長）
初期値	64/—	48/—
グループ先頭音へのアクセント付け	1.3/(1.0)	1.5/(1.0)
山型表現 （階層 1）	1.25/(1.0)	
（階層 2）	1.5 /(1.0)	
タメ 装飾音・連符音	(1.0)/2.0	—
アテンション移動	(1.0)/2.0	(1.0)/1.25
跳躍音程（上行）	(1.0)/1.5	(1.0)/1.0
跳躍音程（下行）	(1.0)/2.5	(1.0)/1.0

() は設定対象外とした項目.

設定した演奏制御パラメータの一覧を**表 3** に示す．表中，1.0 とは音量ないし音長を変化させないことを表す．声部 B における装飾音・連符音については，楽譜中に該当する音符が存在しないため設定の対象から除外された．基本的な演奏速度（テンポ）は，1 分間あたりの四分音符発音数（BPM: Beat per Minute）を 112 とした．装飾音符の楽譜上の音価については，楽典上の規定がなく，MusicXML にも記述されていないため，便宜的に三十二分音符（上述の演奏速度のときに約 67 ミリ秒）と等価とした．

最終的に，この演奏生成では，計 7 種の演奏ルールと 11 の演奏制御パラメータが用意され，このうち 9 パラメータが変更された．この演奏は http://www.m-use.net/research/PopE/ から試聴できる．

文献 22) による「演奏の人間らしさ」についての聴き比べ評価をもとに，**Pop-E** による演奏と各演奏との間で検定を行った（**図 8**）．各演奏の評価値については，**Pop-E** による演奏と他のシステム演奏との間で 1%水準での有意差が確認された．一方，人間の演奏との間では有意差はなかった．なお,「演奏の好み」についても「演奏の人間らしさ」とほぼ同様の結果であった．つまり，この聴き比べにおいて，**Pop-E** による演奏は，人間の演奏とほぼ同レベルの表現を行うことができたといえる．

5.2 演奏再現能力の検討

演奏表現モデルを評価する別の手法の 1 つに，システムの演奏制御パラメータのチューニングによって，与えられた演奏がどの程度再現されるかを評価するという方法がある[3),7)]．ここでは「幻想即興曲」を対象

表 4 各演奏者の再構成演奏に設定された演奏制御パラメータ
Table 4 Control parameters to three performances of reconstruction by **Pop-E**.

	H		N		A	
	声部 A（音量/音長）	声部 B（音量/音長）	声部 A（音量/音長）	声部 B（音量/音長）	声部 A（音量/音長）	声部 B（音量/音長）
初期値（velocity）	40/—	38/—	70/—	45/—	48/—	35/—
グループ先頭音へのアクセント付け	1.2/0.8	1.1/1.2	1.2/1.1	1.2/1.2	1.25/1.3	1.1/1.0
山型表現 （階層 1）	1.5/1.0	1.0/0.8	1.2/0.8	1.0/1.0	1.1/1.0	1.2/1.3
（階層 2）	1.5/1.0	1.3/1.0	1.0/1.3	2.0/1.4	1.7/1.0	1.8/1.7
装飾音	1.1/0.7	—	0.9/1.3	—	1.1/0.7	—
連符音	1.0/1.0	—	1.0/1.0	—	1.0/1.3	—
アテンション移動 A to B	1.8/1.1	1.2/1.2	1.4/1.0	1.1/1.0	1.2/1.1	1.3/1.0
B to A	1.3/1.5	1.1/1.1	1.3/1.6	1.1/1.5	1.2/1.3	1.1/2.0
跳躍音程 上行	1.0/2.7	1.1/1.3	1.1/2.7	1.0/1.0	1.2/3.0	1.0/1.0
下行	1.0/3.4	0.9/2.0	1.2/5.5	1.0/2.2	1.2/4.0	1.0/2.0

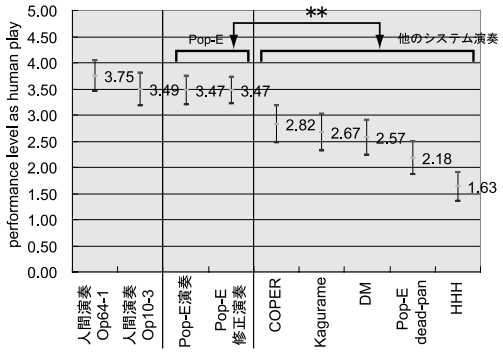

図 8 NIME-Rencon における「演奏の人間らしさ」評価（文献 22) から引用．$**: p < 0.01$）
Fig. 8 Evaluation of performance level as human play at NIME-Rencon Workshop ($**: p < 0.01$).

