

# グルーピング構造に着目した確率的メロディー生成手法の提案

A proposal of a method for generating melodies by using a probability theory and grouping structure

白井 亨\*1      谷口 忠大\*1      西川 郁子\*1  
Akira Shirai      Tadahiro Taniguchi      Ikuko Nishikawa

\*1 立命館大学  
Ritsumeikan University

Recently, music composition using computer software is gathering attentions. Many people want to compose their original music. However, musical composition is still too difficult for beginners to obtain their favorable original music. In this paper, we propose a novel method for generating a melody by using a generative probabilistic model using Gibbs sampler and grouping structure proposed in GTTM. This method easily creates a melody in the motif of an existing music. We also evaluate the probabilistic music composition model by some experiments.

## 1. はじめに

本稿では音楽理論 Generative Theory of Tonal Music[Lerdahl 83](以下 GTTM) によるグルーピング構造を制約として用い、規範とする既存楽曲から新たなメロディーをギブスサンプリングにより確率的に自動生成する手法を提案する。この手法によりユーザは既存の楽曲を基に自分の曲想に合った新しい楽曲を生成することが可能になると考えられる。

## 2. 研究背景

### 2.1 現状の問題点

現在音楽初心者のための作曲支援ツールとして自動作曲システムが存在しており (ACS\*1, jac\*2など), これらのシステムではユーザがパラメータとして曲想 (明るい, 暗い, ジャズ風, 沖縄風など) を選択することでその曲想に合った楽曲を生成することが可能である。しかし, 作曲に使用するパラメータの持つ曲想がユーザの持つ曲想と違った場合, ユーザの望む楽曲を生成できないという問題がある [蓮井 09]。

このような問題に対して対話型進化計算を用いユーザの感性をシステムに取り入れて対話的にユーザの曲想に合う楽曲を生成する研究が行われている [高木 98]。この方法を用いることでユーザは自分の曲想に合う楽曲を生成することができるが, 解として数十秒から数分あるメロディーをすべて聞いて評価することになりユーザの疲労が増大してしまうため, 結果的に世代数を少なくする必要があり適切な解に近づけないという問題がある。これに対し鈴木らは立体音響の中で仮想的な空間を歩くという方法でユーザの負担を和らげつつも直感的な作曲を行うことができる手法を提案している [鈴木 08]。しかし世代が若いときは楽曲の断片のみを評価しなければならないためやはりユーザの負担は大きくなる。対話型進化計算以外の自動作曲手法では嵯峨山らの Orpheus[深山 07] がある。Orpheusでは日本語の韻律を制約として動的計画法により自動作曲を行うという手法を提案している。これによりユーザが与えた歌詞のイメージに合う楽曲の自動生成を行っている。しかし, 楽曲

を最適化してしまうためユーザに選択の余地がないという問題がある。そこで出来る限り適切な楽曲は生成しつつも, ユーザがよい曲を対話的に選択する余地は残す設計が必要になる。

### 2.2 確率的なメロディー生成

メロディーには言語ほどの文法的制約はない。しかし, ポップスにはポップスのメロディー, ジャズにはジャズのメロディーがあるようにある様式を与えることで音楽は楽曲として成立する。様式にはそれぞれ出やすい音, 遷移しやすい音が決まっている。つまりメロディーはある様式がもつ情報をもとに確率的に生成されるものと考えられる。そこで, 本研究では既存楽曲を蓄えたコーパスからの情報をもとにして確率的にメロディーの生成を行う確率モデルを提案する。ただし, 本提案モデルではメロディーを最適化するのではなく事後確率の高い幾つかのメロディーを生成結果としてユーザに提示する。

