

聴取経験に基づく予測補完型音楽生成アーキテクチャに関する考察

片寄晴弘^{†,††} 豊田 健一^{†,††}

本稿では、自動作曲、編曲、パフォーマンスレンダリング等の音楽生成系主要研究を聴取経験に基づく予測補完という視点から分類・整理する。それら音楽生成処理を包括的・統合的に実現するための基本アーキテクチャについて考究する。

A Study of a Music Generation Architecture Complementing and Predicting Musical Passage based on Experience

HARUHIRO KATAYOSE and KEN'ICHI TOYODA

This paper surveys studies of algorithmic composition, arrangement and performance rendering of music. We categorize and examine nature of music production in terms of predictive complement based on listening experience. At the end of this paper, we discuss design of the system architecture to cover plural music production tasks.

1. はじめに

1956年の「イリアック組曲」¹⁾以降、非常に多く音楽生成系システムの開発・研究がなされされてきた²⁾。現在、知的音楽タスクに関連する研究領域としては、「作曲システム」「編曲システム」「表情付けシステム」「伴奏システム」「セッションシステム」などが定着しており³⁾、それぞれの領域を代表するユニーク、かつ、優れた研究が存在する。これらの研究事例を眺めてみると、例えば「アイデアを、そのまま、セッションシステムにも応用してみたい」と思わせる自動作曲システム、あるいは「このアーキテクチャであれば、伴奏システムへの応用は非常に容易であろう」と思わせる表情付けシステムなど、タスクを越えた可能性を感じさせるものが少なくない。

音楽は、その発達の過程において、作曲や演奏などタスク毎への分業が進んだ。それぞれのタスク毎に音楽システムの研究や開発が行われていることはある意味当然である。しかしながら、音楽が発生した状況や、脳内における音楽の処理を考えてみれば、領域毎にシステムの開発を行ったり、また、リアルタイムか非リアルタイムかというような切り口を持つことは、必ずしも本質的ではないと思われる。現在は、「イリアック

組曲」の時代と比べて、背景となる技術レベルも大幅に向上しており、すべての音楽生成系プロセスに適用可能な統合アーキテクチャを考慮していく条件が整いつつあると筆者らは考える。

本稿では、「作曲システム」「編曲システム」「表情付けシステム」「伴奏システム」「セッションシステム」の代表的な研究例を紹介し、聴取経験に基づく予測補完という視点から整理する。その上で、すべての音楽生成系プロセスに適用可能な包括的・統合音楽システムの実現に向けての基礎的な考察を実施する。

2. 音楽生成システムのサーベイ

冒頭でも述べたように生成系の音楽システムの研究は非常に数多く存在する。この章では、議論を進めていく上で重要と思われるものを、抜粋して紹介する。

まず、作曲、編曲に関連して、それぞれ、D. CopeのEMIシステム⁴⁾、平田のパーピーブス⁵⁾を紹介する。これらは、静的なシステムとして動作するものである。続いて、音楽の予測を扱ったものとして、D. Conklinらの音列予測システム⁶⁾、松尾らの研究⁷⁾を紹介する。次に、表情付けシステムとして、奥平らのiFP⁸⁾、C. RaphaelのOrchestra in a Box⁹⁾について紹介する。これらは、表情付けシステムの中でも、プレイヤーの意図を動的に反映できるものである。最後に、インタラクションに重点を置いた研究として、自動伴奏の視点からは、R. Dannenbergのシステム¹⁰⁾、セッションの観点からは、西嶋らのNeuro-Musician¹¹⁾、インタ

[†] 関西学院大学

Kwansei Gakuin University

^{††} 科学技術振興機構さきがけ研究 21

PRESTO, JST

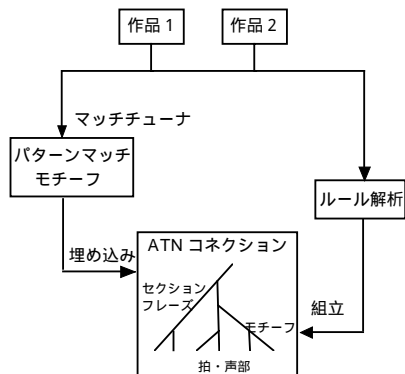


図1 EMI の概要．图中、左サイドがパターンマッチ（モチーフ抽出）プロセス、右サイドがルール解析のプロセスを示している．

ラクティブコンポージングの視点から、R. Rowe の Cypher¹²⁾、F. Pachet の Continuator¹³⁾ を紹介する．

2.1 EMI

EMI は D. Cope によって 1981 年から開始された自動作曲に関するプロジェクトである⁴⁾．Cope は「作曲とは、今までに作られた作品の事例の解析と再合成によってなされる」という考え方を基に自動作曲・編曲システムを構築した．EMI の処理概要を図 1 に示す．EMI は、大きくパターンマッチ（モチーフ抽出）プロセスと、ルール解析のプロセスから構成されている．パターンマッチのプロセスにおいては、ピッチのみ、リズムのみ、ピッチとリズムを合わせたものの 3 つの視点から、同じ、あるいは、同型と考えられるモチーフの発見を行う．一方、ルール解析プロセスではパートの進行方向、繰り返される音の数、和声の概形などの出現確率を計算する．このようにして作品の様式に関する基礎データを取得し、マルコフ遷移を木構造状に配したデータ構造を作成する．これらの基礎データを乱数を用いて、再構成を行うことで作曲が行われる．

