Vol.2011-MUS-91 No.3
2011/7/27

階層 Pitman-Yor 言語モデルを用いたメロディー生成手法の提案

白井　亨[†1]　谷口忠大[†1]

メロディーを生成する手法は数多く提案されているが，既存楽曲の特徴を持つメロディーを生成する事はできていない．そこで，本研究では階層 Pitman-Yor 言語モデルを拡張した可変長 n-gram 言語モデルを用い，学習データの特徴を持つ新たなメロディーを生成する手法を提案する．また，実験を行い，その有効性を検証する．

A proposal of the melody generation method using hierarchical pitman-yor language model

Akira Shirai[†1] and Tadahiro Taniguchi[†1]

Although a lot of melody generation method has been proposed, it is difficult to generate a melody that has the features of existing music. Thus, we propose a method using Variable-order Pitman-Yor Language Model, extension of Hierarchical Pitman-Yor Language Model, which generates new melodies that have the features of the data. We also evaluate this method by some experiments.

1. はじめに

近年，パソコン上で作曲を行う DTM が広く普及したことや作品を発信する場が整っていることなどから新たに作曲活動を始めることが容易になった．しかし，依然として作曲を行う人は少なく日常的に曲を書く人はプロの作曲家など一部に限られている．これは楽曲の構造的な複雑さや作曲行為自体が非常に個人的なものであることが原因だと考えられる．まず，作曲を行うためには様々な音楽的知識が求められる．それらの知識を持っていない音楽初心者は楽曲の複雑な構造を理解する事ができないため楽曲を完成させる事ができない．また，音楽的な活動を行っていない人にとって一般的には個人作業である作曲の現場に触れることは困難である．結果的に間違った学習やノウハウの抱え込みを行ってしまう可能性を含んでいる．これらの問題点から作曲を補助するツール及び作曲行為を共有するソーシャルな場が必要である．そこで本研究では音楽初心者でもイメージ通りの作曲が可能となる手法の提案及びこれを用いたソーシャルメディアの設計を行う．

2. 研究目的

本稿では音楽初心者でもイメージ通りの作曲が可能となる手法を提案する．

音楽的な知識を持たない人でも手軽に作曲を行うことのできるポピュラーな方法として自動作曲がある．自動作曲システムはフリーソフト[⋆1]としても多く広まっておりユーザが楽曲イメージを表現するキーワード（ジャズ風，明るいなど）を選ぶことでそのイメージに合う楽曲を生成することが可能である．しかし，一般的な自動作曲システムは楽曲イメージがあらかじめ用意されたものに限られることや，結果に対して修正することができないことなど自由度が低いという問題がある．また，楽曲イメージとキーワードの対応は開発者が決めたものであり，この対応が必ずしもユーザと一致するとは限らない．そのため，この楽曲イメージとキーワードの対応にズレがあった場合，ユーザはイメージしている楽曲を作曲することができない．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作曲をすることで楽曲イメージを反映させる研究が行われている[2)]．この方法ではユーザの評価を基に遺伝子で表現された楽曲を進化させていくことで感性を柔軟に取り入れている．しかし，このような計算法を用いた作曲手法では解として数十秒から数分ある楽曲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．

この他にも様々な自動作曲手法が提案されている．深山らのオルフェウスは日本語歌詞の韻律などを制約として動的計画法を用いて歌唱曲の生成を行っている[3)]．この手法ではメロディーの生成を各種制約による重みがかかった尤度最大経路の探索問題として扱っている．

†1 立命館大学
Ritsumeikan University

⋆1 音楽研究所: 自動作曲システム ACS,
http://hp.vector.co.jp/authors/VA014815/music
鵜飼利彦: Juice and Candy 3.33,
http://www.vector.co.jp/soft/win95/art

オルフェウスは入力された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．システムは Web 上で公開されており作詞した歌詞を用いて手軽に歌唱曲を生成することが可能であるが，一方で最適化手法を用いていることで多様性を失っているという問題もある．