### 2.3 GTTM による制約

しかしながら統計的な音高遷移のみを用いた楽曲生成では, ユーザの感性を反映させることができないと考えられる。しかし対話型進化計算を用いた楽曲生成の研究からも人間の感性を生成に用いることはユーザの曲想に合う楽曲を生成することにおいて重要である事が示唆されている。そこで, 本研究ではコーパス情報にもとづくメロディー生成に制約として音楽理論 GTTM を用いることでこの解決を目指す。音楽理論 GTTM とは音楽知識のある人間がどのように楽曲の構造を認知しているかに着目した理論であり, 旋律の区切りとリズムや韻律をもとに, 旋律や和声を本質的な部分と装飾的な部分に区分する手順が提案されている。GTTM はグルーピング構造, 拍節構造, 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いる。これらの理論で抽出された情報は人間の感性に基づいていることから, GTTM によってユーザの曲想に合う既存楽曲から抽出された情報を制約とすることで生成されるメロディーをユーザの曲想に近づけることができると考えられる。またユーザの曲想を既存楽曲という形で与えることで対話型進化計算のように繰り返し感性評価を行う必要がなくなるのでユーザの負担を軽減することが可能であると考えられる。

連絡先: 白井亨, 立命館大学, 創発システム研究室, 滋賀県草津市野路東 1-1-1, E-mail:shirai@em.ci.ritsumeikan.ac.jp

\*1 <http://hp.vector.co.jp/authors/VA014815/music/autocomp.html>

\*2 <http://www.vector.co.jp/soft/win95/art/se309509.html>

### 3. 提案手法

#### 3.1 概要

本研究で提案する手法の概念図を図 1 に示す．本研究では

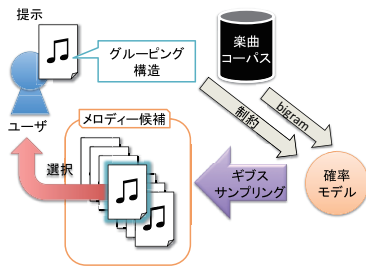


図 1: 提案手法の概念図

既存楽曲を蓄えたコーパス情報にもとづき確率的にメロディーを生成する確率モデルを提案する．具体的にはコーパスより抽出した音高遷移確率の bigram モデルからメロディーが生成される確率モデルを構築する．さらに，この確率モデルにユーザの作曲したい曲想に近い楽曲より抽出されたグルーピング構造を制約として与え，生成確率の高い幾つかのメロディーを生成結果としてユーザに提示する．これらの生成メロディーはユーザの曲想に合うものであることが期待される．

#### 3.2 楽曲コーパスの定義

既存楽曲をデータ化し蓄積したものを楽曲コーパスとして定義する．楽曲は下図 2 のようにメロディーのノート列  $S = (S_1, S_2, \dots, S_N)$  とコード列  $c = (c_1, c_2, \dots, c_N)$  で表現される． $S$  には 8 分音符長  $\delta$  ごとにメロディーの発音 (MIDI ノートナンバー)，持続，休符などの情報が格納される．また， $c$  には  $\delta$  ごとにコード名が格納される．

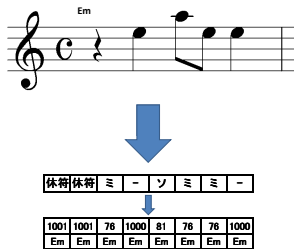


図 2: メロディーをベクトルに変換

#### 3.3 楽曲コーパスからの情報抽出

楽曲コーパスからグルーピング構造と音高遷移確率の抽出を行う．

##### 3.3.1 グルーピング構造の抽出

グルーピング構造の抽出には浜中らの exGTTM[浜中 07] を単純化して用いる．exGTTM ではグルーピング構造を構成するルール GPR1~6 を使用し局所的な境界と高次の境界により階層的なグルーピング構造を求めているが，本研究では GPR1,2a,2b,3a,3d のみを使用し局所的な境界 (0, 1) とその強さ (0~1) を求め，かけ合わせたものをグルーピング構造として使用する．グルーピング構造  $Z = (Z_1, Z_2, \dots, Z_N)$  は以下

の式で求められる．

$$Z_i = \begin{cases} D_{low}(j) \times B_{low}(j) & \text{if } i = \frac{\varepsilon(j)}{\delta} \text{ and } D_{low}(j) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

ここで  $D_{low}(j)$  は  $j$  番目の音符が局所的境界になるかどうか， $B_{low}(j)$  は  $j$  番目の音符の局所的境界の強さ， $\varepsilon(j)$  は  $j$  番目の音符の終音時刻をそれぞれ表している．