2.2 パーピーブ

パーピーブは、2000 年頃、平田によって開発された DOO (Deductive Object-Oriented) という知識表現・推論機構を用いた、原曲のシンプルなコード進行をジャズらしいコード進行にリハーモナイズするシステムである⁵⁾．1) リハーモナイズの実際として、基本コードとその進行とテンションの組とその説明をシステムに与えて、その関係をシステムに覚え込ませる．2) リハーモナイズを行う対象の基本コード進行を与え、データベース中から最も近い、基本コード進行を持つデータを検索し、そのデータのテンションデータを転写する．この 2 つの処理でリハーモナイズを行う．ものごとを基本となる単純型とその派生型あるいは具体化実体（ここでは、基本コード名とテン

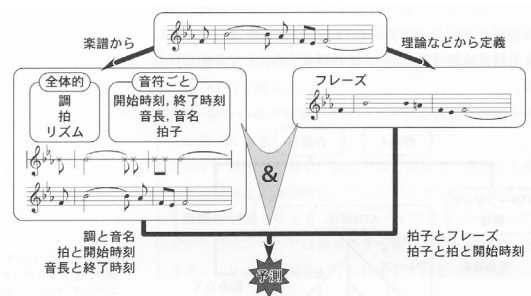


図2 Conklin らによる多視点性を考慮したメロディの予測システム

ション名) という観点から眺め、検索の際には、より具体的なものから照合を行い、見つからない場合には、より抽象度の高いレベル基本での照合を行う形で、合理的な処理が実現されている．また、コード進行を見ながら検索を行うことで、音楽的な文脈が保たれるような配慮がなされている．

2.3 Conklin らのシステム

D. Conklin らは、1994 年頃、メロディを構成するリズム、ピッチを時系列信号ととらえ、適当なスパンで区切った時系列信号とその後続信号の出現確率（推移確率）、エントロピーを計算するプロセスにより、与えた楽曲様式（Bach のコラール）のモデル化を行った⁶⁾．さらに、取得したモデルによって未知曲のメロディを予測させ、実際の進行と比較することで、システムの有効性を検証した．図 2 に示すように、リズムやピッチ以外の複数の視点を導入し、多視点性を考慮したモデル (viewpoint) を用いた．推移確率を用いる手法自体は、珍しいものではないが、予測を行う時系列スパン、複数の視点の組み合せたことにより、Bach のコラールに対する予測性能が、100% に近い値を示したことが着目される．

2.4 松尾らの研究

松尾らは、1998 年に、単旋律のリアルタイム予測を行うシステムを実装した⁷⁾．このシステムは、適当な閾値で拍の量子化を行った上で、自己回帰計算を行い、1) 類似パターンが検出された際には自己回帰計算による出力、2) 類似パターンの検出はなされなかったが、予測が当たった場合は、その予測の継続、3) 上記のいずれにも当たらなかった場合は、予め与えておいた推移率に基づき、上位 3 位までの予測を行う．このシステムでは、記憶可能数（自己回帰計算の範囲）、パターンの最小周期、最大周期、忘却曲線をユーザが与えることが出来る（図 3）．

上記の Conklin らのシステム、松尾のシステムとも、音楽のスタイルや認知に興味を持ち、そのモデル

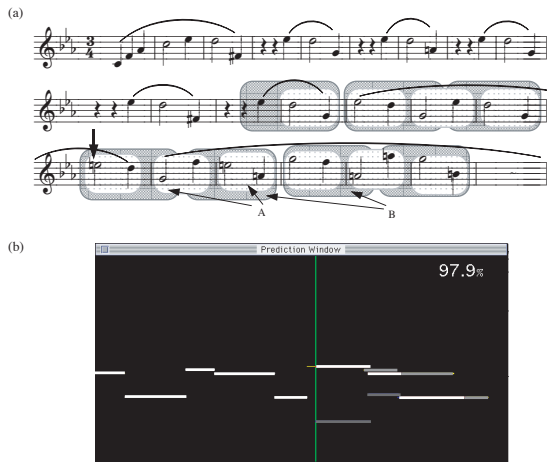


図 3 松尾らによるメロディのリアルタイム予測。(a)の矢印部分に対する予測が(b)．未来の予測は、明るい線ほど、確信度が高い予想である．この事例では、97.9%の確信度でEbを予測している．