白井らは統計的言語モデルとしてよく用いられる n-gram モデルから確率的にメロディーを生成する手法を提案している[1]．この手法では n-gram の n を 2 として既存楽曲からモデルの学習を行い，コード進行と楽曲構造による制約を基にギブスサンプリングを用いて確率的にメロディーの生成を行うことで多様なメロディーを生成，その後に修正することが可能である．また，ユーザの楽曲イメージを反映させる方法として既存楽曲に似ているメロディーの生成を行っている．しかし，n-gram 長が 2 と短く学習データから特徴抽出する機能に乏しいため，楽曲構造による制約を用いて既存楽曲に似ているメロディーの生成を試みているが，良い評価は得られていない．そこで，より長い音符列を学習するために n-gram 長を長くするという方法が考えられるが，n が増えると状態数が指数的に爆発することや学習データがスパースになるという問題がある．また学習データとして与えるコーパス，さらにはメロディーによって必要な n グラム長は異なると考えられ，これを固定することは不自然である．よって，学習データから効率よく特徴を抽出するためには音符列の文脈によって適切な n グラム長を推定しながら，スパースなデータのスムージングを行うモデルが必要である．

そこで本研究では Variable-order Pitman-Yor Language Model(VPYLM) を用いてメロディーのモデル化を行う．VPYLM はノンパラメトリックな確率過程により高精度なスムージング及びデータの持つ潜在的な n-gram 長の推定が可能である．また本研究では自然なメロディーを生成するために学習データからの特徴のみではなくコード進行を制約として与え，歌唱曲を生成するために歌詞を入力した上でサンプリングする手法を提案する．

3. 提 案 手 法

本手法は既存楽曲を蓄えた楽曲コーパスを学習データとして VPYLM の学習を行い，メロディーを生成する．また，ユーザは歌詞の入力とコード進行の設定が可能であり，これらに基づいて生成が行われる．提案手法の概要を図 1 に示す．

3.1 楽曲コーパスの作成

既存楽曲を学習データとして用いるために楽曲コーパスを作成する．メロディーを音符の列 $\mathbf{s} = (s_1, s_2, ..., s_N)$ として N 次元のベクトルで表す．ベクトルの各要素はメロディーを離散的に分割したものであり本研究では 8 分音符単位で分割を行う．よって，ベクトルの各

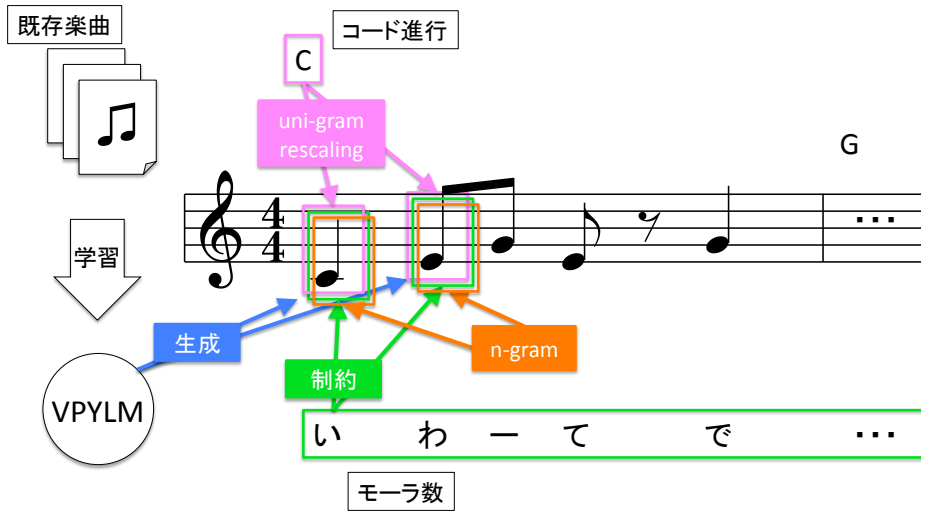

図 1 提案手法の概要図
Fig. 1 The overview of our method.

