

# 旋律予測のコンピューテシヨナルモデルに関する一検討

松尾 聡子<sup>†</sup>, 片寄 晴弘<sup>††</sup> 井口 征士<sup>†</sup>

人間は聴取などの音楽行動において、次に演奏されるべき旋律を予測していると考えられる。したがって、コンピュータを用いた音楽構造の解析や音楽の生成を行うような音楽システムの構築には、旋律を予測する機構の実装が必要とされている。しかし、従来開発されてきた音楽システムのうち、予測機構部を持つものはほとんど見受けられない。本研究では単旋律を対象に、リアルタイムで予測を行うコンピューテシヨナルモデルを実装したシステムの構築を行う。本システムにおいて実装されているモデルは、人間の音楽認知特性を考慮しており、音列対を仮想的に3次元空間上に配置して、それらの距離を求めるという自己回帰手法に基づいて動作する。本システムは人間の記憶や音程知覚に関する特性をパラメータとして与えており、これらのパラメータを変化させることで、用いたモデルの有効性をリアルタイムに検証することが可能である。したがって、本研究は人間の音楽認知・理解に関して考察・検討を行っていく際のビジュアライゼーションシステムとして有用である。さらに本システムはインタラクティブシステムへの応用のほか、それまで演奏されたいくつかの音から予測を行い、予測候補を演奏音として扱うことで、さらに予測を繰り返すことができるため、自動生成すなわち自動作曲システムとしての応用が可能である。

## A Computational Model for Melody Prediction

SATOKO MATSUO,<sup>†</sup> HARUHIRO KATAYOSE<sup>††</sup> and SEIJI INOKUCHI<sup>†</sup>

In musical activities, for example, listening to music, humans predict melodies to be going to be played. Therefore, in order to construct musical systems using computers (analyzer, composer systems, and so on), it is essential to predict melodies. Nevertheless, there are few which have the mechanism of prediction. We construct a system that implements a computational model to predict monophonic melodies in real-time. This system takes multiple viewpoints and hierarchy of music into consideration, and estimates distances between pairs of melodic sequences, based on the self-correlation method arranging them in 3-D space. Moreover, its user can change parameters related to memory and pitch perception. So, our system is able to inspect the introduced models to be effective in real-time. Consequently, it is a useful visualization system in examining human mechanisms of musical recognition. Besides, the function of melody prediction can be applied for algorithmic composition. Repetitive note prediction produce melody sequences.

### 1. はじめに

音楽の鑑賞や演奏において、人間はいくらか先の音を予測しており、演奏が予測どおりだと安心感を抱いたり、予測と異なると意外感を感じたりする。このようにつねに何らかの予測に基づいた聴取を行っていると考えられる。すなわち、音楽を理解することは予測することと表裏一体であるといえる。

リズム・メロディー・ハーモニーは音楽の3要素と呼ばれ、重要視されている。これらの予測を扱ったものとして、Conklin<sup>1)</sup>の研究があげられる。Conklinらは、予測を行う直前までのメロディー、ハーモニーや局所構造といったデータを利用し、情報エントロピーを用いて、次の進行を予測するという手法を示し、人間とシステムの実施能力の比較を行った。しかし、これは楽譜を分析して得られる情報を基にした予測であり、小節や拍の情報(何小節目の何拍目で、音価がいくつなど)が明らかにされた状況での予測であった。一方、Vercoe<sup>2)</sup>は対象をリズムのみに限定しているが、与えられたビート列から自己相関を求めることで次のビートを予測するリアルタイムシステムの開発を行った。それまでの音楽知覚や認知の研究は、ジャンルなどの条件を設定して比較実験を行い、その結果に基づ

<sup>†</sup> 大阪大学大学院基礎工学研究科システム人間系専攻  
Department of Systems and Human Science, Graduate  
School of Engineering Science, Osaka University

<sup>††</sup> 和歌山大学システム工学部  
Faculty of Systems Engineering, Wakayama University  
現在、三菱電機株式会社産業システム研究所  
Presently with Industrial Electronics & Systems Lab.,  
Mitsubishi Electric Corp.

いてモデルを構築するというものがほとんどであり、全曲の情報を利用して予測を行うことが可能であった。その中で、Vercoeの研究はシステムの振舞いを実時間で示すものとして非常に注目を集めた。本研究はこれに音の高さの情報を加えた旋律のリアルタイム予測を扱う。本論文では、調性音楽を対象に、音の近親性に関する知覚特性、記憶可能な音数など、個人に依存する特性をパラメータとして変化させることにより、システムの振舞いがどのように変化するかをリアルタイムに検証しうるモデルの提案を行う。

## 2. 旋律予測に関する考察

音楽は過去の楽句・楽節の反復など、過去のイベントに影響されるという性質を持つ信号である。よって、旋律は過去から将来をある程度予測することができる。本システムはこのような旋律の性質に基づき、過去に演奏された旋律から類似した旋律を抽出し、リズムやメロディーというような韻律構造を明らかにすることにより予測を行う。したがって、旋律に関する類似性を定義するために以下のような音楽の構造や人間の記憶、聴覚上の特性について検討する。

### 2.1 音楽の多視点性、階層性

音楽は図1のように、音楽の3要素であるリズム・メロディー・ハーモニーの側面を持ち、さらにそれぞれが階層を持つような、四面体の階層構造と見なすことができる。ここでは、このように複数の側面からの解釈が可能であるという特性を多視点性と呼び、それぞれの側面が階層になっていることを階層性と呼ぶ。ただし、図1は基本となる考えを概略的に示した図であり、実際にはヘミオラなどの音楽技法を考慮すると、それぞれの側面で異なった階層構造を持つことになる。この問題に関して、実装上での詳細について本節で述べる。

本研究では単旋律を対象としているため、図1の音楽構造のうちリズム・メロディーの2つの視点を扱う。本研究ではリズムあるいはメロディーの一方はよく一致しているが、他方は一致しないような音列ではなく、両者がなるべく一致するような音列を手がかりとした予測を目指しており、これを多視点性を考慮した予測と呼んでいる。

また、音楽は低次からより高次の層に、拍-小節-楽句-楽節-曲全体のように階層構造となっている。このような構造の中で、音楽は小さなまとまりからより大

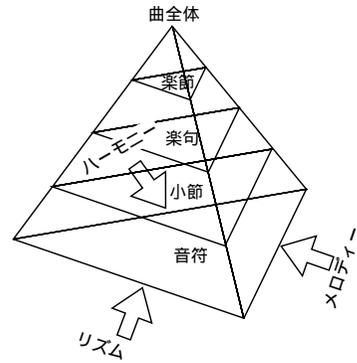


図1 音楽の多視点性と階層性

Fig. 1 Multiple viewpoints and hierarchy of music.

