

音高と音価の隠れマルコフモデルを用いた自動副旋律生成

田村 理遊 但馬 康宏 小谷 善行
東京農工大学 工学府 情報コミュニケーション工学専攻

本研究は、楽曲の音価情報と音高情報を別々の HMM に学習し、副旋律を生成する手法を提案する。副旋律の生成は二段階で行われ、まず音価 HMM を用いてリズムを生成し、次に音高 HMM を用いてリズムに対し音高を当てはめ音符系列を生成する。副旋律生成実験と試聴評価実験を行った結果、事例楽曲の再現性を備え、一定の被験者からの支持を得られる副旋律の生成が実現できた。

Automatic Sub-Melody Generation Using Hidden Markov Models of Pitch and Duration

Riyu Tamura Yasuhiro Tajima Yoshiyuki Kotani
Tokyo University of Agriculture and Technology, Graduate School of Technology
Computer Information and Communication Sciences

We propose a method of sub-melody generation which consists of the two different kinds of HMM, one of them learns duration information and the other learns pitch information. The sub-melody generation procedure has the two steps. In the first step, we use the duration HMM to output rhythms. In the second step, we use the pitch HMM and the rhythms as input to produce notes. The experimental results and evaluation show that our approach can generate sub-melody the listeners prefer.

1. はじめに

本研究では、主旋律と和声で構成された楽曲に対し、新たに調和する副旋律を生成することを目的とする。

旋律の自動生成に関するアプローチには大きく分けて二種類ある。主に対位法などの音楽学上で定められた旋律の組み合わせに関する技法に沿って旋律を生成するルールベースの手法[1]と、事例楽曲から何らかの特徴を抽出し、その特徴を利用し生成モデルを構築する学習ベースの手法[2]である。前者は、人間の作曲技法を直接生成規則に取り入れることができるが、その反面、規則の定義や保守・更新に手間がかかったり、規則の作成者の主観に依存してしまう等の問題点もある。後者は事例楽曲から規則を

生成するため客観性に優れ、音楽理論に限定されない規則の抽出が期待できる。

本研究では、これらの利点に注目し学習ベースの手法を採用する。副旋律の音符の構成要素には音高と音価があり、この二要素それぞれに対し HMM を定義し学習を行う。そして、パラメータの決定した二種類の HMM を使い、まず副旋律の音価系列を求め、次にその音価系列に対し音高を当てはめることで、最終的な副旋律の音符系列を生成する。

2. 楽曲の定義

本稿で扱う楽曲は主旋律・副旋律・和声の三要素から構成されるものと定義する。和声が存在しない楽曲、副旋律が複数パート存在する楽曲、主旋律ま

たは副旋律が長時間の休符状態を含む楽曲は、楽曲の定義から外れるため研究対象には含めない。

本稿で扱う副旋律は、対位法の理論に則って生成される対旋律に限らず、和声進行の影響を強く受けた伴奏に近い旋律や、輪唱に近い形で追唱された旋律なども含める。すなわち、主旋律や和声と調和する広義の旋律と定義する。

主旋律・副旋律・和声は以下の系列で表す。

$$\text{主旋律} : M = m_1 m_2 \cdots m_l$$

$$\text{副旋律} : S = s_1 s_2 \cdots s_l$$

$$\text{和声} : H = h_1 h_2 \cdots h_l$$

$$m_i, s_i \in \{C0, C\#0, D0, \dots, G10, \text{rest}, \text{cont}\}$$

$$h_i \in \{C, Cm, C7, \dots, B6, Bm6\}$$

m, s の要素は音高 128 種類と休符(*rest*)と直前に演奏した音高の継続(*cont*)、 h の要素は和声 168 種類、 l は楽曲の演奏終了時間を表す。

$$M = F4, \text{cont}, G4, \text{cont}, A4, \text{cont}, \text{rest}, \text{cont}, C5, \dots$$

$$S = F3, \text{cont}, C3, \text{cont}, F3, \text{cont}, F3, G3, A3, \text{cont}, \dots$$

$$H = F, F, C, C, F, F, F, F, F, F, Dm, Dm, C, C, \dots$$

図 1 主旋律・副旋律・和声から構成される楽曲の例
(単位時間を八分音符とする)