を実装したものであるが、種になるメロディやハーモニーを与え、逐次、予測を行っていくプロセスによって作曲を行うことも可能である．

2.5 iFP

iFP は、2003 年に奥平らによって実装された、名演奏における elaboration 事例をテンプレートとして利用する演奏インタフェースである⁸⁾．iFP は、指揮ジェスチャ、あるいは、拍打によって与えられるプレイヤーの演奏意図とテンプレート上に記載されるモデルを実時間で合成し演奏を生成する．モデル中の微細な演奏表現が反映されるため、プレイヤーは自身の動きかけを伴って、例えば、ブーニンやアルゲリッチらの名演奏を「なぞる」ことが可能となる．

iFP のスケジューラは予測型のものとして構成されており、プレイヤーからの入力がない場合は、モデルを参照して自走的に演奏が進む．すなわち、プレイヤーは、関与したい部分だけでの介入が可能である．この視点での応用として、iFP を自動伴奏システムとして利用することが可能である．また、演奏モデルのうち、テンポ、拍音量、拍内の微細な演奏表情の3つに対して、どのくらいの重みを持たせて利用するか、また、プレイヤーの意図とモデル上の意図の、どちらをどの程度優先させるかのパラメータを実時間で制御することができる(図4)．

2.6 Orchestra in a Box

Orchestra in a box は、C. Raphael によって、2002 年に、実装されたリアルタイム伴奏システムであり、以下の3つの処理部から構成される⁹⁾．

- Listen: 演奏の音響信号を分析し、隠れマルコフ

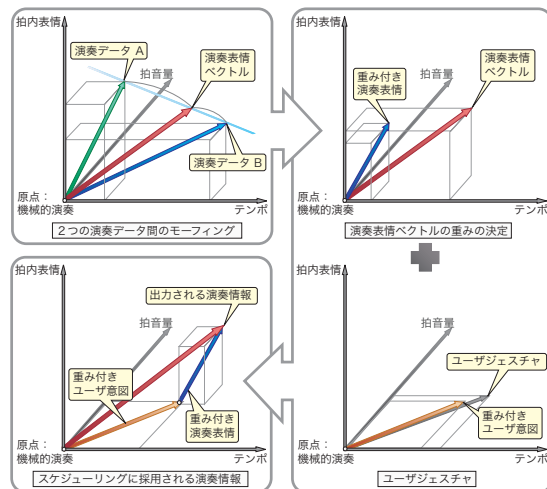


図 4 iFP における演奏表情の計算の概念図

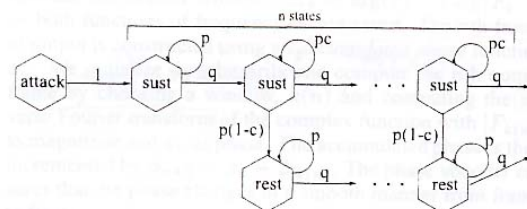


図 5 Orchestra in a Box の Listen 部のモデル

モデルを用いて発音時刻を特定する

- Synthesize: あらかじめ録音されたオーディオファイルを用いて、フェイズボコーダを用いて可変速度で再生する
- Anticipate: Listen と Synthesize を仲介し、各種スケジューリングを行う

これらのうち、Anticipation はシステムの意思決定を行う中枢部であり、予測制御のためにベイジアンネットワークを用いて構成されている．ネットワークの確率値は、あらかじめオフラインで学習されるのみならず、Listen で得られる音響信号をオンラインで解析することでリアルタイムに更新される．図5は、Listen における、ある1つのノートが持つ状態のマルコフモデルを表したものである．

2.7 自動伴奏

自動伴奏システムは、ソロパートと伴奏パートを知っていて、ソリストがどの部分を弾いているかの監視を続けながら、リアルタイムで伴奏パートの演奏スケジューリングを行う．ソリストがつけたテンポの変化、休止時間などにほとんど違和感を感じさせずに追従することを目標とする．スケジューリングには、人

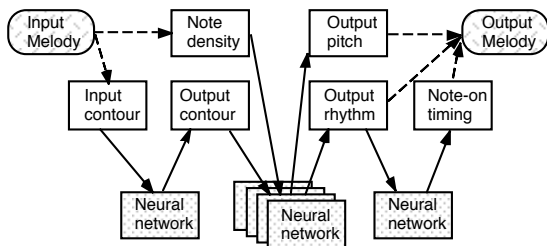


図 6 ニューロミュージシャンのシステム構成図

間の演奏するメロディに、ミスタッチやスキップが混入することを想定する必要がある。R. Dannenberg は、1980年代半ばに、動的計画法 (DTW) を用いたロバストなマッチングによってこの問題に対処した¹⁰⁾。さらに、トリルや和音など時間順序制約が外れる場合に対する (アドホックな) 解決法を示した。