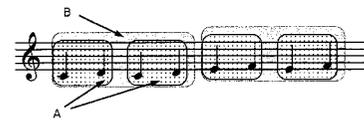


図2 階層性によるグループの包含

Fig. 2 Connotation of group caused by hierarchy of music.

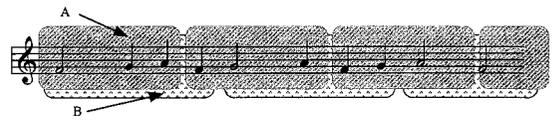


図3 多視点性によるグルーピングの不整合

Fig. 3 Conflict of grouping caused by multiple viewpoints of music.

きなまとまりへと階層的に認識されやすい。このまとまりを本論文では「パターン」と呼んでいる。しかし、拍や小節などの単位でパターン全体の音価や時間的長さ、すなわちパターンの大きさを固定した場合、パターンが小さすぎると高次の層で認識されず、逆に大きすぎると類似性が発見され難い。つまり、いずれにしても予測が困難となる。したがって、パターン全体の音価や時間的長さを固定せずにセグメンテーションを行う機構が必要となる。これには、小さなまとまりで認識し、それを積み重ねて大きなまとまりとする方法や、さまざまな大きさのパターンを同じ尺度で比較し、最も類似度の高いものを選択するという方法が考えられるが、本システムでは後者の手法を採用している。

また、図2は2音ずつのパターンの繰返しであるとともに、それより大きな4音ずつのパターンの繰返しにもなっている。このように大きなパターンがいくつかの小さなパターンを包含することがある。また複数の視点を考慮した場合、視点間でグルーピングの構

hemiola: たとえば、3拍子の曲で、2拍周期のメロディーを用いるような変形拍子。

造が異なることが考えられる．たとえば図3のような旋律の系列において音高によるグルーピングではAのような3音ずつのグループとなる．しかし、リズムによるグルーピングではBのような4音ずつになり、視点間でグルーピングの不整合が起こる．図2や図3のような場合、人間は複数の予測を可能性として持つことがある．したがって、人間の認知モデルとしての旋律予測システムは、単一の結果を正解として断定するのではなく、人間が行う可能性のある多重の解釈を認め、複数の予測候補を示すべきである．

## 2.2 人間の記憶特性

人間の記憶は大別すると長期記憶と短期記憶で表されるといわれている<sup>3)</sup>．音楽の予測においても長期記憶と短期記憶が作用しており、これらの記憶を結合して予測が行われると考えられる．音楽における長期記憶はジャンルや作曲家を同じくする旋律進行の集大成から推論される構造が記憶されており、短期記憶には進行中の曲そのものが記憶される．したがって、短期記憶に基づく予測はそれまでの曲の進行に則したものになる．またこれらの記憶には限界があり、その容量を超えたり一定の時間が経過すると忘却が起こる．人間の認知特性を考慮した旋律の予測には時間の経過や新たな旋律の入力などによる忘却機構を導入する必要がある．本研究ではジャンルなどの音楽的背景に依らない予測、すなわち短期記憶に焦点を当てた旋律予測モデルを扱う．

## 2.3 ピッチクラスに対する人間の知覚特性

人間はピッチクラスにおける音の距離に関して、単純な周波数の高低だけでなく、協和性を加味して知覚していることが知られている．たとえば、オクターブの音程を持つ2つの音はそれより半音狭い音程である長7度よりも近接して感じられることがある．このような音程知覚に関するモデルとして Shepard のモデルがあげられるが、このモデルの詳細は付録に示す．

## 3. 旋律予測の手順

図4に本システムにおける予測の手順を示す．1つの音の演奏が終了するごとに図4の一連の処理が行われ、それに続く何音かの予測結果が出力される．

### 3.1 セグメンテーション

自己回帰を行う前の段階として、演奏された音を何音かずつのまとまりにセグメンテーションする作業が

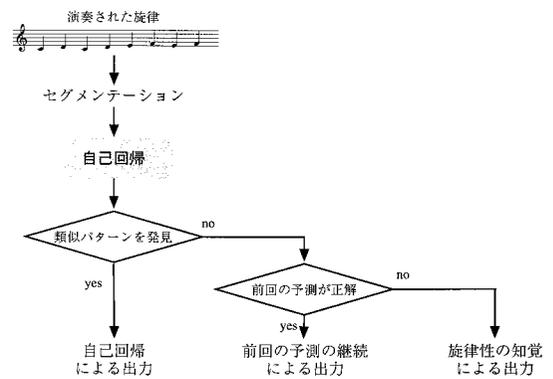


図4 旋律予測手法の処理の流れ

Fig. 4 Process of the prediction method for melodies.

