

自動作曲システムにおける音高決定モデルの獲得

Acquisition of Pitch Determination Model for Automatic Music Composition System

向 瑞稀

MIZUKI, Mukai

概要: 現在、個人の感性を反映した楽曲の自動生成に関する研究が進められている。先行研究では、ある程度の感性が反映できることが評価実験で示されているが、メロディが順次進行を基本として生成され、和音と協和しないアボイドノートが必要以上に含まれるという問題点がある。一般に、順次進行のみで構成される楽曲は、起伏がなく単調に感じられる。また、アボイドノートがメロディ上で長い音や目立つ音として使われると、違和感を覚えることが多い。ユーザが選択した既存楽曲の曲調や雰囲気を含む新しい楽曲の生成が目標であることを鑑みると、アボイドノートなどの出現割合は既存楽曲と類似していることが望ましい。本研究では、既存楽曲におけるメロディの曲調や雰囲気を楽曲に反映することを目的として、マルコフモデルを用いた音高決定モデルの獲得手法を提案する。

Summary: Automatic music composition system that composes music adapting to personal sensibility has been proposed. This research has shown that music can reflect sensibility in evaluation experiments. However, the problem is that melodies are generated on the basis of conjunct motion and contain more avoid notes than necessary, which do not harmonize with chords. Generally, a musical piece that consists of only conjunct motion feels monotonous without any ups and downs. In addition, when an avoid note is used as a long or prominent note in a melody, we often feel uncomfortable. In order to generate new music that reflects the tune of existing pieces of music selected by the user, the rate of occurrence of avoid notes should be similar to that of the selected pieces. This study aims to generate a melody reflecting the tune and mood of the melody in existing musical pieces, and proposes a Markov Model (MM) method for acquiring a pitch determination model in automatic music composition system.

キーワード: メロディ生成・自動作曲・マルコフモデル・Viterbi アルゴリズム

Keywords: Generating Melody, Automatic Composition System, Markov Model, Viterbi Algorithm

1. はじめに

現在、個人の感性を反映した楽曲の自動生成に関する研究[1]が進められている。はじめに、入力曲の共通の特徴から和音進行およびメロディの感性モデルを抽出する。次に、感性モデルや音楽理論を評価基準として、進化計算アルゴリズムを用いて和音進行およびメロディテンプレートを生成する。最後にメロディの音高を決定し、メロディと和音からなる楽曲を出力する。先行研究では、メロディの音高決定過程において2つの問題点が挙げられる。1つ目は、メロディが順次進行になりやすいことである。メロディを構成する各音の音高の候補は、同時に発音される和音の構成音および入力曲の登場音から決定される。音高候補のうち先行する音にもっとも近い音高が選択されるため、順次進行のメロディが構成されやすい。一般的に、順次進行のみで構成される楽曲は、起伏がなく単調になる。2つ目は、和音に協和しないアボイドノートが登場することである。アボイドノートがメロディ上で長い音や目立つ音として使われると、違和感を覚えることが多い。先行研究では、アボイドノートの有無については考慮していないため、他の音高候補と同確率で選択される。生成曲によっては、アボイドノート

が多く出現する可能性がある。2つの問題点が楽曲の印象を悪くしていると考えられる。

本研究では、個人の感性を反映した楽曲の自動生成において、メロディの単調さの回避と違和感の抑制を目的として、より感性に即した楽曲を生成するための音高決定モデルの獲得手法を提案する。

2. 自動作曲システム

自動作曲システムは以下の手順で楽曲を生成する。はじめに、和音進行およびメロディの感性モデルを抽出したのち、進化計算アルゴリズムを用いて、和音進行の感性モデルに適合する和音進行を生成する。感性モデルとは、学習楽曲のうちメロディで共通する音価列や音高の上下および和音進行で共通する音価列や和音進行を表す。次に、進化計算アルゴリズムによりメロディの感性モデルに適合するメロディテンプレートをモチーフごとに生成する。得られた和音進行を用いて、メロディテンプレートに含まれる各音の音高を和音の発音タイミングを加味し決定してメロディを完成させる。おわりに、メロディと和音進行を組み合わせて楽曲として出力する。ここでメロディテンプレートとは、メロディを構成する各音の音高以外の情報、すなわち各音の発音タイミングと音