曲として，3 人のピアニストによる各演奏に対して，**Pop-E** を通じて演奏を再構成し，聴取比較評価と相関係数による評価を行った．

演奏データの生成にあたっては，各演奏の音響信号を手作業で MIDI 情報に変換したものを実演奏☆とし，音量および音長に対する実演奏との相関係数が最大となるような演奏制御パラメータを設定したうえで，**Pop-E** を通じて生成した演奏を再構成演奏とした．この際，山型表現の頂点音については個々の実演奏に最も符号するよう調整したものを用いた．各実演奏に対し設定した演奏制御パラメータを**表 4** に示す．以降，各実演奏を H，N，A，それぞれの再構成演奏を h，n，a と表記する．

聴取比較評価における前提として，被験者は演奏によって異なる微細な表現の差異を聴き分けられる必要がある．一般の聴取者を対象とすると，演奏の違いの

☆ MIDI データとして直接記録されたものではないため，厳密な意味での実演奏ではない．

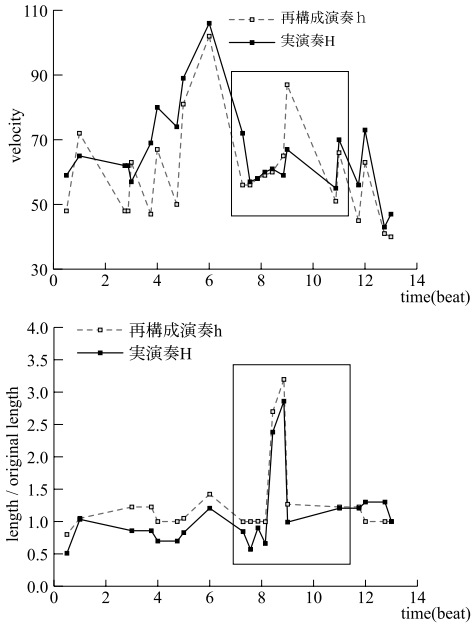

図 9 実演奏 H と再構成演奏 h（上：音量，下：音長）
Fig. 9 Performances of original H and reconstruction h.

判別が困難であることもあるため，ここでは，音楽経験者およびクラシック音楽の愛好者 10 人を被験者とした．演奏者と実演奏との対応関係を開示したうえで，各演奏を繰り返し聴取させ，3 人の実演奏の表現の差異を聴き取れるか，各実演奏と再構成演奏がどのように対応するかについてインタビューを行った．

その結果，全員が，実演奏における 3 人の演奏表現の違いを聞き分け，再構成演奏と実演奏との対応をいい当てることができた．そのうち，5 人の被験者は，再構成演奏に対して「声部 A の 2 小節目七連符から 3 小節目冒頭 A♭ 音のテンポ表現に 3 人の特徴がよく表れていた」と指摘した．

図 9，**図 10**，**図 11** に，声部 A における各再構成

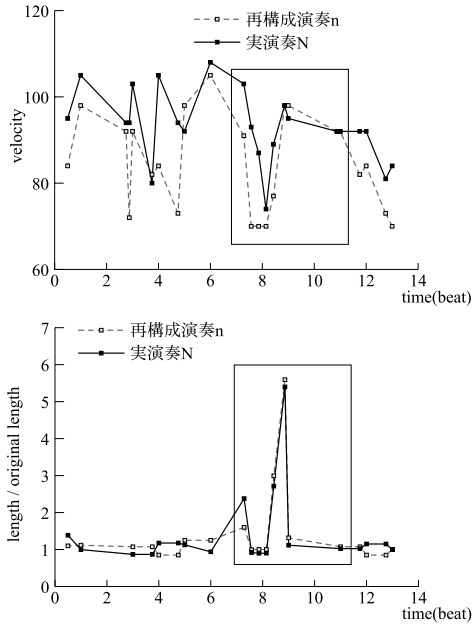

図 10 実演奏 N と再構成演奏 n（上：音量，下：音長）
Fig. 10 Performances of original N and reconstruction n.