必要となる．グループ中の音数に関しては、グループ中の最少音数と最多音数（ユーザがパラメータとして与える）の間のすべての音数によるセグメンテーションを行う．この値を小さくするとより局所的な予測が、大きくすると大域的な予測が行われる．また、より音数の多い類似パターンが優先的に検出されるようにしているため、より高次の階層で知覚される．したがって、階層性を考慮した予測が行われる．

### 3.2 類似性の検出

本研究で提案する自己回帰手法は、直前に演奏された音系列とそれより以前に演奏された音系列を対にし、以下のような処理を行うことにより、それらの類似性を調べるものである．

- (1) 音系列パターンの時間的長さの正規化
- (2) 人間の認知モデルの適用
  - (a) 短期記憶モデルの適用
  - (b) 音程知覚モデルの適用
- (3) 類似度の導出

これらの処理の内容については、4章で詳しく説明する．

### 3.3 予測候補の提示

予測候補は自己回帰の結果により、(1) ある程度類似したパターンが発見された場合（自己回帰による出力）、(2) 類似パターンは発見されなかったが、前の予測が正解で、その予測が継続されると考えられる場合（前回の予測の継続による出力）、(3) 類似パターンがなく、前の予測も不正解である場合（旋律性の知覚による出力）で、予測の手法が異なる．予測候補は複数あげることができ、その数はユーザがパラメータとして設定できる．したがって、2章で述べたようなグルーピングの不整合が生じた場合において、多重の解釈を認めることが可能である．ただし、予測手法の優先順位は(1)、(2)、(3)の順で、自己回帰処理で類似パ

ここでは、同一楽器によるメロディーの構成音、すなわち抽象化されたノートを指すものとする．

本論文中では、比較する2音の音の高さの差を「音程」と呼んでいる．

表 1 旋律性の知覚モデル

Table 1 Perception model of melodiousness.

上昇音程	短2度	2度	短3度	3度	4度	増4度	5度	短6度	6度	短7度	7度	8度
発生確率	0.046	0.066	0.041	0.066	0.065	0.014	0.05	0.028	0.036	0.03	0.03	0.041
下降音程	短2度	2度	短3度	3度	4度	増4度	5度	短6度	6度	短7度	7度	8度
発生確率	0.06	0.065	0.046	0.07	0.046	0.031	0.046	0.027	0.019	0.015	0.019	0.041

ターンが発見された場合は、過去の予測結果の正解・不正解にかかわらず、予測結果が更新される。類似パターンが発見されなかった場合に (2) の処理が、さらに前回の予測が不正解であった場合に (3) の処理が起動される。

3.3.1 自己回帰による出力

予測候補は、4章で述べる自己回帰手法によりパターン間の距離を求め、ある程度距離の近いパターンが発見されると、類似パターンと見なし、それに続く音系列が再現されると考える。この場合、自己回帰で求められた距離の小さいものから順に、反復パターンの第1, 第2, ... 候補と見なし、これに対応する音系列が第1, 第2, ... 予測候補としてあげられる。

3.3.2 前回の予測の継続による出力

反復パターン中の音数が多い場合や類似パターンの出現までに相当の時間を要した場合などは、反復パターンの一部あるいは全部が忘却されてしまい、前回の予測が正解であっても、自己回帰による類似パターンが発見され難い場合がある。この場合、予測は継続されるものとして、前回の予測結果の先頭の音を取り除き、2音め以降の音系列を新たな予測結果とする。したがって、この処理が連続して行われると、予測される音の数は1音ずつ減少する。またこの場合の予測は第1候補のみをあげる。

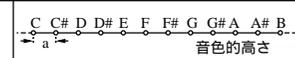
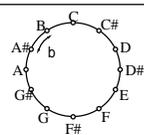
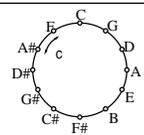
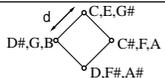
3.3.3 旋律性の知覚による出力

パターンの繰返しが存在しない場合、たとえば曲の初めや曲想の変化する部分でも、人間は妥当な旋律となるような予測を行っていると考えられる。そこで閾値(経験的に得られた数値)を下回る距離値を持つ音系列パターンが存在しない場合、すなわち類似度の高いパターンが発見されない場合は、旋律性の知覚モデル<sup>4)</sup>を適用し、システムは旋律らしくなるような予測結果を出力する。

旋律性の知覚モデルによる予測は、乱数を発生させ、その値によって表1のような確率でそれぞれの音程が採択されるようになっており、直前の音高に対し、採択された音程を持つ音高を予測結果とする処理を繰り返し、複数音先までの予測とする。この場合、音の長さは直前の音と同じ長さとしている。

表 2 音声知覚モデルの構成成分

Table 2 Constituents of pitch perception model.

モデル名	強調される音程	モデル
A	—————	
B	短2度 長7度	
C	完全4度 完全5度	
D	長3度 短6度	

3.4 ユーザにより与えられるパラメータ

(1) 記憶可能な音数

ユーザは記憶可能な音数(単音記憶数)を与えることができる。これにより時間の経過による忘却のほかに、新たな入力という妨害による忘却機構を持たせることができる。

(2) セグメンテーションに関するパラメータ

ユーザはパターン中の最少音数と最多音数、すなわちパターンの最小周期と最大周期を与えることができる。この範囲でパターン中の音数を可変長にし、より多くの音数で知覚されやすくすることにより、階層性を考慮した予測が可能になる。たとえば、単音記憶数12、最小周期4、最大周期8とした場合、まず記憶している12音を連続する8音ずつの音で切り出し、その直前のパターン(ターゲットパターン)とそれ以外のすべてのパターン(ソースパターン)との類似度を計算する(ソースパターンとターゲットパターンの重複は認める)。次に7~4まで、パターン中に含まれる音数を変化させて同じ作業を繰り返すことで、パターンの大きさを可変長にした予測が可能になる。

(3) ピッチクラスに対する知覚特性に関するパラメータ

ユーザは表2に示すピッチクラスの知覚的距離を与えることができる。予測システムはこの値が小さい音

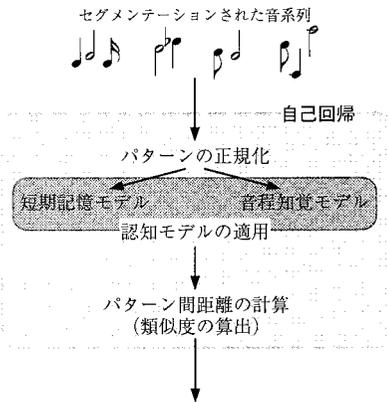


図5 旋律予測モデル

Fig. 5 The model for prediction of melodies.