図1 メロディの例

表1 メロディテンプレート表記の例

	メロディテンプレート表記			
	1/4 拍目	2/4 拍目	3/4 拍目	4/4 拍目
1 拍目	First	0	0	0
2 拍目	-1	0	0	0
3 拍目	3	0	3	0
4 拍目	3	0	0	0
5 拍目	0	0	0	0
6 拍目	-1	0	2	0
7 拍目	3	0	3	0
8 拍目	0	0	0	3

価, 先行音に対する音高の上下を表す. またモチーフとは2小節からなる楽曲を構成する最小単位を表す. メロディの例を図1, 対応するメロディテンプレートの例を表1に示す. メロディテンプレートはシステム内で-1から3の整数列で表現される. 各整数値は1/4拍分の音の状態を表し, それぞれ音を鳴らさないこと, 前の状態を継続すること, 先行音より低い音を鳴らすこと, 先行音と同じ音を鳴らすこと, 先行音より高い音を鳴らすことを意味する.

2.1. 音高決定方法

メロディテンプレートの各音高を以下の手順で決定する. メロディテンプレートの*i*番目の要素の整数値 m_i が2であるときは, 先行音の音高を*i*番目の要素の音高とする. m_i が1または3であるとき, メロディに使用可能な音高を要素とする音高候補集合に基づいて音高を決定する. *i*番目の要素が和音と同時に鳴る場合は, あらかじめ定められた音域のうち, 同時に鳴る和音の構成音と同じ音名の音高をすべて音高候補集合の要素とする. *i*番目の要素が和音と同時に鳴らない場合は, 定められた音域内における音階構成音の音高をすべて音高候補集合の要素とする. *i*番目の要素がメロディテンプレートにおける最初の要素の場合は, 同時に発音される和音の構成音からランダムに選択した音高を決定する. 最初の要素以外の各音の音高は, m_i が1であるとき, 音高候補集合の要素で先行音より低い音高のうちもっとも高い音高, m_i が3であるとき, 音高候補集合の要素で先行音より高い音高のうちもっとも低い音高とする.

2.2. 問題点

メロディを構成する各音の音高の候補は, 同時に発音されている和音の構成音および入力曲の登場音から決定される. 音高候補集合のうち先行する音にもっとも近い音高が選択されるため, 順次進行のメロディが構成されやすい. 一般的に, 順次進行のみで構成される音楽は, 起伏がなく単調になる. より多様性のある楽曲を生成するためには, 跳躍進行も含めメロディを構成することが

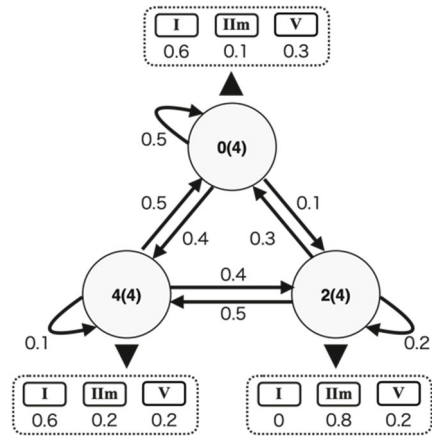


図2 マルコフモデルの例

望ましい. また, 音高候補にはアボイドノートが含まれる場合がある. ここでアボイドノートとは, 和音構成音の半音上の音のことである. 和音の構成音と半音下の関係にある音は, 和音の印象を崩さないことからアボイドノートとは定義されない. 和音発音時にメロディの構成音として, アボイドノートが使用された場合, 不協和な音として聞こえることが多い. 一般的に, アボイドノートを使用する際には, 経過音や刺繍音といった目立ちにくい音高として使われることが多いことから, 本システムでは考慮する必要がある.

