

# FATTA に基づくメロディ予測システム

浜中雅俊<sup>†1</sup> 平田圭二<sup>†2</sup> 東条敏<sup>†3</sup>

<sup>†1</sup>筑波大学大学院システム情報工学研究科

<sup>†2</sup>NTT コミュニケーション科学基礎研究所

<sup>†3</sup>北陸先端科学技術大学院大学

hamanaka@iit.tsukuba.ac.jp

本稿では、自動楽曲分析器 FATTA(Full-Automatic Time-span Tree Analyzer)に基づき、現在演奏中のメロディ(単旋律)の後続音を予測する手法を提案する。従来多くのメロディ予測手法では、統計的学習手法に基づいていたために、学習データに含まれていないメロディに対しては適切でない予測となる可能性があった。これに対し本研究では、音楽理論に基づきメロディの適切さを評価し予測を行う。具体的には、音楽理論 GTTM (A Generative Theory of Tonal Music) および TPS (Tonal Pitch Space) に基づき構築した自動楽曲分析器 FATTA を用いて、複数存在する後続音の候補に対してメロディの安定度を評価する。本手法の特長は、メロディの表層構造だけでなく GTTM および TPS に基づき得られる音楽の深層構造を利用してメロディの安定度を評価している点である。実験の結果、提案手法がメロディの後続音を適切に予測していることを確認した。

## Melody Expectation Method based on FATTA

Masatoshi Hamanaka<sup>†1</sup> Keiji Hirata<sup>†2</sup> Satoshi Tojo<sup>†3</sup>

<sup>†1</sup> Department of Intelligent Interaction Technologies, University of Tsukuba,

<sup>†2</sup> NTT Communication Science Laboratories,

<sup>†3</sup> Japan Advanced Institute of Science and Technology

1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8573 Japan

A method that predicts the next notes is described for assisting musical novices to play improvisations. Melody prediction is one of the most difficult problems in musical information retrieval because composers and players may or may not create melodies that conform to our expectation. The development of a melody expectation method is thus important for building a system that supports musical novices because melody expectation is one of the most basic skills for a musician. Unlike most previous prediction methods, which use statistical learning, our method evaluates the appropriateness of each candidate note from the view point of musical theory. In particular, it uses the concept of melody stability based on the generative theory of tonal music (GTTM) and the tonal pitch space (TPS) to evaluate the appropriateness of the melody. It can thus predict the candidate next notes not only from the surface structure of the melody but also from the deeper structure of the melody acquired by GTTM and TPS analysis. Experimental results showed that the method can evaluate the appropriateness of the melody sufficiently well.

### 1. はじめに

本稿では、音楽理論 GTTM (Generative Theory of Tonal Music) [1]および TPS (Tonal Pitch Space) [2]に基づき、現在演奏中のメロディ(単旋律)の後続音を予測する手法について述べる。本研究の目的は、今演奏中のメロディの後続音の候補を提示することで、音楽初心者

の即興演奏を支援することである。図 1 は、本研究で提案するメロディ予測手法を用いて、音楽初心者の即興演奏を支援するために我々が構築した予測ピアノである。予測ピアノの天板には、縦 32 個、横 25 個のフルカラーLED が配置されており、時間とともに縦スクロールしながら予測結果を表示する。



図 1: 予測ピアノ

我々は、予測ピアノを用いて音楽初心者が即興演奏することを可能とする上で以下の 2 点が重要だと考えている。

- 1) 新規なメロディに対しても適切な予測ができること.
- 2) 予測結果が音楽的観点から見て適切であること.

従来のメロディ予測手法は、統計的学習手法に基づく手法[3,4]と音楽理論に基づく手法[5]の 2 つに大別できる。統計的学習手法に基づくメロディ予測では、予測結果は学習データに依存する。したがって、学習データと予測するデータの作曲家、ジャンル、時代、地域などが異なる場合には、予測精度が低くなるという問題がある。また、学習データにない未知のメロディに対して予測した場合、その結果が適切であるかわからないという問題がある。

これに対し、本研究では音楽理論に基づく予測を実現することで、新規のメロディに対しても適切な予測を可能とすることを目指す。これまで多くの音楽理論が提案されてきたが[1,2,5-8]、音楽理論として GTTM を採用する。その理由は、GTTM は音楽が備える多様な側面を包括的に表象していることと、音楽構造の安定度という概念を持っていることである。音楽構造の安定度という概念を用いると、予測結果が音楽的観点から見て適切であるかどうか評価することが可能となる。