程を類似していると認知する．逆に大きな値を与えると，類似したパターンが発見され難くなる．

#### (4) 予測候補に関するパラメータ

ユーザは予測候補数(予測候補を何位まで提示するか)を与えることができる．これにより多重の解釈が得られる可能性を持たせることができ，人間の音楽に対する多様な解釈の可能性を示すことが可能になる．また，ユーザは予測する音数(何音先まで予測するか)を与えることができる．

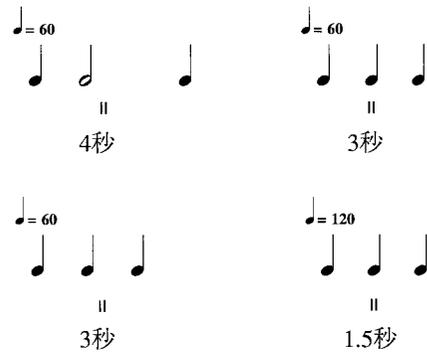
### 4. 旋律予測のコンピューショナルモデル

本研究で提案する旋律予測のコンピューショナルモデルは，セグメンテーションされた音系列パターンの組に対し，図5のような処理を行うものである．個々の処理の内容について，以下の節で詳しく説明する．

#### 4.1 パターンの時間的長さの正規化

人間は曲の演奏中に多少のテンポの揺らぎが起こった場合でも，ほぼ間違いなく同じフレーズであると認識することができる．本研究は人間による演奏を前提としており，テンポの揺らぎが存在すると考えている．そこでセグメンテーションされた音系列の時間的長さに関して，正規化を行っている．すなわち本システムでは，一定倍の音価で表されるフレーズは元のフレーズと同一であるとしている．

演奏された旋律に対しセグメンテーションを行った際に，そのパターン全体の時間的長さが異なる場合が考えられる．たとえば図6(a)のようにテンポを  $\text{♩} = 60$  で一定にしてもパターン内にある音符の音価の違いにより上のパターンの時間的長さは4秒，下図は3秒となる．また図6(b)に示すように，同じ楽譜でも  $\text{♩} = 60$  で演奏した場合と， $\text{♩} = 120$  で演奏した場合ではそれぞれ3秒，1.5秒のパターンとなり，演



(a) 音価の違い

(b) テンポの違い

(a) Difference of time value

(b) Difference of tempo

図6 正規化を必要とする場合の例

Fig. 6 Examples of requiring normalization.

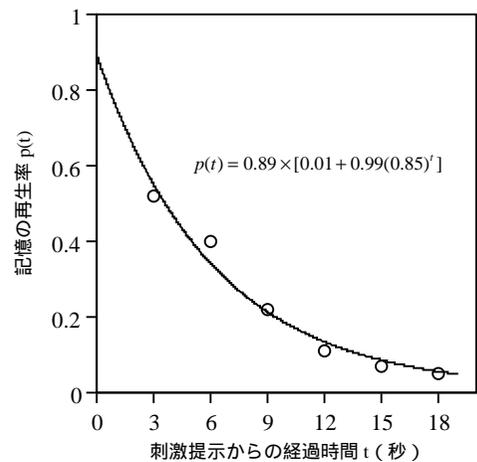


図7 短期記憶モデル

Fig. 7 Model of short-term memory.

奏の速度やテンポの緩急，さらに演奏の揺らぎによってもパターン全体の時間的長さが異なる．このような時間的長さの異なるパターン間での類似度の導出，比較を行うため，本システムではすべてのパターンを拡大，あるいは圧縮して，パターン全体の時間的長さを経験的に100に正規化し，パターン中の各音の長さが100のうち，それぞれいくらを占めるかを計算している．

#### 4.2 人間の認知モデルの適用

##### 4.2.1 短期記憶モデルの適用

ジャンルや作曲家などの背景に依存しない旋律の予測には，短期記憶が支配的であると考えられる．本システムでは時間経過による忘却を表現するモデルとして図7<sup>5)</sup>を導入し，このモデルを定式化した式(1)を用いている．

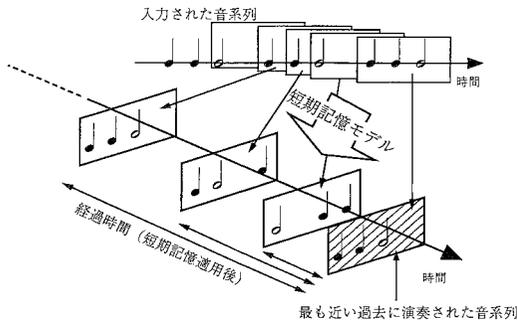


図 8 短期記憶モデルの適用

Fig. 8 Application of model of short-term memory.

$$p(t) = 0.89 \times [0.01 + 0.99(0.85)^t] \quad (1)$$

短期記憶モデルは、式 (1) の  $t$  に、現在から最も近い過去のパターンと、それ以前に演奏された同じ音数の各パターンとを、それぞれ対にして、それらの最後の音の終了時刻（正規化時刻ではなく、実際の時刻）の差を代入し、 $p(t)$  の逆数をその 2 パターン間の時間経過軸上の距離とすることにより適用される。この距離値の小さいものを類似度が高いと見なす。このような処理を図で表現すると、図 8 のように表され、より以前に演奏された音系列パターンは時間経過軸上で、より遠くに配置される。これにより、現在のパターンとの距離値が大きくなり、類似度が低くなる。このようにして時間経過による忘却を表現することができる。また単音記憶数をパラメータとして与えることにより、新たな入力による過去の旋律の忘却を表現することができ、経過時間にかかわらず忘却が起こる。