3. 音価進行のモデル化

旋律の音価進行をモデル化する際は、音価情報にのみ注目し音高情報は除いて考える。音価進行の表現には、音符の開始時かそれ以外かの区別だけが必要となる。したがって、主旋律 M と副旋律 S を「*cont*」 「*cont* 以外」の 2 値の音価系列として扱えばよい。それをふまえ、 M, S を同音符継続数を用いた系列 MC, SC に変換する。同音符継続数は、「現在演奏されている音符 (休符) が過去に継続していた単位時間数」と定義する。 S と SC の対応関係は以下のとおり。

$$S = C, \text{cont}, D, E, \text{cont}, E, F, \text{cont}, \text{cont}, \dots$$

$$SC = 0, 1, 0, 0, 1, 0, 0, 1, 2, \dots$$

3.1 副旋律音価の遷移確率

副旋律が SC で進行する確率を以下に示す。

$$P(SC) = \prod_{i=1}^l P(sc_i | sc_1 \cdots sc_{i-1}) \quad (1)$$

$$sc_i = \begin{cases} sc_{i-1} + 1 & (s_i = \text{cont}) \\ 0 & (s_i \neq \text{cont}) \end{cases}$$

旋律は時間ごとに独立して生成されるのではなく、常に現在時間よりも過去の演奏の影響を受けながら生成される傾向が強い。しかし、時間経過に伴い過去の演奏時間が長くなるにつれ、スパースネスの問題が生じてしまう。そこで、条件項を $i-1$ だけに限定した近似を行う。

$$\prod_{i=1}^l P(sc_i | sc_1 \cdots sc_{i-1}) \approx \prod_{i=1}^l P(sc_i | sc_{i-1}) \quad (2)$$

(2) の bi-gram 確率は、一つ前の事象だけに依存するマルコフ性を持ち、副旋律の同音符継続数を状態とした HMM の遷移確率として用いることができる。

3.2 主旋律音価の観測確率

主旋律 M の音価進行の表現には、同音符継続数の系列 MC の他に、演奏中の音符 (休符) の音価の系列 MD を用いる。 M と MD の対応関係は以下のとおり。

$$M = C, \text{cont}, D, E, \text{cont}, E, E, F, \text{cont}, \text{cont}, \dots$$

$$MD = 2, 2, 1, 2, 2, 1, 1, 3, 3, 3, \dots$$

主旋律は全時間の系列が既知なので、 MD を用いることにより、主旋律の過去や現在の情報だけでなく、未来の情報も利用することができる。和声の変化は主旋律のリズムに包含されていると仮定すると、副旋律が SC で進行するとき MC, MD が観測される確率は以下のように近似できる。

$$P(MC, MD | SC) \approx \prod_{i=1}^l P(mc_i, md_i | sc_i) \quad (3)$$

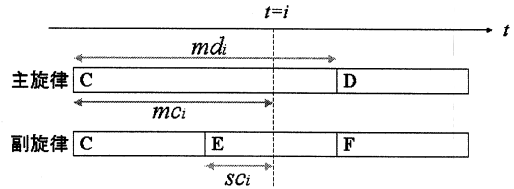


図 2 時間軸上における MD, MC, SC

3.3 音価情報を用いた HMM の枠組み

音価の確率モデルには、(2) と (3) にもとづいた HMM を用いる。HMM の状態は副旋律の同音符継続数とし、観測系列は主旋律の同音符継続数と演奏中の音符の音価とする。同音符継続数は単位時間につき 2 種類の値の変化しか起こらないので、HMM

の全ての状態は遷移先を2状態しか持たない。

通常HMMは、初期確率 π 、遷移確率 a 、観測確率 b から定義される。しかし、楽曲の最終部分は、終止形などの一定の秩序を持った進行になるのが一般的であり、その構造から外れてしまうと、全体的に不自然な楽曲になってしまう。そこでHMMのパラメータに、新たに終了確率 ω を定義する。

- q_i i 番目の状態
- π_i q_i の初期確率
- a_{ij} q_i から q_j への遷移確率 (4)
- $b_j(y)$ q_j における記号 y の観測確率
- ω_i q_i で終了する確率