本稿では、まず 2 節で音楽理論 GTTM および我々がこれまで構築してきた実装可能な音楽理論 exGTTM および自動楽曲分析器 FATTA について説明する。次に 3 節で、メロディ予測手法の提案および予測ピアノの構築について述べる。そして、5 節で実験について述べ、最後に 6 章でまとめと今後の課題を述べる。

## 2. GTTM に基づく楽曲分析器

本研究では、音楽理論 GTTM に基づく楽曲分析器 FATTA を用いてメロディの予測を実現する。本節では、その予備知識として GTTM, exGTTM および FATTA の概略を述べる。

### 2.1 音楽理論 GTTM

GTTM は、音楽に関して専門知識のある聴取者の直観を形式的に記述するための理論として Fred Lerdahl と Ray Jackendoff により提唱された。この理論は、グルーピング構造分析、拍節構造分析、タイムスパン簡約、プロロンゲーション簡約という 4 つのサブ理論から構成されており、楽譜に分析を加えることで、楽譜に内在する様々な階層構造を深層構造として顕在化させる (図 2)。GTTM の各サブ理論は、構成ルール (WFR: well-formedness rules) と選好ルール (PR: preference rules) の 2 種類からなる。構成ルールは、構造が成立するために必要な条件や制約を示すルールで、たとえば、「タイムスパンは階層的な構造となっており、各タイムスパンは 1 つのヘッドを持つ」というような制約がルールとして記述されている。一方、選好ルールは、複数の構造が構成ルールを満たす場合どれが好ましいかを示すルールで、たとえば、「拍節構造が安定するタイムスパン木を優先する」というルールがある。

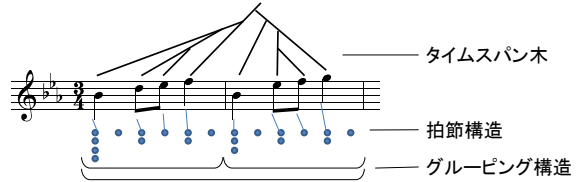


図 2: グルーピング構造, 拍節構造, タイムスパン木

### 2.2 実装可能な音楽理論 exGTTM

GTTM を計算機上へ実装する上での問題は (1) 楽曲分析の曖昧さの問題, (2) 楽曲分析に必要な概念の欠如の問題, の 2 つに大別できる。ここで注意すべきは、(1) 楽曲分析の曖昧さ, もさらに 2 つに分けられることである。音楽理論自体に曖昧さがあるため分析結果が曖昧になることと、楽曲の解釈自体に曖昧性が内在していることで分析結果も曖昧になること, は区別すべきと考える。

上記の問題を解決するため、我々は GTTM の完全パラメータ化および外在化により計算機実装用に拡張した exGTTM を提案した[9,10]。そして、exGTTM を計算機上に実装したシステム ATTA を構築した。計算機実装のための外在化には、大局的な構造に関するルールと局所的な構造に関するルールの両方を適切に組み合わせ階層的な構造を獲得するためのアルゴリズムの導入も含まれる。そのようなアルゴリズムは GTTM[1]では、述べら

れていなかった。また、完全パラメータ化は、ルールの優先度を決め、ルール間の競合を防ぐだけでなく、階層的なタイムスパン木の構造を変化させることを可能とする。もし exGTTM が、楽曲の正しい分析の結果を出力できない場合、我々は新たなパラメータを exGTTM に追加することで、正しい分析結果が出力できるように試みた。このようにして、外在化およびパラメータ化を繰り返し、最終的に、グルーピング構造分析に関する 15 個のパラメータ、拍節構造分析に関する 18 個のパラメータ、タイムスパン簡約に関する 13 個のパラメータを導入した (表 1)。

## 2.3 自動楽曲分析器 FATTA

exGTTM を計算機上に実装したシステム ATTA のパラメータの最適化を自動で行う FATTA を構築した[11]。ATTA は GTTM のタイムスパン木の獲得を可能としていたが、正しい分析結果を得るためには調節可能なパラメータを曲ごとに手動で適切に設定しなければならず、多大な労力がかかるという問題があった。FATTA では、この問題を解決す