要素には音符のピッチ，音が伸びている状態，休符のいずれかが入ることになる．また，同様にしてコード進行を $\mathbf{c} = (c_1, c_2, ...c_N)$ で表す．こちらの各要素にはコードネームが入る．なお，調が C でない楽曲については C 調に移調を行う．

3.2 VPYLM の学習

n-gram モデルでは n-gram 長が短いと，学習データの中で特徴的であるにも関わらず学習できないデータ（例えば頻出する長いフレーズなど）ができてしまう．よって，そこから生成されたメロディーは特徴が失われてしまっていると考えられる．n を増やすことでこのようなデータは学習可能である．しかし，逆に n-gram 長が長くなると，取り得る状態数が指数的に爆発してしまうためスパースなモデルになってしまうという問題がある．よって学習データを適切にスムージングする必要がある．

そこで本手法では Hierarchical Pitman-Yor Language Model(HPYLM)[4] を拡張した VPYLM[6] を用いてメロディーの学習を行う．HPYLM とは Teh らによって提案された n-gram 分布の階層的な生成モデルである（図 2）．HPYLM では Hierarchical Pitman-Yor

Vol.2011-MUS-91 No.3
2011/7/27

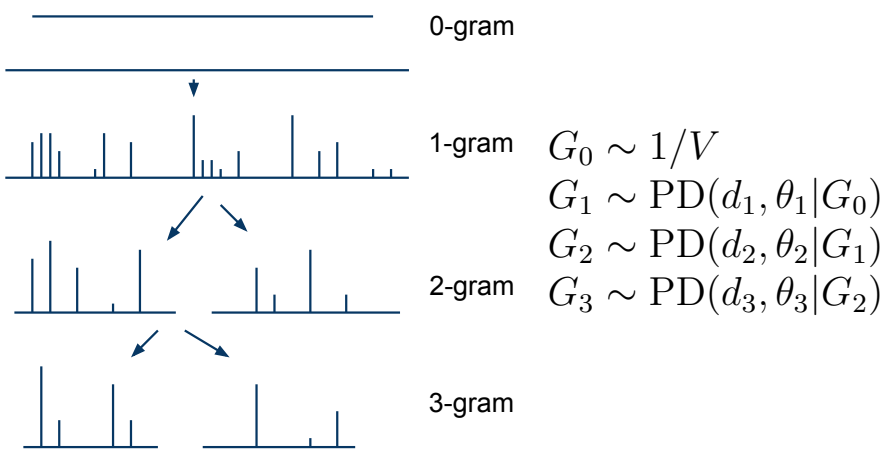

図 2 階層的な n-gram 分布の生成
Fig. 2 The hierarchical distribution of n-gram.

Process と呼ばれるノンパラメトリックな確率過程によって，適切にスムージングされた n-gram 分布を階層的に生成及び推定することが可能である．また，現在数あるスムージング手法の中でも最高性能と言われる Kneser-Ney スムージング[5] はこの確率過程の近似となっている．HPYLM では学習を行う際に確率的に接尾辞木の一つ上の文脈をカウントすることでスムージングを行う．この HPYLM を可変長 n-gram 言語モデルに拡張したものが VPYLM である．VPYLM は n-gram 長を固定せず，学習データから単語の持つ潜在的な n-gram 長を推定する．この拡張によって長い n-gram 長が必要なデータとそうでないデータを推定する事ができ，n を固定した場合よりも効率の良い学習が可能である．