3. マルコフモデルによる手法

既存楽曲の曲調をメロディに反映するためには, 各和音発音時の1ノートに関する状態を考慮する必要がある. メロディの各ノートは, 先行音に依存して決められることから, 各状態はマルコフ性を有しているといえる. メロディは音高を状態, 出力を和音としたマルコフモデルとして定義することができる. マルコフモデルとは, マルコフ性を持つ確率的な状態遷移と出力機能を備えた確率モデルである. 図2は状態と出力がともに3種類のマルコフモデルの例を示す. マルコフモデルを用いた楽曲の生成[2]では, 機械による作曲が人間による作曲と比べて優れていることが被験者実験で実証されている. しかし楽曲の感性の反映については特に言及されていない. 提案手法では, 音高情報を含む和音情報から, マルコフモデルを生成し, 自動作曲システムにより出力されたメロディテンプレートの各音高を決定する. マルコフモデルの各パラメータは, 自動作曲システムに入力した既存楽曲のメロディの音高系列と和音系列に基づいて決定する. 状態系列 s_t はメロディの音高列, 出力系列 x_t はメロディの各音高に対して同時に発音している和音の系列とする. メロディに含まれる休符は, 状態系列に含まない. 各音高, 和音はそれぞれ, 主音との音高差 (オクターブ値), デイグリー表記で表す. 自動作曲システムに入力した既存楽曲 $m_1 \sim m_N$ のメロディに*c*種類の音高 $\omega_1 \sim \omega_c$ が含まれるとき, *c*個の状態 $s_1 \sim s_c$ を持つマルコフモデルを構成する. 楽曲 m_o における音数を $L(m_o)$, 音高 ω_i から音高 ω_j への遷移回数を $A(m_o, \omega_i, \omega_j)$ とすると, 状態 s_i から

状態 s_j への状態遷移確率 a_{ij} は式(1)により求められる。

$$a_{ij} = \frac{\sum_o A(m_o, \omega_i, \omega_j)}{\sum_o (L(m_o) - 1)} \quad (1)$$

楽曲 m_o において音高 ω_i と和音 v_k が同時になった回数を $E(m_o, \omega_i, v_k)$ とすると、状態 s_i における和音 v_k の出力確率 $e_i(v_k)$ は、式(2)により求められる。

$$e_i(v_k) = \frac{\sum_o E(m_o, \omega_i, v_k)}{\sum_o \sum_l E(m_o, \omega_i, v_l)} \quad (2)$$

初期状態が音高 ω_i である確率を ρ_i としたとき、自動作曲システムにより生成された和音進行の初期和音を v_k としたとき、式(3)により求められる。

$$\rho_i = \frac{\sum_o E(m_o, \omega_i, v_k)}{\sum_o \sum_h E(m_o, \omega_h, v_k)} \quad (3)$$

メロディの音高を決定するに先立ち、自動作曲システムにより生成された和音進行とメロディテンプレートから、メロディの音価列と和音列を生成する。はじめに、メロディテンプレートに含まれる各音の音価のみを抽出し、音価列を生成する。このとき、整数値が-1、すなわち休符の音価は音価列には含まない。次に、和音進行から、メロディの音価列に含まれる各音と同時に和音を特定し、和音列を作成する。ただし、各和音はディグリー表記で表す。得られた和音列を観測系列として、構成したマルコフモデルから Viterbi アルゴリズムにより最尤状態系列を探索する。最尤状態系列の各状態を対応する音高に変換し、メロディテンプレートに含まれる各音の音高として割当て、メロディを完成させる。

4. 従来手法とのハイブリッド

第3章で説明した手法を用いて音高決定モデルを学習することに加えて、以下の処理を追加する。和音I7はIを拡張した和音だと考えられる。和音I7の構成音は主音との音高差で表すと{0,4,7,11}、和音Iは主音との音高差で表すと{0,4,7}であり、和音I7の構成音と和音Iの構成音の間には、 $I \subset I7$ の関係がある。和音Aの構成音は和音Bの構成音と $A \subset B$ の関係があるとき、楽曲 m_o において音高 ω_i と和音 v_k が同時になった回数 $E(m_o, \omega_i, B)'$ を式(4)により調整する。

$$E(m_o, \omega_i, B)' = E(m_o, \omega_i, B) + E(m_o, \omega_i, A) \quad (4)$$

学習に用いる楽曲データは、自動作曲システムに入力された楽曲と同一とする。本システムでは、3楽曲を必ず入力するように制限する。第3章で説明したマルコフモデルによる手法では、生成対象楽曲に既知でない和音が含まれる場合に、音高を決定することができない。本手法では、既知でない和音を代理のコードに置き換えることで音高決定に対応する。ここで生成対象楽曲の和音系列を $C = c_1, c_2, \dots, c_n$ 、学習データの和音シンボル列を $V = v_1, v_2, \dots, v_l$ とおく。入力曲の和音 c_k が既知でない