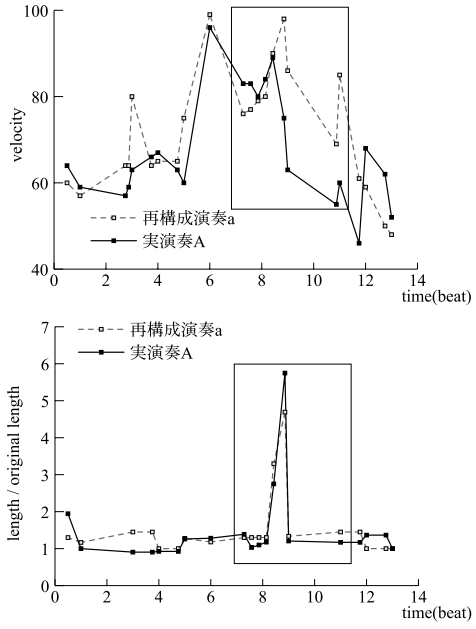

図 11 実演奏 A と再構成演奏 a（上：音量，下：音長）
Fig. 11 Performances of original A and reconstruction a.

演奏を示す．各図中，四角で囲まれた範囲が，3 人の被験者によって指摘された部分である．再構成演奏と実演奏との相関係数（**表 5**）では，声部 A の音長が，声部 B の音長および声部 A の音量より明らかに高い値となった．声部 A の音量についても，同一演奏者に対する相関係数が，声部 B あるいは他の演奏者の音量

表 5 実演奏と再構成演奏との相関係数
Table 5 Correlation coefficient between performances of the original all players and the reconstruction all players[e].

再構成演奏	同一演奏者の実演奏（音量/音長）	異なる演奏者の実演奏（音量/音長）
声部 A	0.70/0.91	0.37/0.86
声部 B	0.53/0.68	0.41/0.38
A&B	0.59/0.76	0.40/0.55

表 6 再構成演奏 h と実演奏との相関（音量/音長）
Table 6 Correlation coefficient between performances of the original H and the reconstruction h.

h	H	N	A
声部 A	0.76/0.94	0.55/0.86	0.37/0.88
声部 B	0.52/0.51	0.20/0.26	0.54/0.63
A&B	0.61/0.62	0.32/0.47	0.48/0.72

表 7 再構成演奏 n と実演奏との相関（音量/音長）
Table 7 Correlation coefficient between performances of the original N and the reconstruction n.

n	N	H	A
声部 A	0.65/0.97	0.60/0.79	0.00/0.97
声部 B	0.36/0.83	0.44/0.17	0.68/0.49
A&B	0.48/0.89	0.50/0.39	0.44/0.66

表 8 再構成演奏 a と実演奏との相関（音量/音長）
Table 8 Correlation coefficient between performances of the original A and the reconstruction a.

a	A	H	N
声部 A	0.66/0.94	0.41/0.78	0.28/0.91
声部 B	0.70/0.69	0.42/0.30	0.20/0.43
A&B	0.69/0.77	0.23/0.60	0.48/0.72

と比べて最も高い 0.70 となっている．さらに，演奏家別に算出した再構成演奏と実演奏との相関係数（**表 6**，**表 7**，**表 8**）においては，声部 A の音長が，h，n，a のいずれも 0.94 以上となり，音量や声部 B に対して明らかに高い．以上の結果は，被験者らが，声部 A のテンポ表現を指摘したことを支持するとともに，各再構成演奏が，特に声部 A の音長について，対応する実演奏を良好に再現できたことを支持すると考えられる．

一方で，音量や声部 B に対する相関係数は，音長に対する値に比べて高いとはいえない．声部 B では類似した音型が繰り返されており，適用される演奏ルールの数が少数に限られたことがその要因である．演奏ルール・制御パラメータの絞り込みには反するが，音型の繰返しに関する演奏ルールについては今後検討していく必要があろう．