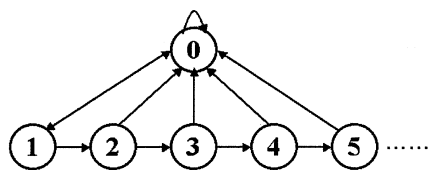


図3 同音符継続数を状態としたHMM

4. 音高進行のモデル化

旋律の音高進行をモデル化する際は、第4章とは逆に音高情報だけに注目し音価情報は除いて考える。副旋律 S では、*cont*は完全に前の音高に依存し、独立した音高情報を持たないので、*cont*が複数回連続しても一つの音高情報として扱うことにする。つまり、副旋律の音符単位の分解能を持った系列に変換して扱うことにする。この副旋律の系列を ST とする。また、主旋律と和声は、副旋律の音符の先頭時間に演奏されている音高の系列に変換し、これらを MT, HT とする。

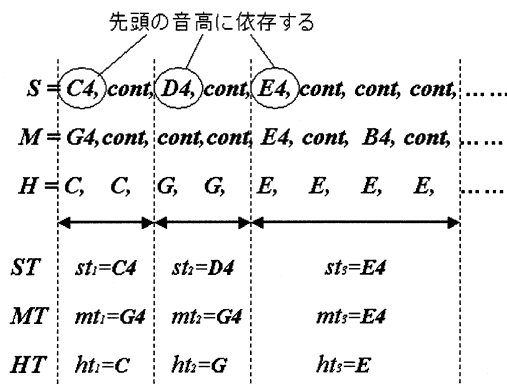


図4 ST, MT, HT の対応関係

旋律の音高は128音階を扱うが、この音階はオクターブ単位で $C \sim B$ までの音高がループを繰り返している二次元的な値として扱うことが可能である。これは音楽ならではの特別な性質であり、二次元的な値として扱う方が、一次元的な値として扱うより音高としてより多くの情報を持っていることになる。

- 主旋律の音高： $m'_i \in \{C \sim B, rest\}$
 - 副旋律の音高： $s'_i \in \{C \sim B, rest\}$
 - 副旋律のオクターブの高さ： $o_i \in \{0 \sim 10, rest\}$
- これらの記号を用いると ST は以下のように表せる。

$$ST = st_1 st_2 \dots st_l = s'_1 o_1 s'_2 o_2 \dots s'_l o_l \quad (5)$$

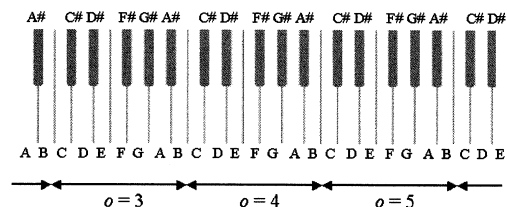


図5 音高のループとオクターブの高さ

4.1 副旋律音高の遷移確率

副旋律が ST で進行する確率を以下に示す。

$$P(ST) = \prod_{i=1}^l P(s'_i o_i | s'_{i-1} o_{i-1}) \quad (6)$$

(6)では(1)と同様にスパースネスの問題が生じるのでbi-gram確率に近似する。

$$\prod_{i=1}^l P(s'_i o_i | s'_{i-1} o_{i-1}) \approx \prod_{i=1}^l P(s'_i o_i | s'_{i-1} o_{i-1}) \quad (7)$$

しかし、(7)では音階を一次元的な値として用いた場合との違いがなく、前述した音楽特有の二次元的な性質を生かしていない。そこで(7)に対し、新たに二種類の考え方にもとづいた近似を行う。

一つ目は、直前の音符と現在の音符の音高差の生起確率に置き換え、なおかつオクターブの高さの差を同一視する近似方法である。

$$\prod_{i=1}^l P(s'_i o_i | s'_{i-1} o_{i-1}) \approx \prod_{i=1}^l P(s'_i o_i - s'_{i-1} o_{i-1} | s'_{i-1}) \quad (8)$$

(8)は音高差の生起確率を求めているが、 $i-1$ における音高とオクターブの高さが定まっていれば、128音

階の音高に置き換えることが可能である。この近似は、 $i-1$ において音高が同じであれば、オクターブの高さに関わらず等しい **bi-gram** 確率を持つ。