るため、ATTA による分析の結果に対して安定度を評価するアルゴリズムを与え、ATTA の各パラメータの値を自動的に最適化することを実現した。具体的には、ATTA に未実装のタイムスパン木・プロロンゲーション木の安定度に関する選好ルール GPR7 および拍節構造の安定度に関する選好ルール TSRPR5 の実装を行った。

### GPR7 (time-span and prolongational stability):

タイムスパン木、プロロンゲーション木が安定するグルーピング構造を優先する。

### TSRPR5 (metrical stability):

タイムスパン木のヘッドを選ぶ際、拍節構造が安定となるものを優先する。

これらのルールは、タイムスパン簡約やプロロンゲーション簡約の結果など GTTM の分析過程の後方で得られる情報と前方にある分析過程とのインタラクションを必要とする。図 3 は ATTA と最適化のためのループからなる FATTA の全体像である。

表 1: exGTTM の調節可能なパラメータ

	パラメータ	説明
グルーピング構造分析	$S_{GPRR}$ ( $0 \leq S_{GPRR} \leq 1$ )	グルーピング構造分析の各ルールの強さを表すパラメータ。値が大きいほど、ルールの強さが強くなる。 $R \in \{ 2a, 2b, 3a, 3b, 3c, 3d, 4, 5, \text{ and } 6 \}$
	$\sigma$ ( $0 \leq \sigma \leq 0.1$ )	GPR5 で用いる平均をグループの中心とする正規分布の標準偏差。値が大きくなるほど正規分布の裾野が広がる。
	$W_m$ ( $0 \leq W_m \leq 1$ )	GPR6 で、各音の発音時刻の類似度と音高差の類似度のどちらを重視するか決めるパラメータ。値が大きいほど、音高差の類似度を重視する。
	$W_l$ ( $0 \leq W_l \leq 1$ )	GPR6 で、並列な区間の長さをどのくらい重視するかを決めるパラメータ。値が大きいほど長い並列的な区間を重視する。
	$W_s$ ( $0 \leq W_s \leq 1$ )	GPR6 で、音符 $i$ が並列な区間の始端あるいは終端のいずれになる方を重視するかを調整するパラメータ。値が大きいほど終端を重視する。
	$T_{GPR4}$ ( $0 \leq T_{GPR4} \leq 1$ )	GPR4 で、GPR2,3 の効果が明白であるかどうかを決める閾値。値が小さいほど、GPR4 が成立しやすくなる。
	$T^{low}$ ( $0 \leq T^{low} \leq 1$ )	低レベルの境界であるかどうかを決める閾値。値が小さいほど、境界と認識されやすくなる。
拍節構造分析	$S_{MPRR}$ ( $0 \leq S_{MPRR} \leq 1$ )	拍節構造分析の各ルールの強さを表すパラメータ。値が大きいほど、ルールの強さが強くなる。 $R \in \{ 1, 2, 3, 4, 5a, 5b, 5c, 5d, 5e, \text{ and } 10 \}$
	$W_m$ ( $0 \leq W_m \leq 1$ )	MPR1 で、各音の発音時刻の類似度と音高差の類似度のどちらを重視するか決めるパラメータ。値が大きいほど、音高差の類似度を重視する。
	$W_l$ ( $0 \leq W_l \leq 1$ )	MPR1 で、並列な区間の長さをどのくらい重視するかを決めるパラメータ。値が大きいほど長い並列的な区間を重視する。
	$W_s$ ( $0 \leq W_s \leq 1$ )	MPR1 で、拍点 $i$ が並列な区間の始端あるいは終端のいずれになる方を重視するかを調整するパラメータ。値が大きいほど終端を重視する。
	$T_{MPRR}$ ( $0 \leq T_{MPRR} \leq 1$ )	MPR の各ルールが成立するかどうかを決める閾値。 $R \in \{ 4, 5a, 5b, 5c \}$
タイムスパン簡約	$S_{TSRPRR}$ ( $0 \leq S_{TSRPRR} \leq 1$ )	タイムスパン簡約の各ルールの強さを表すパラメータ。値が大きいほど、ルールの強さが強くなる。 $R \in \{ 1, 3a, 3b, 4, 8, \text{ and } 9 \}$
	$W_m$ ( $0 \leq W_m \leq 1$ )	TSRPR4 で、各音の発音時刻の類似度と音高差の類似度のどちらを重視するか決めるパラメータ。値が大きいほど、音高差の類似度を重視する。
	$W_l$ ( $0 \leq W_l \leq 1$ )	TSRPR4 で、並列な区間の長さをどのくらい重視するかを決めるパラメータ。値が大きいほど長い並列的な区間を重視する。
	$W_s$ ( $0 \leq W_s \leq 1$ )	TSRPR4 で、ヘッド $i$ が並列な区間の始端あるいは終端のいずれになる方を重視するかを調整するパラメータ。値が大きいほど終端を重視する。