##### 3.3.2 音高遷移確率の抽出

音高遷移確率  $\beta$  の抽出には単語 bigram モデルを用いる．あるコード上での音高遷移確率は以下の式で表わされる．

$$p(S_i | S_{i-1}, c_{i-1}) = \frac{C(c_{i-1}, S_{i-1}, S_i)}{C(c_{i-1}, S_{i-1})} \quad (2)$$

ただし  $S, c$  はそれぞれノート列，コード列を表し  $C(S)$  はコーパス上に存在する  $S$  の数を表す．しかし，このままではコーパス上に存在しない遷移の 2-gram 確率が 0 になってしまうという問題が発生する．そこで本研究では以下のようにベースライン補正を行いこれを遷移確率として使用する．

$$p(S_i | S_{i-1}, c_{i-1}) = \frac{C(c_{i-1}, S_{i-1}, S_i) + 0.1}{C(c_{i-1}, S_{i-1}) + (\lambda \times 0.1)} \quad (3)$$

ここで  $\lambda$  はノートの種類数を表す．

#### 3.4 メロディー生成

グルーピング構造を制約とする確率モデルを考える前にまず楽曲コーパスより抽出した音高遷移確率  $\beta$  をもとにして確率的にメロディー  $S$  が生成され，そこからグルーピング構造  $Z$  が抽出されるグルーピング構造の生成モデルを考える．提案するグラフィカルモデルを図 3 に示す．

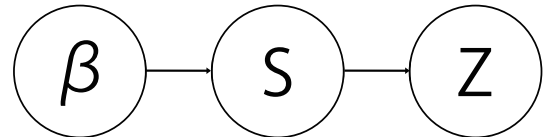


図 3: グルーピング構造の生成モデル

ここでグルーピング構造は本来決定論的に求まるものであるが確率モデルに組み込むためにここでは下式のようにガウス分布から確率的に生成されるものと仮定する．

$$p(Z|S) = \frac{1}{(\sqrt{2\pi})^N \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(Z - \mu(S))^t \Sigma^{-1} (Z - \mu(S))\right) \quad (4)$$

ここで  $\mu(S)$  は生成されたノート列  $S$  から抽出されたグルーピング構造を表し  $\Sigma$  の値を操作することで制約の強さを変えることができる．このグルーピング構造の生成モデルにベイズの定理を用いることで下式のようなグルーピング構造を制約としたメロディーの生成モデルを得ることができる．

$$p(S|Z, \beta) = \frac{p(Z|S)p(S|\beta)}{\sum_{S'} p(Z|S')p(S'|\beta)} \quad (5)$$

しかし， $S$  を変化させる度に分母を計算しなおしていると計算量が膨大になってしまう．

そこで本研究ではギブスサンプリングを用いることでノート列  $S$  の推定を行う．ギブスサンプリングとは 1 つの確率変

数のみを順次変化させてサンプリングを行う手法であり少ない計算量でサンプリングが可能である．まずは上式を次のように変化させる部分  $S_i$  とそれ以外の部分  $S_{\setminus i}$  に分ける．

$$p(S_i = k | S_{\setminus i}, \mathbf{Z}, \beta) \quad (6)$$

ギブスサンプリングの為に  $p(S_i = k | S_{\setminus i}, \mathbf{Z}, \beta)$  と  $p(S_i = l | S_{\setminus i}, \mathbf{Z}, \beta)$  の比が求まればよいので分母や  $k$  に依存しない項は無視することができる．よって上式は次のように表すことができる．

$$p(S_i = k | S_{\setminus i}, \mathbf{Z}, \beta) \propto p(\mathbf{Z} | S_i = k, S_{\setminus i}) p(S_i = k | S_{i-1}, \beta) \times p(S_{i+1} | S_i = k, \beta) \quad (7)$$

上式にギブスサンプリングのアルゴリズムを用いることでメロディーの生成を行うことができる．

## 4. 実験

本提案手法を用い以下の設定，条件の下でメロディーの生成実験を行った．

### 4.1 実験設定

#### 4.1.1 楽曲コーパス

楽曲コーパス蓄える楽曲として日本の j-Pop アーティスト Bump of Chicken, Spitz, the Gospellers, Kinki Kids らの楽曲より 35 曲分の A メロ, B メロを使用した．本研究では簡単化のため分解能を 8 分音符, ノートを 1 オクターブ (60 ~ 71) とした．