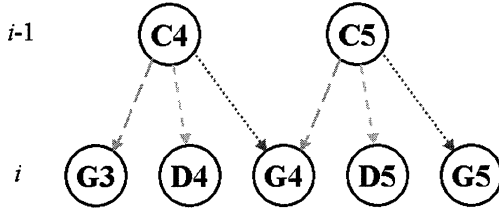


図6 音高差に置き換えた近似の例
(同種の破線の遷移確率は等しい)

なお、休符状態(*rest*)には音高自体が存在しないため、音高差を取ることが出来ない。そこで特例として、休符状態への遷移確率は直接求め、休符からの遷移確率は休符の一つ前の状態からの遷移確率と同一とみなす。

$$P(\text{rest} - s'_{i-1}o_{i-1} | s'_{i-1}) \rightarrow P(\text{rest} | s'_{i-1})$$

$$P(s'_i o_i - \text{rest} | \text{rest}) \rightarrow P(s'_i o_i - s'_{i-2} o_{i-2} | s'_{i-2})$$

(8)の近似の長所は、全ての音域の旋律進行情報を共有しているのので、限られた楽曲データから効率的に情報が得られると同時に、スパースネスの低減をしている点にある。本来、音高が同一でもオクターブの高さが異なれば、当然その音自体は別物として扱われる。しかし、旋律進行の特性はオクターブの高さによって大きく変わることは無い。たとえば、 $C3$ から $D3$ への遷移が許可される場合に $C4$ から $D4$ への遷移が許可されない可能性は低い。(8)はこのような考えにもとづいた近似である。

しかし、オクターブの高さが省略されたことにより、現在の音高がどの辺りの音域に属しているのかの情報が失われてしまうと、旋律が時間経過にともない高音域や低音域に発散する等、実際の楽曲ではありえない進行になってしまう可能性がある。オクターブの高さ o は、旋律進行を一定の音域内に安定させるために必要な情報であり、(8)の近似では何らかの形でその情報を補完する必要がある。

そこで、(7)に対する二つ目の近似を用いる。

$$\prod_{i=1}^l P(s'_i o_i | s'_{i-1} o_{i-1}) \approx \prod_{i=1}^l P(s'_i o_i) \quad (9)$$

(9)は、直前の事象を完全に無視しており、旋律進行に関する情報はほぼ皆無である。旋律の音符は単位時間ごとに独立して生起しているわけではなく、音符同士の前後関係に依存しているので、(9)単体では ST の進行を表現するには不適切である。だが、過去の

事象に捕らわれることなく、出現頻度の高い音高へ副旋律を収束させる効果が期待できるため、(8)の問題点を補完するには有用と思われる。

そこで、(8)と(9)の積に重み付けしたものを、最終的な(7)の近似とする。

$$\prod_{i=1}^l P(s'_i o_i | s'_{i-1} o_{i-1})$$

$$\approx \prod_{i=1}^l P(s'_i o_i - s'_{i-1} o_{i-1} | s'_{i-1})^\alpha P(s'_i o_i)^{1-\alpha} \quad (10)$$

重み: α ($0 \leq \alpha \leq 1$)

4.2 主旋律と和声の観測確率

副旋律が ST で進行するとき、主旋律 MT と和声 HT の組が観測される確率は以下のように近似する。

$$P(MT, HT | ST)$$

$$= P(mt_1 \dots mt_l, ht_1 \dots ht_l | st_1 \dots st_l)$$

$$\approx \prod_{i=1}^l P(m'_i, ht_i | s'_i) \quad (11)$$

主旋律と副旋律の関連性は、オクターブ単位の差よりも音高どうしの調和に起因するところが大きいという仮定する。(11)では主旋律・副旋律のオクターブの高さは省略し、音高だけに絞った観測確率として扱う。

4.3 音高情報を用いた HMM の枠組み

音高の確率モデルは、(10)と(11)にもとづいた HMM を用いる。HMM の状態は副旋律の音符の音高とし、観測系列は主旋律の音高と和声の組とする。HMM のパラメータは、終了確率 ω を定義した(4)に準ずる。