2.3.1 Tonal Pitch Spaceに基づくGPR7の実装  
GPR7は、タイムスパン簡約およびプロロンゲーション簡約からグルーピング構造分析へのループに関するルールであり、タイムスパン木(TS木)やプロロンゲーション木が安定するグルーピング構造を優先する。 $D_{GPR7}$ はGPR7が成立する度合いで、0から1までの連続値をとる。 $D_{GPR7}$ を次式のように定義する。

$$D_{GPR7} = \frac{\sum_i \text{distance}(p(i), s(i)) \times \text{size}(i)^2}{\sum_i \text{size}(i)^2} \quad (1)$$

$i$ は、TS木のヘッドを表す。TS木は2分木であるから、各ヘッドは重要な枝 *primary* と重要でない枝 *secondary* をもつ。そして、 $p(i)$  は  $i$  の *primary* の音符、 $s(i)$  は  $i$  の *secondary* の音符を表す。 $\text{distance}(x, y)$  は音符  $x$  と  $y$  の距離を TPS(Tonal Pitch Space)[14] に基づき求めた値である。0から1の値となるよう正規化している。 $\text{size}(i)$ は、タイムスパンの長さで、2乗したものを重みとして掛けて正規化している。 $\text{size}(i)$ や  $\log \text{size}(i)$  など様々なものを試した結果、一番性能が良くなったため、2乗したものを採用した。

TPSにおいて和音  $x=C_1/R_1$  と和音  $y=C_2/R_2$  の距離は次式で定義される。

$$\delta(x \rightarrow y) = i + j + k \quad (2)$$

$i$ は調距離、 $j$ はコード距離、 $k$ は基本空間距離である。調距離は  $R_1$  と  $R_2$  の距離、コード距離は、 $C_1$  と  $C_2$  の距離、基本空間距離は、 $x$  の基本空間と  $y$  の基本空間の違いで、それぞれ TPS で定義されている。式1は、TS木の各ヘッドの音符間の距離 (TPSで定義される距離) が近いほど大きな値となるため、たとえば、 $i$  が直上のタイムスパンの *primary* のヘッドと等しいなら距離はゼロとなり最も安定となる。

### 2.3.2 TSRPR5の実装

TSRPR5を実装するためには、タイムスパン簡約の結果がどのくらい拍節的に安定しているかを評価する必要がある。本研究では、タイムスパン木と整合性の高い拍節構造が安定な拍節構造であると考え、次式でタイムスパン簡約の結果がどのくらい拍節的に安定しているか、すなわち TSRPR5が成立する度合いを定義する。

$\text{dot}(x)$ は拍節構造分析の結果として求まる各音符の拍点の数である。 $D_{TSRPR5}$ は0から1の値を出力し、タイムスパン簡約の結果重要と判断とされた音符が拍節構造分析でも重要と判断されているほど大きな値となる。

### 2.3.3 適切なパラメータセットの獲得

タイムスパン木の安定度  $D_{GPR7}$  および拍節構造の安定度  $D_{TSRPR5}$  の両方が大きくなるような ATTA の各パラメータの値を適切なパラメータセットとして獲得する。ここでは、 $D_{GPR7}$  と  $D_{TSRPR5}$  の相加平均が最大となるようなパラメータの値を探索する。ATTAには46個の調節可能なパラメータがあるため、探索の際、全探索を行うと効率が悪い。そこで以下のようなアルゴリズムを用いることにした。

- (1) あるひとつのパラメータを値域の最小から最大まで動かし、 $D_{GPR7}$  と  $D_{TSRPR5}$  の相加平均が最大となるような値に設定する。ただし、パラメータの値は値域の10分の1ずつ変化させるものとする。相加平均が最大となるような値がいくつもある場合には、値域の中央値に近いほうの値とする。
- (2) (1) をすべてのパラメータについて行う。
- (3)  $D_{GPR7}$  と  $D_{TSRPR5}$  の相加平均が前回の繰り返しよりも大きくなるまで (2) を繰り返す。