3.3 メロディーの生成

本手法ではユーザの入力した歌詞とコード進行を基にメロディーの生成を行う．具体的には歌詞の文字数による制約とコード進行による制約をかけながらサンプリングを行う．

3.3.1 Gibbs sampler によるメロディーの生成

ある文脈 $\mathbf{h}$（ここでは音符の並び）の後に音符 s が出現する確率は，n-gram 長 n を隠れ変数とみなして

$$p(s|\mathbf{h}) = \sum_n p(s, n|\mathbf{h}) \tag{1}$$
$$= \sum_n p(s|n, \mathbf{h}) p(n|\mathbf{h})$$

のように予測を行う．ここで，第一項はオーダーを n とした HPYLM の予測確率，第二項は文脈 $\mathbf{h}$ の持つ n-gram 文脈長分布である．式 (2) より，メロディー全体の事後確率（メロディーの生成確率）は

$$p(\mathbf{s}) = \sum_i^N p(s_i|\mathbf{h}) \tag{2}$$

のように求めることができる．ギブスサンプリングするためには生成確率の比が求まればよいので式 (2) から音符 s_i 以外の音符 $\mathbf{s}_{-i}$ が決まっているとに s_i が音符 k になる確率は

$$p(s_i = k|\mathbf{s}_{-i}) \propto p(s_i = k|\mathbf{h}) p(s_{i+1}|\mathbf{h}') \times ... \times p(s_{i+n_{\max}}|\mathbf{h}^{(n_{\max})}) \tag{3}$$

で求めることができる．ここで $n_{\max}$ は学習する際に決めた VPYLM の最大 n-gram 長である．式 (3) を用いることでメロディーの Gibbs Sampler を構成することができるが，本手法ではさらに歌詞とコード進行によって重み付けを行う．

3.3.2 歌詞の割り当て

歌詞の総モーラ数[⋆1]を M，メロディーに含まれる伸ばしている状態，休符以外の総音符数を N_{note} とする．ここではモーラ数は音符数を上回ることはないとし，共に同じ数存在する状態を理想型と考える[⋆2]．よって M が N_{note} より大きくなると急激に確率が下がり，同じ時に最大値をとり，M が小さくなると徐々に確率が下がるような制約を導入する．本手法ではこのような制約を以下の式で表す．

$$p(M, N_{\mathrm{note}}) = \begin{cases} \exp((N_{\mathrm{note}} - M)\alpha) & \text{if } N_{\mathrm{note}} < M \\ \exp(\frac{M - N_{\mathrm{note}}}{\beta}) & \text{otherwise} \end{cases} \tag{4}$$

ここで α 及び β は指数関数の減衰率を調整するパラメータである． 3 にこの制約の振る舞いを示す．式 (3) に式 (4) を適用すると

⋆1 一定の時間的長さをもった音の文節単位．音節とは異なり長音，促音，撥音もモーラ数に含まれる．
⋆2 モーラ数が音符数より小さい場合，モーラを発音している間にピッチが変わることを許す．