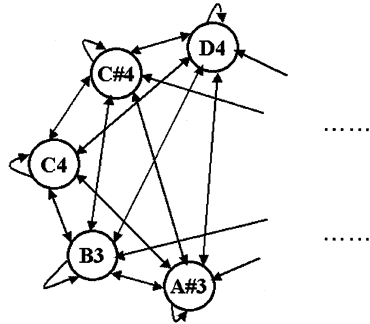


図7 副旋律の音高を状態とした HMM

5. 副旋律生成アルゴリズム

作曲者によって作曲手順には個人差があるが、まず主旋律と和声を作成し、その後それらに調和する副旋律を付加するのが一般的である。その手順にもとづき、本研究では主旋律と和声によって副旋律が

定まると仮定する。与えられた主旋律と和声に対しもっとも出現する尤度の高い副旋律は以下のとおり。

$$\arg \max_S P(S|M, H) \quad (12)$$

Bayes の定理より(2)の等式が成り立つ。

$$P(S|M, H) = \frac{P(S)P(M, H|S)}{P(M, H)} \quad (13)$$

$P(M, H)$ は S に関係無く値が定まるため無視する。したがって、(12)を求めるには $P(S)$ と $P(M, H|S)$ の積を最大化する S を求めればよい。

本研究ではこの考えにもとづき副旋律生成を行う。まず二つの HMM の各パラメータを事例楽曲から学習する。そして、音価情報を用いた HMM によって SC の最尤系列を探索する。SC からは S を「cont」「cont 以外」の 2 値で表現した音価系列が得られる。さらに音高情報を用いた HMM によって「cont 以外」の音高 ST の最尤系列を探索し、副旋律の音符系列を生成する。探索アルゴリズムには Viterbi アルゴリズムを用いる。生成例を以下に示す。

<Step.1 最尤 SC を探索>

0, 0, 0, 1, 2, 0, 0, 0, 1, 0

<Step.2 2 値の音価系列に変換>

?, ?, ?, cont, cont, ?, ?, ?, ?, cont, ?

<Step3. 最尤 ST を探索>

$E4, D4, C4, C4, E4, G4, G4, F4$

<Step4. 音価系列に ST の音高を付加>

$E4, D4, C4, cont, cont, C4, E4, G4, G4, cont, F4$

6. 自動副旋律生成実験

実験用の楽曲は、童謡・民謡・歌謡曲から 46 曲用意した。その中から、テストデータとなる楽曲を 1 曲選択し、残りの 45 曲を事例楽曲として用いた。これを、選択する生成対象楽曲を変更しながら全楽曲に対して行った。また、調の主音が C になるように、楽曲全体の音高をシフトした。単位時間は十六分音符の音価とし、(10)の重み値は $\alpha = 0.8$ とした。

実験結果の例を譜面形式で図 8 に示す。譜面の一段目が原曲の主旋律と和声、二段目(副旋律 A)が生成結果の副旋律、三段目(副旋律 B)が原曲の副旋律である。全体として生成結果は、とくに目立った不協和音は存在せず、楽曲として聴けるレベルは実現できた。リズムに関しては、ある程度主旋律に合わせて、四分音符前後の音価を持った音符を生成している。音高に関しては、不自然な発散等は見られず、安定した音域での進行となっている。

The figure displays two musical scores side-by-side. The left score is for the song 'Utsunomiya' (「鳩」) and the right score is for 'My Home in Utsunomiya' (「峠の我が家」). Each score consists of three staves: the top staff is the original melody and harmony, the middle staff is the generated counter-melody A, and the bottom staff is the original counter-melody B. The generated counter-melody A shows a sequence of notes that closely follows the rhythm and pitch contour of the original counter-melody B, demonstrating the effectiveness of the generation process.