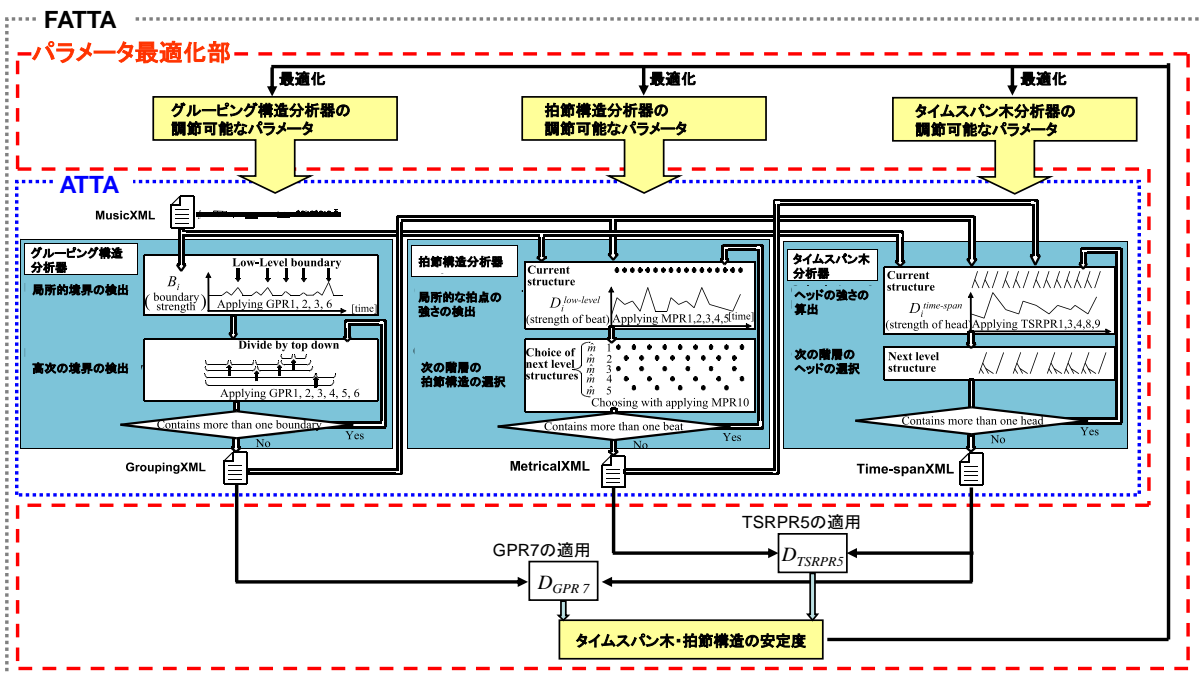


図3: FATTAの処理の流れ

### 3. メロディ予測手法

本メロディ予測手法では、FATTA に基づき算出されるタイムスパン木の安定度を用いて、今のメロディの後続音の候補を予測する。FATTA は、単旋律の西洋調性音楽を対象としているため、本手法も同様に単旋律の西洋調性音楽の後続音の予測を対象とする。我々は、唯一の後続音を予測するという事は困難であると考え、安定度の高い音を候補として複数提示することにした。そしてそのような提示を可能とするインタフェース予測ピアノを構築した。

#### 3.1 FATTA に基づくメロディ予測

音楽に興味深くしている理由の一つは、予測した音が遅れるなど、聴取者の予測と一致しないことが、むしろ緊張と弛緩をうみ出していることである。作曲家は、曲が単調とならないよう注意深く曲を作っているため、初めて聴く曲の後続音を正確に予測することは一般に困難である。たとえば、予測した音の前にしばしば装飾音が挿入されることがある。このような場合でも、我々が提案する予測手法では、後続音の候補を適切に予測することができる。なぜなら、FATTA はタイムスパン木全体の安定度を算出しており、しかもそのタイムスパン木では、重要な音が幹に装飾的な音が葉に接続されている。したがって、装飾音を予測できなかった場合でも、その次にくる重要な音を予測することが可能となる。