2.8 Neuro-Musician

Neuro-Musician は、人とコンピュータの間での即興ソロ演奏の掛け合いを目的に、西嶋らによって 1991年に作られたシステムである¹¹⁾。このシステムでは、ソロ演奏と、それに対応する (後続の) ソロ演奏をニューラルネットワークによって学習しておく。ネットワークの初段に与える特徴量は、メロディの輪郭、音の密度である (図 6)。MIDI 楽器によって、8 小節程度の演奏が入力されるたびに、その演奏をモチーフとした掛け合い演奏が即興的に出力される。

2.9 Cypher

Cypher は、コンピュータとインタラクションを通じて音楽を創作することを目的として 1990 年代初頭に R. Rowe が開発した作曲環境である¹²⁾。Cypher は listener 部と player 部から成り立つ。listener 部ではリアルタイムで MIDI イベントを解析し、音の密度、音域、強弱など低次の音楽要素からコード、フレーズなどの高次の音楽要素までの認識を行う。player 部では、listener 部で認識された音楽要素に対する反応の仕方を記述したルールを発火させることで演奏を生成する。この両者の機能を組み合わせることにより、Cypher はインタラクティブコンポージングツールとして使用されるほか、究極的にはインプットなしで新しい音楽を作りだしていくこともできる。リスナー部においては、一つ一つをとると比較的シンプルな、音の密度、音域、局所的コードなどを抽出するエージェントが存在する。それぞれのエージェントを組み合わせることで、より高度 (複雑) な、音楽知覚・認知処理を実現することができる。

2.10 Continuator

Continuator は、2002 年に、F. Pachet によって、

開発されたセッション/インタラクティブ作曲環境である¹³⁾。その基本的なデザインは、プレイヤーの与えたフレーズの区切り (例えば 250ms に設定) をトリガーとし、直前に受け付けた数音を種として、それに続く確率的に最もふさわしい後続音をコーパスから検索することによって、後続フレーズを生成するというものである。演奏された音列を逐次解析し、音列の推移確率を蓄えておくことによりコーパスを生成していく。この際、Conklin らのシステムと同様、複数の視点に基づいて音列の特徴をモデル化している。Continuator では、コーパス上の音の推移確率と、直前のプレイヤー演奏の特徴 (傾向) に対して、どちらにどの程度、バイアスをつけるかを定めるパラメータを用意し、いわゆる「ライブ性にかかわるテイスト」を設定できるようにしている。

基本的な利用形態は、プレイヤーとシステムが交互にソロ演奏を行うというものであるが、ソロに続く伴奏をパートをシーケンスとして登録 (学習) しておき、オーバーラップしながら演奏生成を行うことにより、伴奏システムとして利用することもできる。

Continuator は、統合的な音楽生成処理のアーキテクチャを考えていく上で、非常に示唆に富んだシステムである。以下の章でも、再度取り上げ、考察を進めていく。

3. 考 察

3.1 予測補完という視点に関する考察

それぞれのシステムの特徴を、作曲、演奏表情における生成対象、記憶の観点から整理したものを表 1 に示す。この表において、記憶変数は、インタラクション型のシステムにおけるユーザの操作対象の意味で用いている。作曲、演奏表情の欄にあげている後続補完とは、入力に対しその後のシーケンスを出力するもの、縦型補完とは、当該拍のデータ (和声、演奏表情等) を生成するものを表している。

機能面に着目すると、すべてのシステムに、予め与えられた、あるいは、何らかの手段で内発的に生成された“種”に対して、長期記憶を参照しながら、データ (種) を詳細化するという共通の性質がある。詳細化過程は“補完”と読み替えることが可能である。補完過程自体は、未来の処理であっても現在の処理であっても、同種の処理である。Continuator や iFP などのように、インタラクティブ型であって、かつ、長期記憶による自律的なデータ生成とリアルタイムに入力される信号への反応のバランスの重み制御を行う機構を持つシステムは、作・編曲システムや音楽解釈シス