数値上の比較としては，石川らが，21 種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，本論文の音量・音長に対応する相関係数として，それぞれ，0.655，0.273 を得ている[7]．**Pop-E** は 9 種の演奏ルールを声部別に振り分けた 18 の演奏制御パラメータを用いて，音量・音長に対応する相関係数 0.59，0.76（表 5）が得られ，音長については石川らの報告した精度を上回った．対象曲が異なるため単純な比較はできないが，この結果は，演奏ルールの種類が少なくても，演奏解釈の本質をとらえうることを示す 1 つの判断材料となると考えている．

また，前述したように，声部 A の音長を除く実演奏と再構成演奏との相関は必ずしも高いとはいえないが，この精度の演奏でも，被験者らは各演奏を区別できた．前述の結果をいい換えると，被験者らは，この楽曲に対して，音量よりも音長（テンポ表現）を手がかりとして演奏表現を推定したと思われる．このことを確認するために，a と h の演奏制御パラメータを 10 段階に内挿した演奏を新たに生成し，4 人の被験者（3 人は音楽経験者，1 人は未経験者）を対象に，どちらの演奏に近いと判断されるかについての聴取判別実験を行った．この結果，すべての被験者に共通して識別空間は，テンポ優位で構成されることが確認された．この結果が他の楽曲に対してもいえるかについては議論の余地があるが，今後，より効率的な演奏デザイン支援を進めていくうえで，聴取者の演奏に対する聴き方や，音量表現とテンポ表現とのバランスについても検討する必要があるだろう．

6. ま と め

本論文では，(1) 複数旋律に対する自然な演奏表現の生成，(2) 演奏デザインの効率的な支援を目標とした複数旋律音楽の演奏表現モデル **Pop-E** について述べた．本システムの利用にあたって，ユーザは，あらかじめグループや頂点・アテンションパートなどの構造情報を入力する必要があるものの，演奏ルールを厳選したこと，グループ構造に基づいた声部間同期処理を行うことにより，全体としては，直感的な表情付け手法による複数旋律の自然な演奏表現を生成することが可能となった．

提案手法によって生成した演奏は Rencon Award の対象となった．また，3 人の演奏表現の違いがモデル上で表現できたことから，**Pop-E** は演奏表現の本質の一端をとらえているものと思われる．

今後は，本論文では取り扱わなかったユーザによる音楽構造情報に対する入力支援ツールや，消音時刻とペダリングの制御機構の実装を進める予定である．

謝辞 この研究は，科学技術振興機構さきがけ研究 21「協調と制御」領域の支援を受けて実施されました．現在は，CrestMuse プロジェクトの研究として実施されています．

参 考 文 献

1) 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).
2) Frydén, L. and Sundberg, J.: Performance Rules for Melodies. Origin, Functions, Purposes, *Proc. Intl. Computer Music Conf.* (*ICMC*), pp.221–224 (1984).
3) Clynes, M.: A Composing Program Incorporating Microstructure, *Proc.Intl.Computer Music Conf.* (*ICMC*), pp.225–232 (1984).
4) Lerdahl, F. and Jackendoff, R.: *A Generative Theory of Tonal Music*, MIT Press (1983).
5) Narmour, E.: *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*, the University of Chicago Press (1991).
6) Widmer, G.: Learning Expressive Performance: The Structure-Level Approach, *Journal of New Music Research*, Vol.25, No.2, pp.179–205 (1996).
7) 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.268–276 (2002).
8) Arcos, J., de Mantaras, R. and Serra, X.: SaxEx: A Case-Based Reasoning System for Generating Expressive Musical Performances, *Journal of New Music Research*, Vol.27, No.3 (1998).
9) 平賀瑠美，平田圭二，片寄晴弘：蓮根：めざせ世界一のピアニスト，情報処理，Vol.43, No.2, pp.136–141 (2002).
10) Raphael, C.: Orchestra in a Box: A System for Real-Time Musical Accompaniment, *Working Notes of IJCAI-03 Workshop on methods for automatic music performance and their applications in a public rendering contest* (2003).
11) 奥平啓太，片寄晴弘，橋田光代：音楽演奏インタフェース iFP—演奏表情のリアルタイム操作とビジュアライゼーション，情報処理学会研究報告音楽情報科学 2003-MUS-51, Vol.82, pp.39–44 (2003).
12) Bresin, R. and Battel, G.: Articulation Strategies in Expressive Piano Performance. Analysis of Legato, Staccato, and Repeated Notes in Performances of the Andante Movement of Mozart's Sonata in G Major (K. 545), *Journal of New Music Research*, Vol.29, No.3, pp.211–224 (2000).