#### 3.2 FATTA のリアルタイム化

GTTM に基づく楽曲分析器 FATTA を用いてメロディの予測をするためには、FATTA をリアルタイムで動作させる必要がある。しかし、FATTA は 8 小節の楽曲を分析するのに数分かかり、リアルタイムで動作することが困難である。そこで我々は、リアルタイム化のためにアルゴリズムを改良した。まず第 1 に、2.3.3 で述べた繰り返し処理の収束が早くなるよう、パラメータの初期値を直前の音が入る前のパラメータセットを用いることにした。第 2 に、ATTA での分析時間を短縮するため、窓を設定した。窓のサイズは、現在から始まるグループで 16 小節以内の最大のものとした。16 小節以内にグルーピング境界がない場合、窓のサイズは 16 小節とした。

#### 3.3 FATTA および TPS に基づく安定度の算出

本手法では、FATTA を用いてメロディの安定度を算出することで、後続音の各候補に対し

てその適切さを評価する。安定度としては、2.3.3 と同様に  $D_{GPR7}$  と  $D_{TSRPR5}$  の相加平均を用いる。図 4 は、先頭から現在の音までの安定度を算出したものである。GTTM のルールを適用するためには、最低 4 音必要であるため、安定度は 4 音以降から算出される。図 4 では、第 1 音は調の主音、同様に最後の音も主音となっており、安定度はメロディの終わりで大きな値となり、途中では小さな値となっている。つまり、調の主音に関係の深い音ほど安定であると言えることができる。メロディの調およびコード進行は、GPR7 を適用するために実装した TPS が自動推定している。

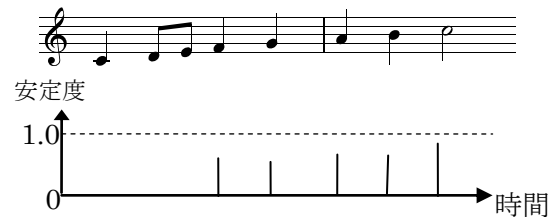


図 4: 現在の音までの安定度

#### 3.4 予測ピアノ

演奏初心者の即興演奏を支援する目的でピアノの天板上に予測結果を表示する予測ピアノを構築した。初心者が即興演奏中に次の音を迷った場合には、天板上に表示された予測結果の通りに演奏することで、演奏を続けることができる。天板の下には、鍵盤の幅に合わせて横 25 個×縦 32 個のフルカラーLED が配置されており、ピアノロール形式で予測結果が表示できる。25 は鍵盤の数、32 は 2 小節 (16 分音符の分解能) を表している。

図 5 は、予測ピアノの処理の流れを表している。まず、鍵盤から MIDI 信号が計算機に送られ、音価の量子化が行われ、演奏中のメロディが MusicXML の形で出力される。そして、次に来る可能性のある音候補 (25 音×32) の全部に対して FATTA を用いてリアルタイムで安定度を計算し、天板上に安定度の高い音ほど明るく (高い輝度で) 表示する。予測結果は、演奏に合わせて一定スピードで縦スクロールする。

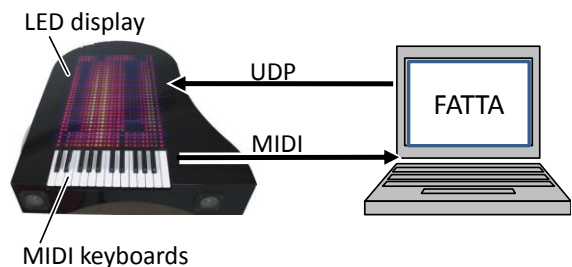


図 5: 予測ピアノ



## 4. 実験結果

従来のメロディ予測システムの多くは統計的手法に基づいているのに対し、本システムは音楽理論に基づいており、アプローチが完全に異なる両者を実験により比較することは難しい。ここでは、メロディ予測結果の例を挙げそれについて検討する。