表 1 生成的音楽システムの分類

	作曲	演奏表情	長期記憶 (静記憶)	記憶変数	特徴
EMI ⁴⁾	ホモフォニー	-	モチーフ・推移率	-	乱数利用 (非決定論的)
Neuro-Musician ¹¹⁾	後続補完	-	音列	音列	ソロの掛け合い
Conklin コラール予測 ⁶⁾	単旋律後続補完	-	多次元視点推移率	-	予測型, エントロピー
松尾メロディ予測 ⁷⁾	単旋律後続補完	-	旋律推移率	音列	実時間予測, エントロピー
パービーブ ⁵⁾	縦型補完	-	単一曲和声	-	演繹的推論
Cypher ¹²⁾	後続補完	-	ユーザプログラム	音列	インタラクティブ作曲環境
Adliband ¹⁴⁾	後続補完	-	和音, スケール制約	音列	セッション作曲, 音響解析
Continuator ¹³⁾	後続補完	フレーズ転写	音列マルコフ連鎖	音列 (+リズム)	インタラクティブ, 重み制御
MIS ¹⁵⁾	-	縦型補完	数曲までの演奏表情	-	演奏ルールの抽出と適用
Dannenberg 自動伴奏 ¹⁰⁾	-	縦型補完	単一曲	テンポ	ミスタッチに対応
Orchestra in a Box ⁹⁾	-	縦型補完	単一曲表情	テンポ・音量	インタラクティブ
iFP ¹⁶⁾	-	縦型・後続補完	演奏表情	テンポ・音量	インタラクティブ, 重み制御

テムなどの静的なタスクを実施するシステムを包含する, と考えて良いだろう。

続いて, 人間の処理の視点から, 縦型補完と後続補完の意味合いについて考察を行う。ある音に合わせて音を出すのと, 未来の音列を組み立てるのでは, 感覚的に大きな差があるように感じられるかもしれない。しかしながら, 最近の脳科学の考え方に従えば, 行為としての処理は双方とも, “予測処理” が本質であり, 合わせて弾いたという感覚とそれにかかわる “現在・同時性” の知覚は「脳の合理化」過程が生み出した解釈 (イリュージョン) ということになる。あるタイミングに合わせて音を鳴らす際, 意識にあがることはほとんどないが, その前から, 動作のスケジューリングがなされている。加えて重要なのが, 認識にかかわる時間遅れの問題である。B. Libet の, 前腕と脳の直接刺激による順序知覚の実験や, 最近の TMS (経頭蓋的磁気刺激法) を用いた実験によって, 視聴覚・触覚刺激が知覚されるまでには無視できない遅れがあり, 脳はその遅れを感じさせないようにさせる補償処理を行っていることが明らかになっている¹⁷⁾¹⁸⁾。つまり, 縦型補完と後続補完にしても, プロセスの実施内容という視点では, 予測処理が中心的な役割を果たしており, 音楽とその行為に対しての質感は事後にもたらされた感覚ということになる。

では, ゆっくりと時間をかけて実施する作曲や音楽解釈の場合はどうであろうか? この場合は, 上での議論のようなシビアな実時間処理の制約はない。とはいっても, その作業において, 一瞬で音楽の全体像を俯瞰したり, リダクション構造や任意の場所で切り取った音楽的断片を思い浮かべたりすることは極めて困難である。作曲, 音楽解釈においても, 我々人間

は, 時系列を参照した思考に頼らざるを得ない。そこでの作業, 少なくとも, 評価にかかわるプロセスは, 音楽聴取と同様のものであると思われる。Meyer もかつて指摘したように, 「予測しつつ聴く」は, 音楽 (聴取) における根幹的な態度の一つである¹⁹⁾。次の進行を予測 (補完) するという処理は, 音楽にかかわる処理の中でも最も根幹的なものと位置づけて間違いないだろう。

3.2 機能とデータ表現に関する考察

予測補完型のアーキテクチャを採用するとして, 次に問題となるのはデータ表現である。その際,

- (1) 長期記憶のモデルと利用
- (2) 記憶変数の反映と操作性
- (3) 時間記述とリズムの表現
- (4) 創造性・決定性の扱い

について, 考慮していかなければならない。

これらの事項について, Continuator と iFP が採用している手法を紹介し, 次章でのアーキテクチャの提案につなげていく。

3.2.1 Continuator

長期記憶のモデルと利用

ツリー構造のマルコフ連鎖によってコーパスを記述している。各ノードにおける基本データは, 演奏に関するデータとして, pitch と duration (親ノードからの) delay タイム, 子ノードへのポインタがある。一旦, 生成 (continuation) のプロセスが始まると, コーパス内の最長リストに一致するノードを探索し, そこからの後続音を選択し, 後は, 逐次的に後続音を生成していく。