Vol.2011-MUS-91 No.3
2011/7/27

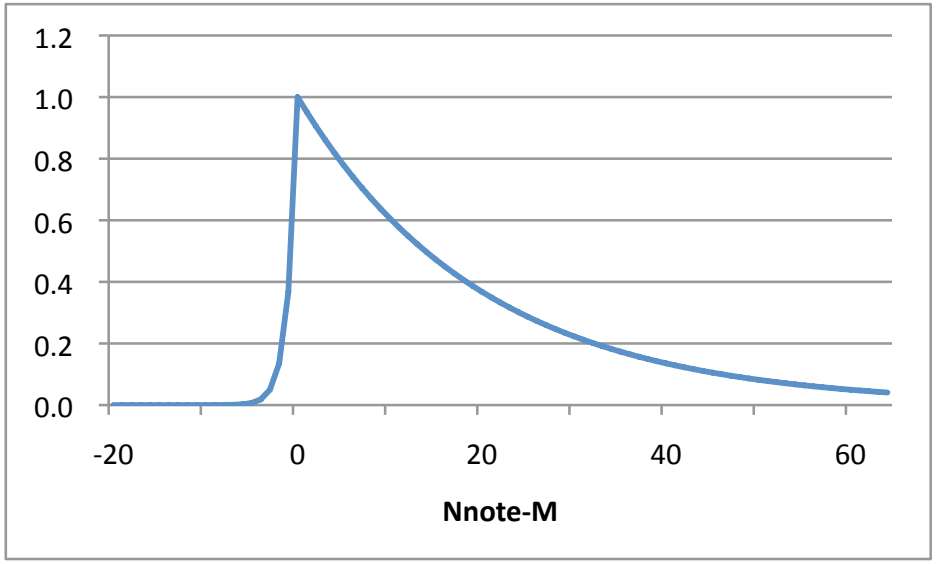

図 3 モーラ数制約（$\alpha = 1, \beta = 20$）
Fig. 3 The restriction of mora ($\alpha = 1, \beta = 20$).

$$\hat{p}(s_i = k|\mathbf{s}_{-i}) = p(s_i = k|\mathbf{s}_{-i})p(M, N_{\text{note}}) \tag{5}$$

となり，歌詞を考慮したメロディーの Gibbs Sampler を構成することができる．

3.3.3 uni-gram rescaling を用いたコード進行のトピック適応

コード進行による制約をトピック適応の問題と捉え uni-gram rescaling を用いて適応を行う．コードを適応すると

$$p(s_i|s_{i-1}^{i-n}, c_i) \propto \frac{p(s_i|s_{i-1}^{i-n})p(s_i|c_i)}{p(s_i)} \tag{6}$$

のように表すことができる．ここで $p(s_i|c_i)$ はコード c_i 上で音符 s_i が出現する確率であり，楽曲コーパスから学習する．式 (5) に式 (6) を適用すると

$$\begin{aligned} p(s_i|\mathbf{s}_{-i}, \mathbf{c}) \propto\ & p(s_i = k|\mathbf{h}, c_i)p(s_{i+1}|\mathbf{h}', c_{i+1}) \\ & \times \ldots \times p(s_{i+n_{\max}}|\mathbf{h}^{(n_{\max})}, c_{i+n_{\max}}) \times p(M, N_{\text{note}}) \end{aligned} \tag{7}$$

となる．ここで $p(s_i|\mathbf{h}, c_i)$ は

$$p(s_i|\mathbf{h}, c_i) = \frac{\sum_n p(s_i|n, \mathbf{h})p(n|\mathbf{h})p(s_i|c_i)}{p(s_i)} \tag{8}$$

と表させる．式 (8) を用いて Gibbs Sampler を構成することで歌詞及びコード進行を考慮したメロディーを生成する．

4. 実　　験

提案手法の妥当性を検証するために 20〜30 代の男性被験者 5 名に対して評価実験を行った．

4.1 実験条件

以下の条件で実験を行った．

学習データ 楽曲コーパスとして J-Pop アーティスト 4 組の楽曲から A，B メロ部分を 35 曲使用した．コーパス量が少ないため対策として同じデータを 10 個分学習させた．総単語数は 37760，総語彙数は 30 となった．

学習設定 1000 回のギブスサンプリングを行い，n-gram の最大長 $n_{\max}$ を 10 とした．

生成設定 VPYLM から 100 回の burn-in の後 200 回のギブスサンプリングを行い最大生成確率のメロディーを生成した．メロディーの長さ N を 32（4 小節）とした．また歌詞の制約パラメータを $\alpha = 10, \beta = 50$，コード進行を (C,G,Am,G) とした．歌詞は「岩手で出会った君と恋した」とした[⋆1]．

4.2 実験結果

図 4 は学習した VPYLM の n-gram 長分布である．5, 6-gram 文脈になる単語が多く，比較的長い系列を扱えることがわかる．

図 5 は VPYLM からメロディーを生成する際に生成確率がどのように変化するかを表している．縦軸は生成確率の対数をとったもの，横軸はサンプリング回数を表している．最適化ではないため値は常に変動するが徐々に事後確率の高いサンプルが生成されていることがわかる．