4.2.2 音程知覚モデルの適用

付録に示したモデルにより、正規化時刻において対応して発音されている音のピッチクラスにおける知覚的距離を求める。ただし一定音高、上下させてできるパターン（移調など）は同じものと見なすため、パターン中の最初の音の高さが等しくなるように平行移動している。本システムでは、音色的高さ、音楽的高さ、完全 5 度、長 3 度の音程成分を表 2 のようなモデルで表し、それぞれの音程成分に対する知覚的距離を表 2 のモデルにおける幾何学的距離で示している。本システムの音程モデル上での距離は、半音（オクターブを考慮しない）、半音（オクターブ帰帰）、5 度、3 度の音程モデル上での距離をそれぞれ  $a, b, c, d$  とすると、 $\sqrt{a^2 + b^2 + c^2 + d^2}$  を計算した値を採用している。たとえば、C-C の距離は、 $\sqrt{a^2 + b^2 + (5c)^2 + d^2}$  と計算される。本システムは、この  $a, b, c, d$  の値をパラメータとして与えており、優位に働く音程成分を変化させることができる。

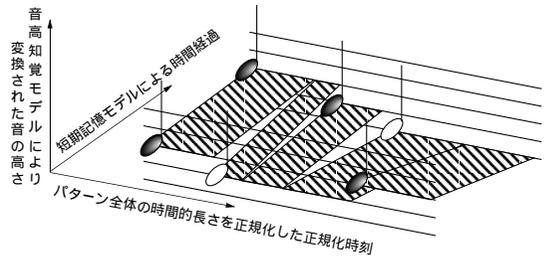


図 9 音形列の 3-D 空間表現

Fig. 9 3-D representation of sequence patterns.

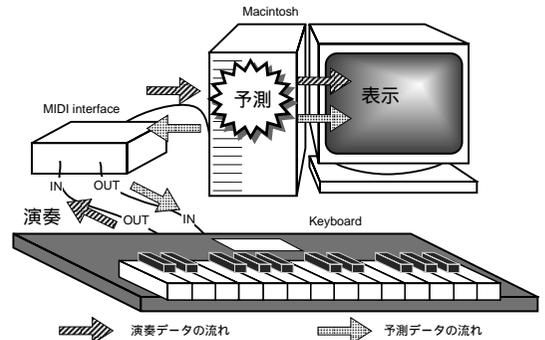


図 10 システムのハードウェア構成

Fig. 10 Hardwares of our system.

4.3 類似度の導出

予測の直前に演奏された音系列パターンとそれより以前に演奏された音系列パターンを対にして、短期記憶モデル、音程知覚モデルを適用し、2 パターン間でそれぞれの正規化時刻において発音されている音のピッチクラスにおける距離を求め、累積する。ただし音のない正規化時刻は前の音で内挿する。したがって、パターン間の距離の算出は図 9 の斜線部の長方形の面積の和を求めることに相当し、この面積が小さいものほど類似度が高いと解釈する。またこの際、類似度が等しいものが複数あればパターン中の音数の多いものを優先する。

5. システムの概要

5.1 ハードウェア構成

本システムのハードウェア構成を図 10 に示す。演奏者は MIDI (Musical Instrument Digital Interface) 音源付きキーボードで演奏を行い、演奏された旋律は MIDI インタフェースを介して Macintosh に入力される。Macintosh で予測された結果はモニタに表示されるとともに、予測の第 1 候補は再び MIDI インタフェースを通してキーボードで出力される。ただし、演奏音と予測音が重なって演奏されると両者が区別し

にくいため、予測音は1オクターブ低い音で演奏される。また、予測が正解でないことが明らかになるとその予測音の演奏を中止するようになっている。

## 5.2 結果の表示

キーボードからノート・オフの情報が入力されるごとに先行きいくつかの音の予測が行われ、結果はシーケンサのように表示される(図11参照)。図11の縦方向は音高を示しており、上部に行くほど高い音を示している。図11の横方向は時刻を示している。中央部の縦線は現在の時刻を表し、縦線より左がすでに演奏された音系列を、右が予測された音系列を表している。これらの音列は時間の経過とともに右から左へスクロールされる。また表示される結果は予測が行われるたびに、すなわちノート・オフされるごとに更新される。つまり、表示は時間の経過、演奏の進行とともに、時々刻々と変化していく。したがって、パラメータを変化させた場合に、その影響を動的に観測することができる。また図11では予測候補は第3位まであげられており、可能性の高い順に赤、黄、青で色分けされているため、多重の解釈がなされる場合もそれぞれの予測を視覚的に観測することができる。

## 5.3 確信度

図11の右上の百分率は、マッチングの第1候補パターンと直前のパターンとの距離値、すなわち図9の斜線部の面積の累積と閾値(経験的な値:500)から式(2)のようにして計算される値で、第1予測候補に対する確信度と見なすことができる。

$$\left\{1 - \frac{D}{T}\right\} \times 100 (\%) \quad (2)$$

( $D$ : 第1予測候補との距離値,  $T$ : 閾値 = 500)

ただし、この値はそれぞれのパターンの距離値  $D$  に対して計算される値であり、すべての予測候補に対する確信度の合計は100%にはならない。

## 6. 実験と結果

### 6.1 実験 I: 階層性に関する実験

図12(a)のように6音周期のパターンAがパターンBを包含するような階層性を持つ旋律を演奏し、予測を行った。図12(a)における予測結果の出力画面を図12(b)に示す。このとき、パラメータは単音記憶数30、最小周期2、最大周期24、音程知覚モデルの構成成分比(表2)  $A:B:C:D=2:4:2:1$  を与え、先3音の予測を第3候補まであげている。