#### 4.1.2 評価関数

生成メロディーを評価するための指標としてグルーピング構造の類似度とメロディーの生成確率を用いる．

グルーピング構造の類似度は生成されたメロディーから抽出されたグルーピング構造  $\mu(\mathbf{S})$  と与えた楽曲のグルーピング構造  $\mathbf{Z}$  が似ているかを表す指標であり, これを評価するために 2 乗誤差を用いる．2 乗誤差  $e$  は以下の式で表される．

$$e = \sqrt{\sum_i^N (Z_i - \mu(\mathbf{S})_i)^2} \quad (8)$$

また, 生成されたノート列  $\mathbf{S}$  の生成確率を評価関数  $\psi(\mathbf{S})$  とし, 生成確率の代替値として用いる．

$$\text{生成確率} \propto \psi(\mathbf{S}) = p(\mathbf{Z} | \mathbf{S}) p(\mathbf{S} | \beta) \quad (9)$$

#### 4.1.3 実験方法

まずは楽曲コーパスより似せたいと思う楽曲を選択する．この後, 決められたステップ数のギブスサンプリングを行いメロディーの生成を行う．生成されたメロディーの評価値及びメロディー自体を聞き本提案手法の精度を評価する．

### 4.2 提案手法に対する検証

#### 4.2.1 実験条件

以下の設定条件の下でメロディーを生成し生成確率がどのように変化するかを観察を行った．

設定 1 楽曲コーパスから "Spitz" の "ロビンソン" の A メロを選択し, コード進行は原曲と同じものを与え共分散行列  $\Sigma$  の値を 0.2 としてギブスサンプリングのステップを 2000 回試行し 8 小節分のメロディーを生成する．

設定 2 設定 1 の共分散行列の値を 1.0 として生成を行う．

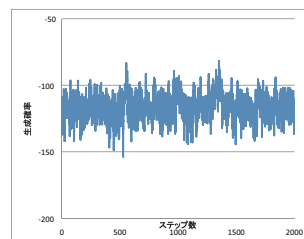


図 4: 生成確率の推移

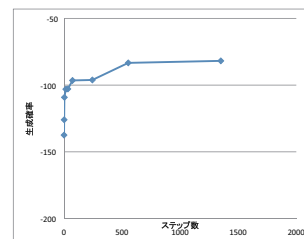


図 5: 最大生成確率の推移

### 4.2.2 結果

図 4 は設定 1 の下で生成されたメロディーの生成確率, 図 5 は最大生成確率を追ったものである．縦軸は生成確率の対数, 横軸はギブスサンプリングのステップ数を表している．図より, サンプリング回数が増えるに連れて徐々に生成確率の高いサンプルが得られるようになっていくことがわかる．

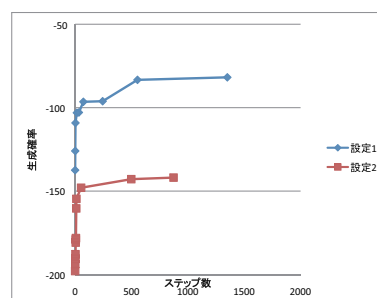


図 6: 最大生成確率の推移 (設定 1, 設定 2)

図 6 は最大生成確率の推移を設定 1 と設定 2 で比較したものである．図より, 設定 1 に比べ設定 2 のほうが全体的に生成確率が低いことがわかる．これは  $\Sigma$  の値を大きくしたことから, グルーピング構造による制約を表現する確率密度関数の値が変化したこと起因するものと考えられる．

### 4.3 生成メロディーに対する検証

#### 4.3.1 実験条件

以下の設定条件の下で生成されたメロディーがグルーピング構造による制約によってどのような影響を受けているか検証を行った．

設定 楽曲コーパスから "the Gospellers" の "ひとり" の A メロを選択し, コード進行は原曲と同じものを与え共分散行列の値を 0.2 としてギブスサンプリングのステップを 2000 回試行し 8 小節分のメロディーを生成する．生成されたメロディーの中で生成確率最大, 中くらい, 最小のものを比較する．