とき、以下の手順に従い、代理のコードに置き換える。和音 c_k の構成音と $A \subset c_k$ の関係にある和音Aが存在するとき、代理のコードを和音Aと置き換える。和音 c_k の構成音と $A \subset c_k$ の関係にある和音Aが存在しないとき、 c_k の根音と同じ根音を持つ和音がシンボル列に含まれる場合、該当和音を代理のコードとし、存在しない場合、和音シンボル列Vに含まれるダイアトニックコードをランダムに代理のコードとする。さらに置き換えた和音のインデックス k を代理和音インデックス列Iへ追加する。

4.1. Viterbi 経路の再推定処理

Viterbi アルゴリズムでは、遷移可能な音高が存在しないとき、 i 番目の音高探索が終了する問題がある。探索終了箇所から同音が続くことで、意図しない音高の遷移が発生する。意図しない音高の遷移を防ぐため、Viterbi 経路の再推定の処理を追加する。和音 x_t が観測されたとき音高 ω_i から遷移する ω_j が存在しないとき、和音系列 $X = x_1, \dots, x_n$ 、式(3)から初期確率 ρ を再度求め、Viterbi 推定をする。次に再推定で探索された状態を $S' = s'_1, \dots, s'_n$ と置いたとき、状態 $s_i = s'_i$ と置き換える。

4.2. 反復される同音進行と音高の2音間遷移の置換

同音進行の継続数または2音間の音高遷移数が多い場合、楽曲が単調になることが考えられる。過度な同音進行や音高の2音間遷移を防ぐため、反復される同音進行と音高の2音間遷移の置き換える処理を追加する。楽曲 m_o において音高 ω_i から音高 ω_j への遷移回数が $A(m_o, \omega_i, \omega_j)$ と表されるとき、同音進行の最大継続数は $\max_o A(m_o, \omega_i, \omega_i)$ とする。また楽曲 m_o で各状態状態が以下の式が成り立つとき、 i 番目の音高より2音間遷移の繰り返し回数を $M(m_o, \omega_i, \omega_j)$ と表し、2音間遷移の最大繰り返し回数は $\max_o M(m_o, \omega_i, \omega_j)$ となる。音高置き換え処理の対象は以下の条件に当てはまる音高とする。

1. 対象楽曲の同音進行の継続数が $\max_o A(m_o, \omega_i, \omega_i)$ より大きい。
2. 対象楽曲の2音間遷移の繰り返し回数が $\max_o M(m_o, \omega_i, \omega_j)$ より大きい。
3. k が代理和音インデックス列Iに含まれる。
4. s_k が対象楽曲の音域外である。

上記の条件が当てはまるとき、音高の置き換え処理の対象となり、対象の音高には2.1節で説明した方法を用いて、音高を決定する。

5. 評価実験

評価実験では、聴者に全10曲の未知楽曲を聴かせ、楽曲生成対象の感性を形容語対で表現された評価項目に関して、各曲の印象をSD法により7段階尺度で評価させる。評価には「明るい-暗い」、「好き-嫌い」、「嬉しい-悲しい」、「優しい-厳しい」、「穏やか-激しい」の計5対10語の形容語を用意した。ここで楽曲生成対象とする感性は、感性の評価軸上で正方向と負方向のいずれに位置するかが定められているものとする。「明るい」など正方向の形容語では値が高いほど、「暗い」など負方向の形容語