このとき第1候補(赤)は、A音(4分)の演奏後、しばらく休符が続いた後、88.7%の確信度でG(2分)-D(4分)が演奏されると予測しており、6音周期の旋律が半音ずつ上昇して演奏されていると認識している。本システムには音楽的妥当性(調号など)を検証する機構がないため、このG音のようにその調では現れないはずの音が予測されることがある。調の考慮は今後の課題として導入すべき機構であると考えている。第2候補(黄)は直前の「F(2分)-C(4分)」の音型が繰り返されるとして、E(2分)の次に「A(4分)-C(2分)」を予測したが、その後はまだ演奏されていないため、予測の最初の音(A音)をもう一度、予測として出力している。第3候補(青)は「G(2分)-D(4分)-F(2分)-C(4分)」を2度下げて繰り返すものとし、E音が演奏された後「B(4分)-D(2分)-A(4分)」が演奏されると予測している。ただし第2、第3候補の予測1音目のA音、B音は出力画面を撮った時点ですでに演奏され終わっていると予測されているので、表示されていない。

### 6.2 実験 II: 多視点性に関する実験

図13(a)のように2音周期のリズムパターンと3音周期の音高パターンを持つ旋律を演奏し、予測を行った。図13(a)における予測結果の出力画面を図13(b)に示す。このとき、パラメータは単音記憶数12、最小周期2、最大周期8、音程知覚モデルの構成成分比(表2)  $A:B:C:D=2:4:2:1$  を与え、先3音の予測を第3候補まであげている。

このとき第1候補(赤)は「E(2分)-D(4分)-G(2分)」を97.9%の確信度で予測しており、2音周期のリズムと3音周期のメロディーを認識している。実際には、その後、E音が演奏されているが、本システムは次の音の正解と予測音との差が半音なら正解と見なしている。したがって、図では第1候補の1音目はE音として描かれている。これは調の考慮によりある程度解決される問題であると考えている。第2候補(黄)は「E(4分)-D(2分)-G(4分)」を予測しており、ずっと以前の、2つの分休符+3音周期のパターンを繰り返すと認識している。第3候補(青)は直前の4音を完全5度下げたパターン、すなわち「C(2分)-A(4分)-G(2分)」が演奏されると予測している。

正確には第1候補も正解ではないが、第2候補のようなずいぶん過去のパターンを手がかりにした予測や、第3候補のようなリズムだけを手がかりにした予測ではなく、リズムとメロディーの両方を手がかりに、なるべく新しいパターンの再現を行っているという点で、第1候補の予測は妥当な結果であると考えられる。

図12-図16の結果に関しては、音名を示した方が結果が理解しやすいと思われるため、実際には出力画面に表示されない音名を併記した。

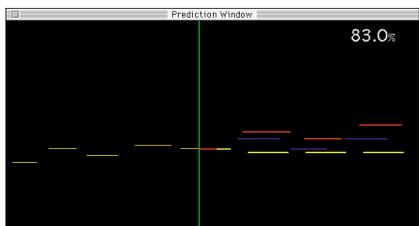


図 11 予測結果表示画面

Fig. 11 Display window for prediction.

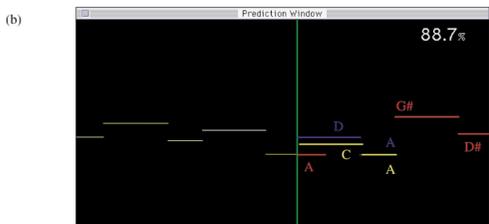
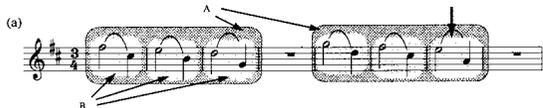


図 12 実験 I. (a) 演奏した旋律と (b) 結果

Fig. 12 (a) Melody and (b) result for Experiment I.

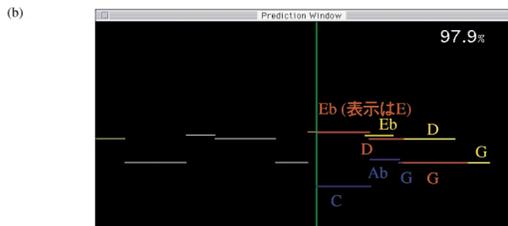
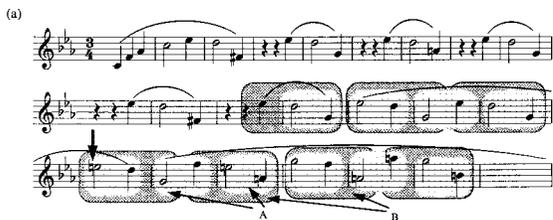


図 13 実験 II. (a) 演奏した旋律と (b) 結果

Fig. 13 (a) Melody and (b) result for Experiment II.



図 14 実験 III, IV で演奏した旋律

Fig. 14 Melody for Experiment III, IV.

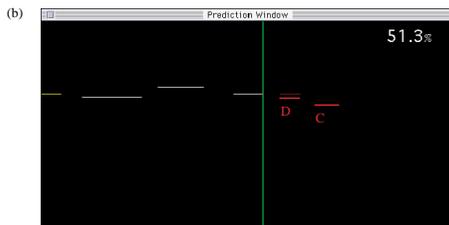
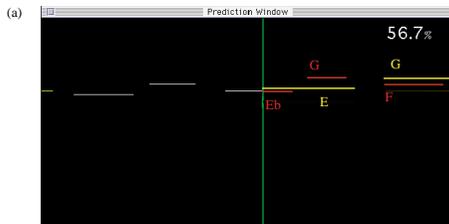


図 15 実験 III の結果

Fig. 15 Result of Experiment III.