図 6 は生成されたメロディーと，そこに含まれているコーパス内の楽曲のフレーズを表している．図のように生成メロディーは複数楽曲の部分的なフレーズを含んでいる事が分かる．

⋆1 歌詞と伴奏を加えた完成版の楽曲は YouTube にアップロードしている．
http://www.youtube.com/watch?v=n8zYgDwulZg

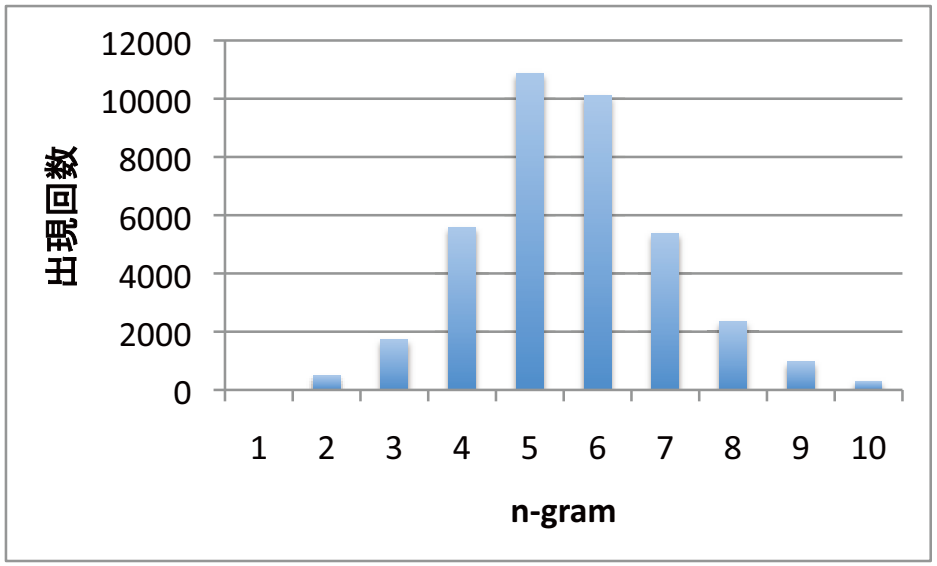

図 4 *n*-gram 長の分布
Fig. 4 The distribution of the *n*-gram order.

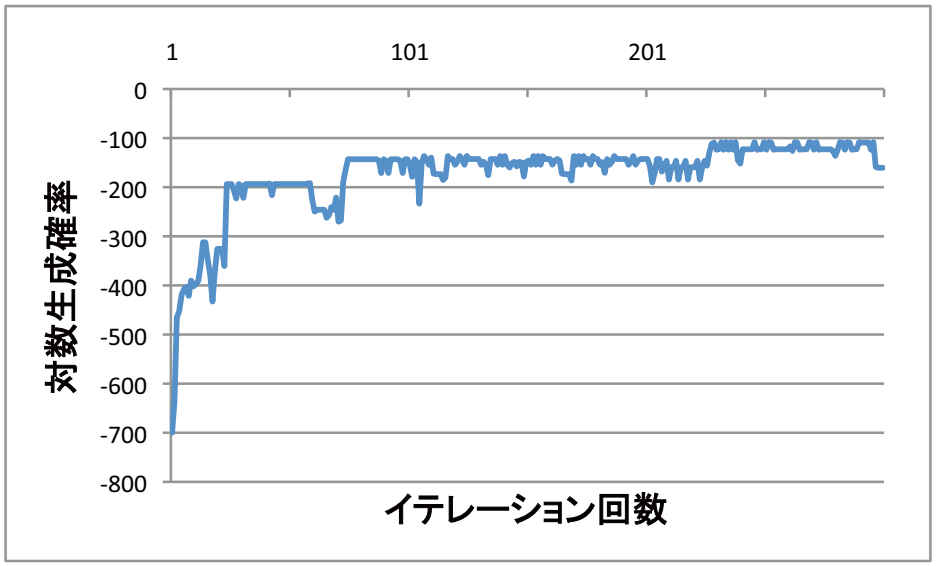

図 5 生成確率の推移
Fig. 5 The transition of a generation probability.