図 7 に "ひとり" の楽譜を示す．赤い示された音符は直後にグルーピング構造の境界が検出されることを表している．

#### 4.3.2 結果

図 8, 9, 10 はそれぞれ生成されたメロディーの中で生成確率の一番低いもの, 中くらいのもの, 一番高いものの楽譜である．また赤く示されている音符は直後にグルーピング構造の境界が現れる音符である．これらは見本楽曲の楽譜に似ているとは言い難い．しかし, 境界の位置が同じような場所に現れるなど見本曲のグルーピング構造による制約をしっかりと

ひとり A



図 7: 見本楽曲 (The Gospellers 「ひとり」より引用)

生成確率 (低)



図 8: 生成確率 (低):1588 ステップ目

生成確率 (中)



図 9: 生成確率 (中):90 ステップ目

生成確率 (高)



図 10: 生成確率 (高):1733 ステップ目

受けていることが見て取れる。また、これらの図から生成されたメロディーはどれも同じ音での繰り返しが多く、特に生成確率が高いメロディーはその傾向が強いことが分かる。これは bigram モデルの影響を強く受けていることが原因と考えられる。本来メロディーは長い流れをもつが、bigram では 1 つ前の音しか考慮しないためこのような結果が現れたと考えられる。また、生成されたメロディーは跳躍を多く含んでしまっている。bigram による影響もあるだろうがグルーピング構造の GPR3(音高差) による影響もあると考えられる。これは図のようにグルーピング構造の境界が検出された場所に音高の変化が

多くみられるからである。

## 5. まとめ

本研究では、は GTTM によるグルーピング構造を制約として用い、規範とする既存楽曲から新たなメロディーをギブスサンプリングにより確率的に自動生成する手法を提案し当該手法の検証を行った。その結果当該手法を用いることで規範楽曲による制約及び遷移確率から複数のメロディーを確率的に自動生成可能なことが実証された。

今後の課題としては、まずは楽曲コーパスの増強があげられる。また、楽曲コーパスからの情報抽出方法を考え直す必要があるだろう。現在はコーパスに存在しない bigram をベースライン補正しているが、これをスムージング技法などを用いてより精度の高い遷移確率を求めることが要求される。さらに、bigram モデルの影響で長いメロディーが現れないという問題については trigram を用いるなどの対応が考えられる。trigram の使用については bigram 以上にコーパス不足が懸念されるため前述の課題を早急に解決していきたい。

また、今回は紙面の都合上被験者による生成メロディーに対する評価実験を載せることができなかった。本稿で述べた楽曲の類似度であり見本楽曲と生成されたメロディーのグルーピング構造が類似しているからと言って実際に人間が聞いて似ていると思うかは分からない。この部分については実際にユーザに生成メロディーを評価してもらうことが必要になるだろう。さらにフィードバックに基づいて、自動作曲装置が新たな曲を生成するというインタラクティブなプロセスを楽曲生成過程に取り込んでいくことが必要となるだろう。

## 参考文献

- [Lerdahl 83] Lerdahl, F. and Jackendoff, R.: A Generative Theory of Tonal Music, The MIT Press, Cambridge (1983).
- [蓮井 09] 蓮井洋志: 作曲モデルを利用した対話型作曲支援システムの作成, 知能と情報 (日本知能情報ファジィ学会誌), Vol.21, No.2, pp.247-255 (2009).
- [高木 98] 高木英行, 畷見達夫, 寺野隆雄: 対話型進化計算法の研究動向, 人工知能学会誌, Vol.13, No.5, pp.692-703 (1998).
- [鈴木 08] 鈴木鈴壘, 三輪真生, 有田隆也: 歩いて創る音楽 - 立体音響を活用した音の適応度地形上の歩行に基づく対話型進化計算 -, MYCOM2008.
- [深山 07] 深山覚, 中妻啓, 米林裕一郎, 酒向慎司, 西本卓也, 小野順貴, 嵯峨山茂樹: Orpheus: 歌詞の韻律に基づいた自動作曲システム, 情報処理学会研究報告, 2008-MUS-76, pp.179-184, 2008.
- [浜中 07] 浜中雅俊, 平田圭二, 東条敏: 音楽理論 GTTM に基づくグルーピング構造獲得システム, 情報処理学会論文誌, Vol.48, No.1, pp284-299 (2007).