記憶変数の反映と操作性

覚構造に訴える楽譜は, 構造を一瞬で俯瞰する可能性を与えている。

文献には詳細な記述は無いが, 分解能は音符単位と思われる

サバンの中にはそのような能力を持った人がいるそうである。視

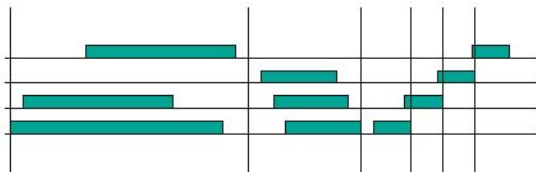


図 7 continuator のスパン．縦軸が pitch，横軸が時間．それぞれのスパンがノードに相当する．

プレイヤーはフレーズを演奏することによって、マルコフ連鎖の初期値 (continuation の種) を与える．データ探索は、例えば、

- (1) $\text{pitch} * \text{duration} * \text{velocity}$
- (2) $\text{small pitch region} * \text{velocity}$
- (3) $\text{small pitch regions}$
- (4) large region

などのように、具体的なものから、より抽象レベルの高いものへと、幅を広げていく形で実施される．この操作は、sparseness 問題の解決策にもなっている．

時間記述とリズムの表現

コーパスからの探索においては、図 7 中の各区間が一つのノードとして扱われる．探索時に利用されることはないが、各ノード中の各音に、親ノードからの delay タイムが記述されており、生成フェーズで利用することも可能である．

生成フェーズのリズムについては、1) ノードを統合せずにそのまま再生、2) ノードの順序性を保ち、すべて同じ音価で再生、3) 入力フレーズのリズムのコピー、4) 量子化処理に基づいた拍節構造の利用、の 4 つのタイプをユーザが選べるようになっている．

創造性・決定性の扱い

Continuator では、occurrence 頻度の重み付けによる乱数選択で、次音の選択を行っている．出力は決定的なものではない．また、マルコフ連鎖による探索に加え、直近のプレイヤー演奏の特徴 (傾向) に対して、バイアスをかけるパラメータを用意している．これにより、例えば、プレイヤーが直近で、弾いたスケール (音) を優先させて出現させるといった操作が可能となっている．

3.2.2 iFP

長期記憶のモデルと利用

iFP は、特定楽曲の演奏システムである．拍毎の制御情報と、拍以下の微細な deviation を分離して、T-Tree 構造でデータを表現している．特定楽曲の演奏システムであるため、音の並びの構造は、楽曲毎に固有である．演奏データは、正規化されたデータとそこからの deviation の組みとして記述されている．演

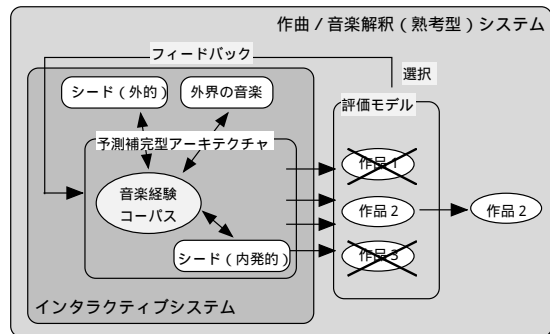


図 8 予測補完型音楽生成のコンセプト図

奏データベースから、任意の条件により選択された複数演奏の平均をテンプレートモデルとして利用することができる．また、二つのテンプレートのリアルタイムモーフィングが可能である．

記憶変数の反映と操作性

iFP では、拍打、あるいは、指揮ジェスチャにおけるテンポ、ダイナミクスの変化を通じて、演奏の制御を行う．現在のプレイヤー意図、数個の履歴平均、直近と現在の比較からの予想の 3 つの要素の重みを与えることにより、応答性にかかわるテイストを設定する．プレイヤーが明示的に演奏をストップ させない限り、iFP は演奏を続ける．また、プレイヤーは任意の場所で介入が可能である．

時間記述とリズムの表現

いわゆる楽譜表記に相当するリズム情報を有し、テンポはタクトス毎に与えられている．テンプレートモデルは、少数のガイド音を用い、DP と HMM のハイブリッド処理によって生成される²⁰⁾．

4. 予測補完型音楽生成アーキテクチャ

この章では、自動作曲・編曲、パフォーマンスレンダリングを統合的に実現するための基本アーキテクチャの設計について論じる．

4.1 基本的な考え方

前章までの議論により、アーキテクチャの中核には、予測補完処理を据える．また、長期記憶による自律的なデータ生成とリアルタイムに入力される信号への反応のバランスの重み制御を行うものとする．より人間の認知・知覚に留意したモデルを形成する場合には、これに加えて評価機構が必要となる (図 8) ．