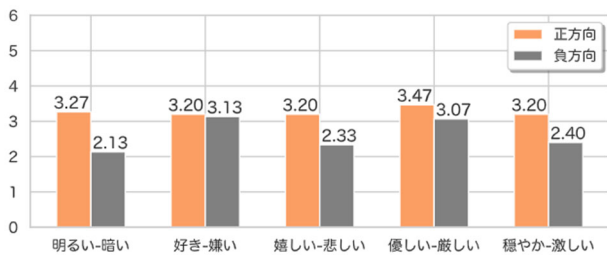


図3 提案手法の評価結果

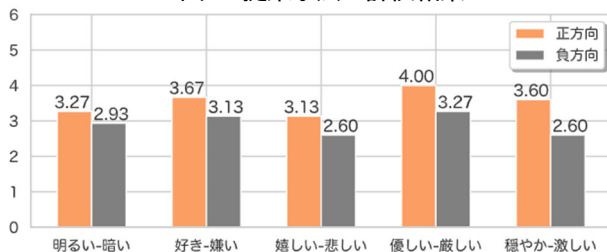


図4 従来手法の評価結果

では値が低いほど感性を反映しているとする。はじめに、自動作曲システムの学習データを選択するため WEB アンケート形式で予備実験を実施した。被験者は計 10 曲の未知楽曲を視聴したのち、5 対の形容語に関して評価する。得られた評価値を用い、各形容語について上位 3 曲を自動作曲システムの入力データとして感性モデルを獲得し、従来手法による音高決定過程、および本研究でのマルコフモデルおよび従来手法のハイブリッドによる音高決定手法を用いて、8 小節の楽曲を 20 曲生成した。本実験では、被験者は生成曲全 20 曲を視聴したのち、5 対の形容語に関して評価し、各楽曲に関しての視聴した感想を自由に記述する。また全楽曲を聞いた感想を自由に記述する。

6. 考察

従来手法および本研究における提案手法により生成された楽曲の評価結果をそれぞれ図 3、図 4 に示す。ここでの評価値平均とは、ある形容語に基づいて生成された曲の対象形容語に対応する評価項目の評価値の全被験者平均を意味する。評価結果からすべての形容語対の評価項目において正方向および負方向に差異があらわれた。両手法とも「嫌い」、「厳しい」に関して正方向に高い評価結果となり負方向の感性を反映できてはいないが、他の項目では、正方向の感性語では標準値の 3 を上回り、負方向の感性語では標準値の 3 を下回る結果となった。各評価項目において、正方向と負方向の評価に関する有意差の有無を t 検定により調査したところ、従来手法においては、「優しい-厳しい」に関して有意水準 5% で有意差がみられ、提案手法においては、「明るい-暗い」に関して有意水準 5% で有意差がみられたことから、両手法ともある程度感性を反映しているといえる。各評価項目について従来手法と提案手法には有意差はみられなかったが、「暗い」「悲しい」「激しい」の負方向の項目で提案手法は比較的良い評価値が得られている。また「好き」「優しい」「穏やか」の正方向の項目では、提案手法は従



図5 生成曲の例

来手法より悪い結果となっている。すなわち提案手法は負方向の感性を反映できるが、正方向の感性を反映できていないといえる。楽曲評価の自由記述より、「従来手法は提案手法と比べて、後半の極端な音高変化がない」「提案手法の音高が不安定な箇所がおもしろい」「強い不協和音が多い」という意見から、提案手法は楽曲上不安定な印象を与えるメロディが登場するため、正方向の評価を妨げていると考えられる。しかし両手法において「嫌い」とい評価項目に違いがないことから、メロディの不安定さが楽曲印象に影響しているとはいえない。

本実験の生成曲の例を図 5 に示す。(1) は提案手法によって生成されたメロディ、(2) は従来手法によって生成されたメロディ、(3) は和音進行を表す。1 小節目 3 番目の音高は、従来手法では音高 F# がベース曲に登場しなかったため音高 G が選択されているが、提案手法では先行音の E から経過音として正しく使われている。提案手法はベース曲の音高のみでなく、学習曲の音高を使用することでより多様性のあるメロディを生成できた。

7. おわりに

本研究では、個人の感性を反映した楽曲の自動生成において、メロディの単調さの回避と違和感の抑制を目的として、従来手法を拡張した確率的な手法、和音進行およびメロディの音高列から状態遷移を学習したマルコフモデルを用いた手法、従来手法とマルコフモデルのハイブリッド方式の手法の 3 手法について検討し、より感性に即した楽曲を生成するための音高決定モデルの獲得手法を提案した。評価実験の結果から「明るい-暗い」に関して感性を反映することができた。また従来手法では登場し得なかった経過音を表現することができ、より多様性のある楽曲を生成できた。しかし評価実験では従来手法より有用性を示すことができなかった。また生成曲に関して、「音高が安定しない」などの意見があり、不協和な音高が過度に出現する問題があるため改善の必要がある。自動作曲システムは、アプリ化するなど実用化に向けた取り組みが進められている。人間が作成した楽曲と遜色ない楽曲を生成することを目指す他、人間には生成しえない新たな楽曲の発見に繋がることを期待する。

参考文献

- [1] Otani, N., Okabe, D., Numao, M.: "Generating a Melody Based on Symbiotic Evolution for Musicians' Creative Activities", Proc. of GECCO' 2018, pp.197-204, 2018.
- [2] Schulze, W., Merwe, V. B.: Music Generation with Markov Models, IEEE Multi Media, Vol.18, No. 3, pp.78-85, 2011.