13) Hamanaka, M., Hirata, K. and Tojo, S.: Automatic Generation of Grouping Structure based on the GTTM, *Proc. Intl. Computer Music Conf.* (*ICMC*), pp.141–144 (2004).
14) Widmer, G. and Tobudic, A.: Playing Mozart by Analogy: Learning Phrase-level Timing and Dynamics Strategies, *Proc. ICAD 2002 Rencon Workshop*, pp.28–35 (2002).
15) 橋田光代，野池賢二，平賀瑠美，平田圭二，片寄晴弘：FIT 2002 RENCON Workshop—報告と課題，情報処理学会研究報告音楽情報科学 2002-MUS-48, Vol.123, pp.35–40 (2002).
16) 片寄晴弘，平田圭二，平賀瑠美：IJCAI-RENCONの報告と課題，情報処理学会研究報告音楽情報科学 2003-MUS-52, Vol.111, pp.149–152 (2003).
17) Snyder, B.：音楽と記憶：認知心理学と情報理論からのアプローチ，音楽之友社 (2003).
18) Aiello, R. (Ed.)：音楽の認知心理学，誠信書房 (1998).
19) 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).
20) 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).
21) 柴田南雄，遠山一行（総監修）：ニューグローヴ世界音楽大事典より［楽句］，講談社 (1993).
22) 野池賢二，橋田光代，平田圭二，片寄晴弘，平賀瑠美：NIME04 RENCON 開催報告と次回への課題，情報処理学会研究報告音楽情報科学 2005-MUS-59, Vol.14, pp.71–76 (2005).

(平成 18 年 5 月 8 日受付)
(平成 18 年 10 月 3 日採録)

橋田 光代（正会員）

国立音楽大学音楽デザイン学科卒業．平成 13 年同大学院音楽研究科音楽学専攻音楽デザインコース修了．SIGGRAPH'98 Television, ICMC'99, June in Buffalo 2001 にて作品入選．平成 18 年和歌山大学システム工学研究科博士後期課程修了．修士（音楽），博士（工学）．現在，関西学院大学理工学研究科/ヒューマンメディア研究センター博士研究員．音楽情報科学，音楽理論，コンテンツ Design & Creation の研究に従事．感性工学会会員．

長田 典子（正会員）

昭和 58 年京都大学理学部数学系卒業．同年三菱電機（株）入社．産業システム研究所を経て，平成 8 年大阪大学大学院基礎工学研究科博士課程修了．工学博士．平成 15 年より関西学院大学理工学部情報科学科助教授．専門はメディア工学，感性情報処理等．電子情報通信学会，電気学会，日本認知心理学会，日本顔学会，IEEE 等各会員．

河原 英紀（正会員）

昭和 52 年北海道大学大学院工学研究科博士課程修了．工学博士．同年電電公社武蔵野電気通信研究所入所．ATR 人間情報通信研究所を経て，平成 9 年より和歌山大学システム工学部教授．聴覚機能の数理的解明と工学的表現の研究に従事．平成 9 年音響学会佐藤論文賞．平成 12 年 EURASIP 最優秀論文賞受賞．IEEE，ASA，日本音響学会，神経回路学会，認知科学会各会員．

片寄 晴弘（正会員）

平成 3 年大阪大学大学院基礎工学研究科博士後期課程修了．工学博士．オージス総研，イメージ情報科学研究所，和歌山大学を経て，現在，関西学院大学理工学部教授．ヒューマンメディア研究センターセンター長．音楽情報処理，コンテンツ Design & Creation，心理計測の研究に従事．CREST「時系列メディアのデザイン転写技術の開発（CrestMuse Project）」研究代表者．元さきがけ研究 21「協調と制御」領域研究者．