以上のことを、M. Minsky が Emotion Machine

一定時間、打鍵状態をキープする、あるいは、指揮動作で手を下方にキープする

<http://web.media.mit.edu/~minsky/>

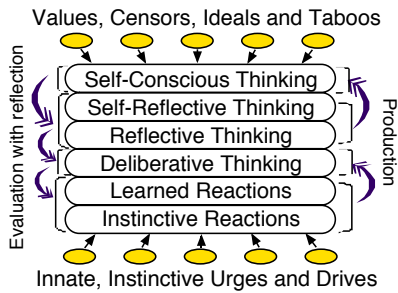


図 9 Emotion Machine の 6 つの階層にフィードバックループを付加

の中で提案している 6 層モデル (図 9) に対応付けて考えてみる。予測補完型アーキテクチャの内部状態が、下位 2 層 (主として、Learned Reaction 層) に対応する。予測補完型のアーキテクチャが可能世界を予測するというプロセスが、Deliberative Thinking 層に相当する。評価モデルのうち、自身の美的感覚に基づくフィードバックが、Self-Reflective Thinking の層、外部での評判想定等を含めた評価が Self-Conscious Thinking の層に相当する。

なお、操作性の良いインタラクティブな音楽ツールは、Self-Reflective Thinking での思考を外在化させる、すなわち、Deliberative Thinking 層での思考に対するウェイトを高めることで、ユーザの創造性を高めていると考えることができる。

4.2 システムデザイン

どのような音楽タスクにしても知識の蓄積と利用を考える必要がある。Conklin の研究は、これまで音楽知識と呼ばれてきた情報が、viewpoint と与えたデータ事例から学習される (形成される) スタイルに分離できることを示唆している。Emotion Machine の考え方に基けば、viewpoint 自体が学習の対象となるが、現時点でそこまでの自己組織化を考慮することが合理的であるとは思われない。音楽ドメインに共通する viewpoint に関しては、そのものか、その形成を促す仕組みを、我々 (システム開発者) が用意してもよいと考える。類似性の判定に関連するリダクションの種類と構成についても、現時点では、我々 (システム開発者) が与え、しかるべき道筋がついた後、学習の対象とするべきだと考える。以上のような考え方に基いたシステムデザインを以下に示す。

まず、図 11 のようにオブジェクト同士のネットワークからなるデータベースを構築する。ただし、図中の横軸 $\Delta flame$ は、あるフレームと他のフレームとの間の距離を表しており、アークの始点は全て $\Delta flame = 0$ のオブジェクトである。

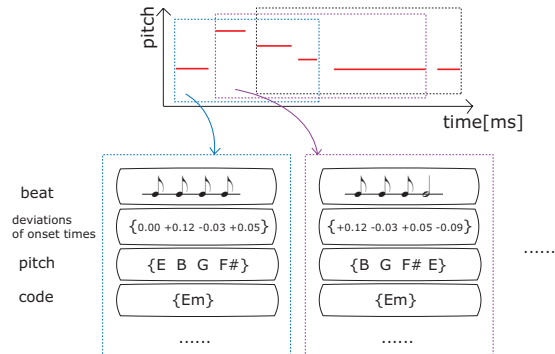


図 10 フレーム分割によるリダクション

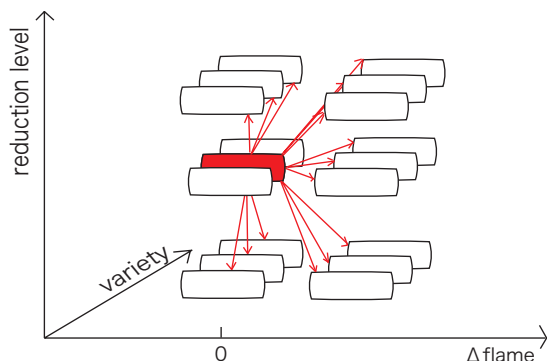


図 11 オブジェクトネットワークデータベース

このネットワークデータベースでは、リダクションレベルの異なるオブジェクト同士の接続 (図 11 の縦軸方向) を許している。量子化された発音時刻など、入力される学習データそのものからは直接得られないオブジェクトもあるが、抽出対象とするオブジェクトを観測されたデータの隠れ状態ととらえることで、同一のデータベースを用いて、ベイズ推定による補完が可能である。

生成処理の流れを図 12 に示す。先行音のフレームをオブジェクトに分割し、それに後続するフレームのオブジェクトを、データベースを参照することで予測し、生成する。ここで、予測の種となるオブジェクトおよび出力対象とするオブジェクトは、ユーザが指定できるものとする。この組み合わせにより、システムの聴取にかかわる性格と役割が決まる。

以上のアーキテクチャにより、例えば「メロディの履歴を参照して和声付けを行う」「リズムにふさわしい deviation を付加する」といった処理も同一の枠組みで包括・統合的に扱うことが可能になるとともに、一連の処理をリアルタイムに行うことで、伴奏システム、セッションシステムとしての動作も可能である。ただし、実装の際にはレイテンシを削減するために何らか

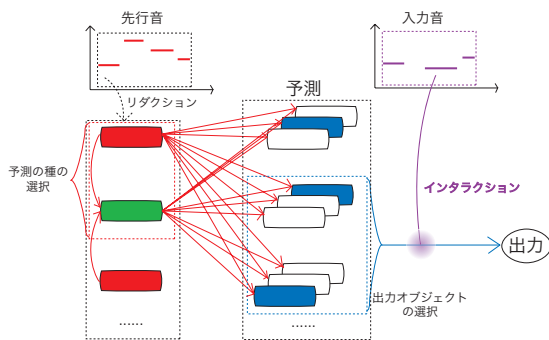


図 12 生成処理の流れ

の簡約, 近似が必要と思われる. また, いかに入力音とのインタラクションをはかるかについても, 様々な方法が考えられる. これらについては, 今後検討していきたい.