図6は、ハイドン作曲のアンダンテを演奏したときの予測結果である。楽譜の下の棒グラフは、それぞれの音符の安定度を示している。音符の下の数字は、鍵盤にある25音の中での安定度の順位を表している。図6の結果では、7小節目のF#の音の安定度が最低となっていなかったり、5小節目のCの音はこの部分がCメジャーコードであるにもかかわらず4小節目のGの音に比べて安定度が低くなっているなど、我々の直感と異なる部分がある。これは、安定度の計算が、現在の1音のみを評価しているわけではなく、曲の始めから現在までの全体からなるタイムスパン木を評価しているためである。4小節目と8小節目の終わりでは安定度が高くなっているが、これは、ドミナントコードがトニックコードに解決したがる性質を持っているためである。一方、5小節目の始めでは安定度が比較的低いが、これは、トニックコードの主音が5小節目の頭で登場しているため、後続音に様々な可能性が生じ、相対的に安定度が低くなっているためである。

上記の通り図6の結果では、音楽的な観点から見て適切な予測ができていたことが確認できた。

## 5. おわりに

本稿では、音楽理論 GTTM および TPS に基づき、演奏中のメロディの後続音を予測するメロディ予測手法について述べた。本手法に基づくメロディ予測結果を、予測ピアノの天板上に表示することで初心者即興演奏を支援することが可能となる。実際に分析例を確認したところ、音楽的な観点から見て適切な予測ができていたことを確認した。今後、予測ピアノによる即興演奏支援について評価していく。また、GTTMのプロロンゲーション簡約を含めた予測手法についても検討していく。

## 参考文献

- [1] Lerdahl, F. and Jackendoff, R.. A Generative Theory of Tonal Music. MIT Press, Cambridge, Massachusetts, 1983.
- [2] Lerdahl, F. Tonal Pitch Space, Oxford University Press, 2001.
- [3] Mozer, M. Neural Network Music Composition by Prediction: Exploring the Benefits of Psychoacoustic Constraints and Multi-Scale Processing. Connection-Science. Vol. 6, No. 2-3, pp. 247-280, 1994.
- [4] Yen, G., Lucas, S., Fogel, G., Kendall, G., Salomon, R., Zhang, B., Coello, C., and Runarsson, T. "Evolving Musical Sequences with N-Gram Based Trainable Fitness Functions", Proceedings of the 2006 IEEE Congress on Evolutionary Computation, pp. 604-614, 2006.
- [5] Cooper, G., and Meyer, L. B. The Rhythmic Structure of Music. The University of Chicago Press, 1960.
- [6] Narmour, E. The Analysis and Cognition of Basic Melodic Structure. The University of Chicago Press, 1990.
- [7] Temperley, D. The Cognition of Basic Musical Structures. MIT press, Cambridge, 2001.
- [8] Larson, S. Musical Forces and Melodic Expectations: Comparing Computer Models with Experimental Results. Music Perception. 21/4, 457-498.
- [9] Hamanaka, M., Hirata, K., and Tojo, S.: Implementing "A Generating Theory of Tonal Music", Journal of New Music Research (JNMR), Vol. 35, No. 4, pp. 249-277, 2006.
- [10] Hamanaka, M., Hirata, K., and Tojo, S.: "ATTA: Automatic Time-span Tree Analyzer based on Extended GTTM", Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR2005), pp. 358-365, 2005.
- [11] Hamanaka, M., Hirata, K., and Tojo, S.: "FATTA: Full Automatic Time-span Tree Analyzer", Proceedings of the 2007 International Computer Music Conference (ICMC2007), Vol. 1, pp. 153-156, 2007.

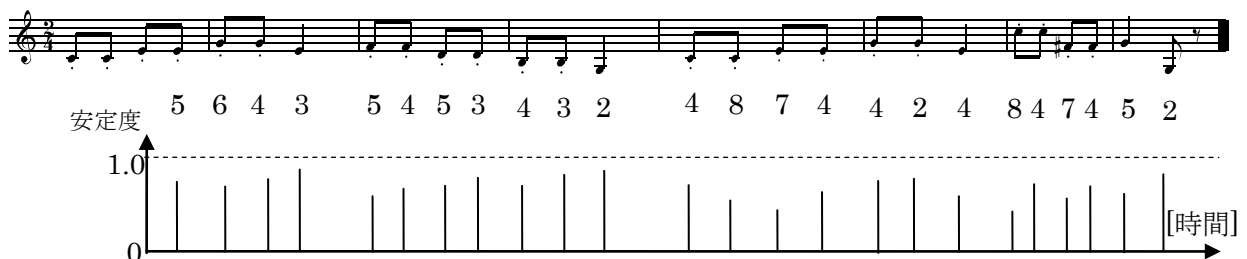


図6: メロディ予測結果