図 8 副旋律生成結果 (左は「鳩」、右は「峠の我が家」の冒頭部)

原曲の副旋律との近さを測る指標として、副旋律の絶対再現率と音高再現率、そして和音構成音の比率を調べた。絶対再現率は生成結果と原曲で副旋律の音高とオクターブの高さの両方が一致した割合、音高再現率は少なくとも音高は一致した割合を意味する。和音構成音は、根音・和声音（根音は含まず）・非和声音・休符について調べた。

表1 生成結果の再現率

絶対再現率	音高再現率
18.1%	35.0%

表2 副旋律の和音構成音の比率

	根音	和声音	非和声音	休符
生成結果	18.4%	35.0%	40.0%	6.6%
原曲	18.2%	34.1%	42.2%	5.5%

もっとも出現確率の高い根音だけを出力した場合、音高再現率は18.2%なのに対し、生成結果は絶対再現率18.1%、音高再現率35.0%に達しており、ある程度再現性を持ったモデルが学習されていることが分かる。また、和音構成音の比率が概ね原曲に近い値になっている点からも同様のことが分かる。

7. 試聴評価実験

試聴評価実験は、被験者が楽曲を試聴しながら自分の感じた印象をアンケートで回答する形式で行った。評価対象は、生成結果の副旋律、BL(ベースライン)用の副旋律、原曲の副旋律の三種類。これらの副旋律に共通の主旋律・和声を付けた楽曲を試聴させた。BL用の副旋律には、主旋律の音高を三度下げただけの、容易に生成可能でそれなりに調和が期待できるものを用いた。アンケートでは三種類の楽曲に対し、「気に入った曲の順位」と「音楽的に自然と思う曲の順位」を回答させ、順位の根拠や感想等の定性的な評価も文章で記入させた。

このアンケートに対する有効回答は38件得られた。本研究では副旋律の優劣は次元の尺度上で決定できると仮定する。したがって、回答の順序データは対比較データに変換できる。この対比較データを、サーストンの比較判断の法則[3]にもとづいて間隔尺度で表した。

表3 被験者が気に入った評価の相対距離

	生成結果	BL	原曲	平均値
生成結果	0.00	0.80	-0.41	0.130
BL	-0.80	0.00	-2.22	-1.007
原曲	0.41	2.22	0.00	0.877

表4 被験者が自然に感じた評価の相対距離

	生成結果	BL	原曲	平均値
生成結果	0.00	1.00	-0.41	0.243
BL	-1.00	0.00	-1.12	-0.707
原曲	0.27	1.12	0.00	0.463

気に入った曲・自然に感じた曲の評価結果は、原曲、生成結果、BLの順に良かった。しかし、自然に感じた曲の評価は気に入った曲の評価に比べ、原曲と生成結果の相対距離が近い結果となった。この結果は文章回答にも見られ、原曲に対する感想には「上手く調和している」「好き」「凝っている」という回答が目立ったのに対し、生成結果に対する感想には「単調」「平坦」という回答がある一方、「自然である」「和音が正しい」などの回答も目立った。以上の結果から、本研究で用いた副旋律生成手法は、人間の作成した事例楽曲から、音楽的に自然と感じられる要素を抽出し、主旋律や和声との調和の取れた自動副旋律生成を実現することができたと考えられる。

8. おわりに

本研究では、楽曲の音価情報と音高情報を別々のHMMに学習し、主旋律・和声から構成される楽曲に対し新たな副旋律を生成する手法を提案した。そして副旋律生成実験を行い、結果から得られた統計量や試聴アンケートを用いて定量的かつ定性的な評価を行った。

本研究で提案した手法にもとづいて生成した副旋律は、複数の被験者から支持されており、音楽としての自然さを感じられるだけの能力は備えていると言える。とくに、主旋律や和声と調和の取れた進行が実現できた。また、その副旋律の音高構成等を分析した結果、音高再現率が35.0%に達し、原曲との和音構成音比率の差も1~2%程度に収まっており、事例楽曲に対する再現性を備えた学習がされていることも確認された。

参考文献

- [1] 中瀧昌平, 西本卓也, 嵯峨山茂樹: 動的計画法と音列出現確率を用いた対位法の対旋律の自動生成, 情報処理学会研究報告, 2004-MUS-56.(2004)
- [2] 紙谷元喜, 木村英志, 阿江忠: GPを用いたメロディー作成の一手法, 情報処理学会研究報告 2002-MUS-45.(2002)
- [3] 難波精一郎, 桑野園子: 音の評価のための心理学的測定法, コロナ社, 1998.