4.3 考　　察

生成結果より，生成モデルとして VPYLM を用いることで楽曲コーパス内の長いフレーズや短いフレーズが入り交じったメロディーを生成することが可能となっている．これらのフレーズはコーパス内の特徴的なフレーズだと考えられるが，これについては評価実験を行いコーパスの特徴を反映できているか調査する必要がある．また，今回は同じコーパスを複製して学習させたため比較的長いフレーズが多く現れている．特に最後のフレーズは殆どがコーパス内の楽曲と同じフレーズになってしまっている．しかし，原曲が表拍から入っているのに対して生成メロディーは裏拍から入っているため違った印象を与える可能性がある．

5. ま と め

本稿では HPYLM を可変長 *n*-gram モデルに拡張した VPYLM を用いたメロディー生成手法を提案し，有効性について検証を行った．実験結果より学習した VPYLM よりサンプリングを行うことで事後確率の高いメロディーを得ることが可能であることが実証された．また，HPYLM による *n*-gram 分布の階層的なスムージングにより従来手法で存在した学習データがスパースになる問題を緩和したと言える．さらに，生成されたメロディーはコーパス内の複数楽曲の部分的なフレーズを含んでいる事が分かった．しかし，特徴的なフレーズを含むメロディーがコーパス内の楽曲と似ているかどうかは分からない．これについては評価実験を行い検証する必要がある．また，本手法では歌詞の割り当て位置の推定は行っていないため，歌詞のリズムがおかしい箇所が多く見られた．今回のモデルではリズムパターンなどは考慮していないため今後改善する必要がある．

今後，VPYLM を基に本手法を拡張していくことで，よりユーザの満足を得られる自動作曲システムに近づけたい．

参　考　文　献

1) 白井亨，谷口忠大: ギブスサンプリングを用いたインタラクティブ作曲システムの提案，ヒューマンインタフェースシンポジウム 2010，論文集，3412 (2010).

Vol.2011-MUS-91 No.3
2011/7/27

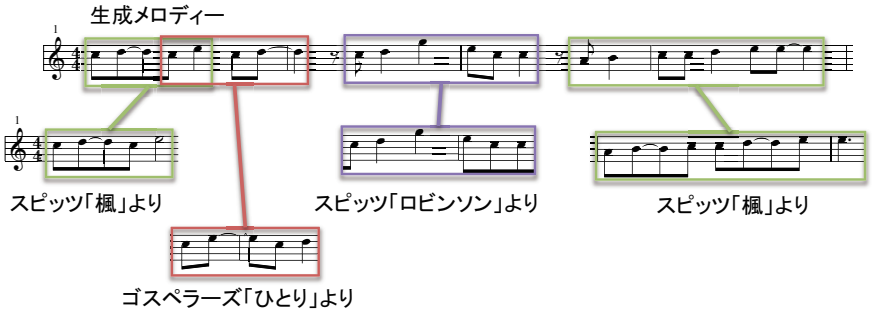

図 6 生成メロディー
Fig. 6 The Generated melody.

2) 高木英行，畝見達夫，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).
3) 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.
4) Teh, Y.W.: A Hierarchical Bayesian Language Model based on Pitman-Yor Processes, Proc.COLING/ACL 2006, pp.985-992 (2006).
5) Kneser, R. and Ney, H.: Improved backing-off for m-gram language modeling, Proc. ICASSP, Vol.1, pp.181-184 (1995).
6) 持橋大地，隅田英一郎: 階層 Pitman-Yor 過程に基づく可変長 n-gram 言語モデル，情報処理学会論文誌，Vol.48, No.12, pp4023-4032 (2007).