5. ま と め

本稿では、「作曲システム」「編曲システム」「表情付けシステム」「伴奏システム」「セッションシステム」の各領域において実施されてきた代表的な研究例を紹介した. それら研究例のタスクの特徴, データ構造, 処理の特徴に関して考察を行った結果, 音楽生成系タスクの統合化に向けては, “聴取経験に基づく予測補完” という視点が重要であるという結論に達し, その視点に基づいたシステムデザインの一例を示した.

現状では, hard-coded されたプロトタイプシステムしか出ていない. また, ここで示したシステムデザインには, まだまだ改良の余地があるが, 本論文で論じた課題は, 音楽情報科学領域の中で, 大きな発展性を秘めた研究対象であると考えている. 今後, さらに, 研究を進めていきたい.

謝 辞

本研究は, 独立行政法人科学技術振興機構戦略的創造研究推進事業さきがけタイプ「協調と制御」領域の研究テーマとして実施されました.

参 考 文 献

- 1) Hiller, L. and Isaacson, L.: *Experimental Music*, McGraw-Hill (1959).
- 2) Roads, C.: コンピュータ音楽 (青柳, 小坂, 平田, 堀内 訳・監修), 東京電機大学出版局, pp. 678-752 (2001).
- 3) 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).
- 4) Cope, D.: *Computers and Music Style*, A-R

EDITIONS (1991).

- 5) 平田圭二, 青柳龍也: パーピーブ: ジャズ和音を生成する創作支援ツール, 情報処理学会論文誌, Vol. 42, No. 3, pp. 633-641 (2001).
- 6) Conklin, D. and Witten, I.: Multiple Viewpoint Systems for Music Prediction, *Journal of New Music Research*, Vol. 1, pp. 51-73 (1995).
- 7) 松尾聡子, 片寄晴弘, 井口征士: 旋律予測のコンピュータシミュレーションモデルに関する一検討, 情報処理学会論文誌, Vol. 41, No. 2, pp. 498-508 (2000).
- 8) Katayose, H. and Okudaira, K.: Using an Expressive Performance Template in a Music Conducting Interface, *Proc. of NIME04*, pp. 124-129 (2004).
- 9) Raphael, C.: Orchestra in a Box: A System for Real-Time Musical Accompaniment, *Proc. of The IJCAI-03 Workshop on methods for automatic music performance and their applications in a public rendering contest - Rencon-*, pp. 5-10 (2003).
- 10) Dannenberg, R.: *Real-Time Scheduling and Computer Accompaniment*, Current Directions in Computer Music Research, MIT Press, pp. 161-183 (1989).
- 11) Nishijima, M. and Watanabe, K.: Interactive music composer based on neural networks, *Proc. of ICMC*, pp. 53-56 (1992).
- 12) Rowe, R.: *Interactive Music Systems: Machine Listening and Composing*, The MIT Press (1993).
- 13) Pachet, F.: The Continuator: Musical Interaction With Styler, *Journal of New Music Research*, Vol. 32, No. 3, pp. 333-341 (2003).
- 14) 青野裕司, 片寄晴弘, 井口征士: アコースティック楽器を用いたセッションシステムの開発, 電子情報学会論文誌, J82-DII, No. 11, pp. 1847-1856 (1999).
- 15) Katayose, H. and Inokuchi, S.: Learning Performance Rules in a Music Interpretation System, *Computers and the Humanities*, Vol. 27, pp. 31-40 (1993).
- 16) 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, Vol. 44, No. 11, pp. 2728-2736 (2003).
- 17) Pollack, R.: *The Missing Moment*, Houghton Mifflin (1999).
- 18) Carter, R.: ビジュアル版 脳と意識の地形図 脳と心の地形図 (藤井 留美訳), 原書房 (2003).
- 19) Meyer, L. B.: *Emotion and Meaning of music*, University of Chicago Press (1956).
- 20) 豊田健一, 野池賢二, 片寄晴弘: 演奏 deviation データベースの作成と公開, 情報処理学会研究報告 2004-MUS, No. 55, pp. 9-14 (2004).