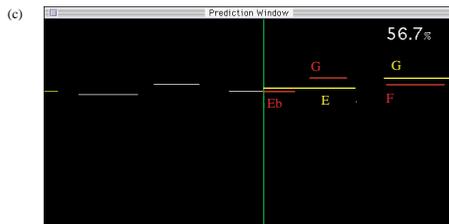
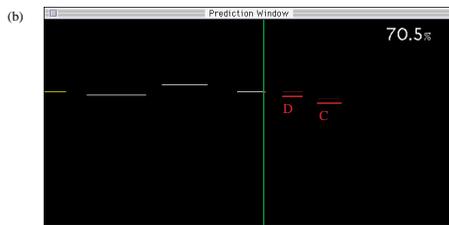
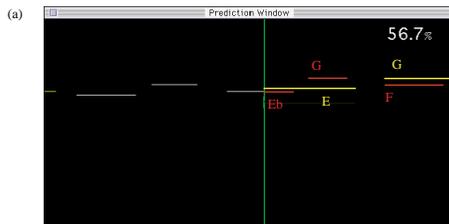


図 16 実験 IV の結果

Fig. 16 Result of Experiment IV.

### 6.3 実験 III：新たな入力による忘却に関する実験

新たな入力による忘却に関する実験として、図 14 の旋律を演奏し、予測を行った。図 14 の緑返し時における予測結果の出力画面を図 15 に示す。図 15 (a) は単音記憶数 24, (b) は単音記憶数 12 としており、下段の方が記憶容量の小さい場合を示している。それ以外のパラメータは、いずれも最小周期 2, 最大周期 12, 音程

表3 実験IVにおける音程知覚モデルの構成成分比  
Table 3 Proportion constituents pitch perception model for Experiment IV.

音程 \ グループ	(a)	(b)	(c)
音色的高さ(A)	2	4	1
音楽的高さ(B)	4	2	4
5度(C)	2	1	2
3度(D)	1	1	2

知覚モデルの構成成分比(表2)A:B:C:D=2:4:2:1を与え、先3音の予測を第2候補まであげている。

このとき(a)は繰返し記号による繰返しを認識し、第1候補として、56.7%の確信度で「E(2分)-G(4分)-F(2分)」を予測している。確信度が低いのは類似パターンが演奏されてからの経過時間が長いためである。また第2候補は「E(2分)-G(全音符)-F(あるいはG(2分))」が予測されている。これは前の小節の「E-D」の繰返し、すなわち2音周期のパターンを認識している。また、E音にはスタッカートが付いており、F音より短く演奏されていたため、以降は約2倍の長さで演奏されると予測している。これに対し、(b)については「E(短めの4分)-D(短めの4分)-C(短めの4分)」が予測されている。ただし最初のE音はスナップショットの時点ですでに演奏が終わっていると予測されているため、図には表されていない。この予測は「前の小節「D(2分)-F(4分)」は、そのさらに1小節と1拍前の「F(4分)-A(8分)」の音長を倍にして繰り返したものであると認識したことによるものであろうと考えられる。(b)は記憶容量が小さく、繰返し前のパターンを忘却しているため、第1候補のみの予測となった。

#### 6.4 実験IV:音程知覚モデルの構成成分比に関する実験

近接して感じられる音程には個人差がある。実験IVでは音程知覚モデルの構成成分比(表2)を変化させ、図14の旋律を演奏し、予測を行った。図14の繰返し時における予測結果の出力画面を図16に示す。音程知覚モデルに関するパラメータを表3のように設定した。それ以外のパラメータはいずれも最小周期4,最大周期12とし、先3音の予測を第2候補まであげている。

(a),(c)は図15の(a)と、(b)は図15の(b)と同じ予測結果となった。このように強調する音程の変化を予測結果に反映させることができる。これは近接していると知覚する音程が異なると、予測結果も異なったものになることを示している。

音程知覚モデルに関しては、モデルによって、音の

高さの近接性がどのように変化するのか、ということがまだ明らかでなく、今回与えた数値も実験的なもので、現段階ではパラメータの違いによって予測も変化することを示すにとどまっております、この実験に関する考察は今後の文献で示していきたい。

#### 7. おわりに

本研究では演奏された旋律から、次に演奏される旋律のリズムとメロディーをリアルタイムに予測するシステムの構築を行った。本システムは音楽に対する人間の認知機構をモデル化しており、パラメータを変化させることで、その予測の多様性を示すことができることを示した。

本研究では、予測に関連の深い、記憶のしやすさという要因について触れていない。これを実現するには、情報理論などを用いた記憶・予測のしやすさ<sup>6)</sup>や(音楽に関する)ゲシュタルト理論<sup>7)</sup>を考慮して、現段階では単音記憶数として固定している記憶の容量を自動的に変化させる適応型予測機構の導入を図ることが求められる。これには小さなまとまりを1つのチャンクとして記憶させ、より大きなまとまりで認識させるというような手法が有効であると考えている。また、本研究で扱っている旋律の類似性や旋律の予測に関しては、Meyer<sup>8)</sup>やその流れをくむNarmour<sup>9)</sup>らの「暗意(implication)-実現(realization)モデル」が確立されている。本研究はそのような理論をはじめ、ゲシュタルトや音楽的な慣例を含め、スキーマとして明示的にあげられているものより、信号に近い情報を対象に実験を行ったが、理論との比較検証を行っていく必要があると思われる。さらに、本論文では本システムがパラメータを変化させることで予測が変化することを示し、パラメータのチューニングによって、人間に近い予測を行わせることが可能であると述べた。しかし、それをチューニングし、人間の音楽予測モデルとするには、未発表の曲を対象とし、実際の人間を用いた被験者実験が必要であるため、今後検討していくつもりである。また、本システムは信号処理的な方法で予測を行っており、音楽的妥当性を検証していないため、調性に合わない音が予測されることがある。今後、音楽理論などを取り入れ、より音楽的に妥当な予測がなされるようにしたいと考えている。

本システムはさまざまな音楽分野において発展が期待される。まず、音楽生成の分野においては、本システムはリアルタイム動作が可能であり、MIDIキーボードの音量の情報もパターン間の距離の導出に用いることができるため、たとえば拍や小節の先頭は強拍と

なりやすいという性質を利用することで、拍や小節の認識を行うことが可能であると考えられる。したがって、近年さかんに行われてきたビートトラックの予測システムとしての発展が可能である。また本システムはリズムだけでなく、メロディーの予測もあわせて行われることから、セッションシステムの予測機構としての応用も可能である。また、音楽認知の分野においても応用が考えられる。本システムは音程知覚モデルや記憶に関するパラメータを変化させることで、それらの要素の予測に与える影響をリアルタイムに知ることができる。したがって、これまでのオフライン研究では明らかにされ難かった認知的要素が旋律予測に与える影響を、本システムではリアルタイムで観測することができる。さらにパラメータの変化、すなわちチューニングの様子をリアルタイムにフィードバックし、より人間に近いモデルとしていくことで、これまで多くの認知科学者により研究されてきた記憶の容量、記憶のしやすさなどの認知モデルの真偽をリアルタイムに実証的に知ることができるようになり、新たな知見を得るために有用なシステムとなるであろう。

謝辞 本研究の一部は(財)日産科学振興財団の支援を受けています。

### 参 考 文 献

- 1) Conklin, D. and Witten, I.H.: Multiple Viewpoint Systems for Music Prediction, *Journal of New Music Research*, Vol.24, No.1, pp.51-73 (1995).
- 2) Vercoe, B.: Perceptually-based music pattern recognition and response, *Proc. 3rd International Conference for the Perception and Cognition of Music*, pp.59-60 (1994).
- 3) メイヤー, R.E. (著), 多鹿秀継 (訳): 認知心理学のすすめ, pp.31-36, サイエンス社 (1983).
- 4) 梅本 堯夫: 音楽心理学, pp.82-83, 誠信書房 (1978).
- 5) Peterson, L.R. and Peterson, M.J.: Short-term refraction of individual verbal items, *Journal of experimental Psychology*, Vol.58, pp.193-198 (1959).
- 6) Katayose, H., et al.: A Model of Pattern for Music, *Proc. ICMC*, pp.505-506 (1995).
- 7) 平賀 譲: 楽曲分析における認知, 音楽と認知, 波多野 誼余夫 (編), pp.12-17, 東京大学出版会 (1987).
- 8) Meyer, L.B.: *Explaining Music*, Univ. of Chicago Press (1973).
- 9) Narmour, E.: *The Analysis and Cognition of Melodic Complexity*, Univ. of Chicago Press (1992).

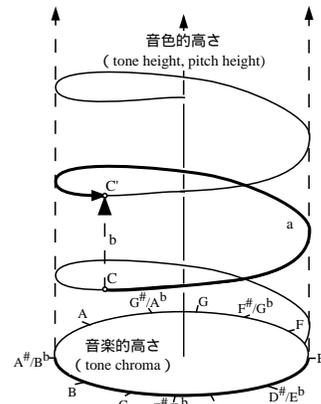


図 17 音程知覚に関する単純螺旋モデル

Fig. 17 Simple helix model of pitch perception.

- 10) Shepard, R.N.: Structural Representations of Musical Pitch, *The Psychology of Music*, Deutsch, D.(Ed.), pp.343-390, Academic Press, New York (1982).

### 付録 Shepard の音程知覚モデル<sup>10)</sup>

Shepard は音楽における音の高さについて、知覚的に重要な音の間の関係が空間的近接関係として直接反映されるような認知構造モデルを示した。

人間は物理的な音高差とは別の尺度で音程の知覚を行っている場合があると考えられる。たとえばオクターブの音程を持つ 2 つの音はそれより半音狭い音程である長 7 度よりも近接して感じられる。このオクターブ類似性を考慮すると単純螺旋構造の音程知覚モデルが導出される(図 17 参照)。すなわち図 17 中の C-C' はオクターブの関係を表しているが、経路 a と比較して経路 b を短くすることでオクターブ類似性を表現することができる。

この螺旋構造で高さ方向に表されている直線的成分と円周方向で表されている循環的成分は、一般にそれぞれ音色的高さ (tone height) と、クロマあるいは音楽的高さ (tone chroma) と呼ばれている。さらにオクターブ以外にも、たとえば完全 5 度や長 3 度のような音程も特別な関係があることが音楽理論、音響心理学的理論、経験的データなどによって示されており、それらの関係も単純な円成分や直線成分から導出することが可能である(表 2)。また、どの音程成分がより優位に働くかは円の直径や直線の長さを変化させることで表現できる。

(平成 11 年 1 月 8 日受付)

(平成 11 年 11 月 4 日採録)



松尾 聡子(正会員)

1966年大阪大学基礎工学部システム工学科卒業。1998年同大学院修士課程修了。在学中、音楽情報処理の研究に従事。1998年三菱電機(株)に入社。産業システム研究所に勤務。主に、発電プラントのヒューマンファクター、ヒューマンインタフェースの研究に従事。



片寄 晴弘(正会員)

1986年大阪大学基礎工学部制御工学科卒業。1991年同大学院博士課程修了。工学博士。同年(株)オーヂス総研入社。1992年度より(財)イメージ情報科学研究所にて、音楽情報処理、感性情報処理、インタラクティブアート制作の研究に従事。主任研究員。1997年度より、和歌山大学システム工学部助教授。1990年情報処理学会学術奨励賞受賞。電子情報通信学会、人工知能学会、ICMA各会員。



井口 征士(正会員)

1940年生。1962年大阪大学工学部電気工学科卒業。1964年同大学院修士課程修了。1965年基礎工学部制御工学科助手。1969年助教授。1984年教授(現大学院基礎工学研究科システム人間系)(財)イメージ情報科学研究所関西研究センター所長兼任。1966~67年フランス・サクレ原子力研究所。1979~80年南カリフォルニア大学画像処理研究所客員。工学博士。専門はパターン計測、特に三次元画像・音響信号等の計測。感性情報処理。著書「三次元画像計測」(昭晃堂)、「感性情報処理」(オーム社)等。電子情報通信学会、計測自動制御学会、システム制御情報学会、人工知能学会、IEEE等会員。