

ベイズ推定による楽曲間内挿に基づく 和音モーフィング法の提案

本城 伸隆¹ 榎田 皓太¹ 徳永 旭将^{1,a)}

概要：本稿では、コンピュータグラフィックス分野で研究されている metropolis procedural modeling と同様のコンセプトに基づき、複数のコード進行間を内挿する和音モーフィング法を提案する。Metropolis procedural modeling は確率的文法に基づく再帰的手続きとベイズ推定を組み合わせることで、ユーザが所望するプロパティをもつ 3D モデルを自動生成する手法である。提案手法では、参照楽曲のコード進行を局所的に変化させ、ターゲット楽曲の雰囲気反復的に近づけていくアプローチを採用する。モーフィングプロセスの実現には、確率的文脈自由文法を model prior として用い、Metropolis-Hastings アルゴリズムにより事後分布からのランダムサンプリングを行う。具体的な手続きについては、簡単な評価実験により示す。

An chord morphing method based on an interpolation of existing chord progressions by Bayesian inference

NOBUTAKA HONJO¹ KOHTA ENOKIDA¹ TERUMASA TOKUNAGA^{1,a)}

Abstract: In this paper, we propose a chord morphing method by interpolating existing two chord progressions. The proposed method relies on the Bayesian computation technique called metropolis procedural modeling, which has been established for generating computer graphics models of desired property in a 3D space. We built a generative model for deriving new chord progressions based on a stochastic context-free grammar. In the morphing process, local structures of chord movements are iteratively changed based on Metropolis-Hastings rule. The morphing procedure is demonstrated with a simple experiment.

1. はじめに

統計的機械学習を駆動力とした音楽信号処理が目覚ましい発展を遂げている一方、言語的な側面に着目した音楽情報処理技術も盛んに研究されている。中でも、隠れマルコフモデル (HMM) のような確率的言語モデルは、自動作曲や自動伴奏技術で広く用いられてきた [1] [2]。HMM は音や和音の遷移にマルコフ性を仮定した単純明快なモデルであるが、リズムを明示的に取り扱えないという欠点もある。リズムの記号的な取り扱いについては、リズムの階層性を木構造で表現する Generative Theory of Tonal Music

(GTTM) が古くから知られており [3]、メロディモーフィングなどに応用されている [4]。これらの先進的応用として、近年では確率的文脈自由文法 (PCFG) に潜在変数を組み込んだメロディへの自動伴奏付与などの研究などが行われている [5]。

このように、統計的言語モデルに基づく音楽生成は応用範囲を拡大しつつあるが、既存のコード進行から新たなコード進行を派生する試みはあまり研究されていない。コード進行はそれ単体でも楽曲のジャンルや雰囲気、個性を強く決定づける要素である。また、現代の音楽制作の現場では、メロディより先にコード進行が決められるケースは珍しくない。さらに、ポピュラー音楽では、様々な楽曲で使われてきた定番のコード進行がいくつも存在する。これらのことから、既存のコード進行からそれと類似した新たなコー

¹ 九州工業大学
Kyushu Institute of Technology, Kawazu 680-4, Iizuka,
Fukuoka, Japan

^{a)} tokunaga@ces.kyutech.ac.jp

ド進行を派生させる技術は、音楽制作の現場で有用と思われる。

画像生成分野では、複数の画像を貼り合わせて新たな画像を生成するイメージ・スティッチング技術や [6]、複数の人物の顔画像から中間的な顔画像を生成するフェイス・モーフィング技術 [7] が盛んに研究されている。また、医療画像処理分野では、自由変形する物体の歪みや位置変化を画像処理によって補正するイメージ・アライメント技術が広く研究されている [8]。画像生成や画像処理では、対象物の幾何学的な特徴を学習する際に、画像ピラミッドと呼ばれるような多解像度処理を行うことが多い。前述のように音楽情報処理においても、リズムの階層性を表現する言語モデルが用いられてきた。従って、複数のコード進行から新たなコード進行を派生させる和音モーフィング技術も、同様のアイデアにより開拓できるはずである。本研究の目的はその実証の第一ステップとして、我々が考える和音モーフィング技術の概念を示し、簡単な評価実験によりその有効性を検証することである。

図 1 に、提案手法の概念を示す。本稿では和音モーフィング技術を「参照和音系列を局所的に変化させ、音楽的に不自然なコード進行にならないよう注意しながら、ターゲット和音系列に徐々に近づけていく処理」と仮定する。ただし、参照和音系列をターゲット和音系列に記号的な意味で一致させるのではなく、同様の特性をもった和音系列を生成することを目的とする。それにより、参照和音系列とターゲット和音系列双方の印象を備えた新たな和音系列を多数生成する。コード進行は記号により完全に表現できる要素であることから、確率的文脈自由文法のような言語モデルと、計算ベイズの確率的な進化発展により実現することができるとされる。コンピュータ・グラフィックス分野では、手続き型の言語モデルとベイズ推論を組み合わせた方法論 (metropolis procedural modeling) が研究されている [9]。本稿では、この metropolis procedural modeling に着想を得た和音モーフィング法を提案する。

2. 和音モーフィング手法

2.1 和音系列生成のための手続き型モデル

まずはじめに、和音系列を生成するためのモデルについて述べる。本研究で用いる確率的文脈自由文法 G を次のように表す：

$$G = (V, \Sigma, R, S) \quad (1)$$

ここで、 V は非終端記号の集合、 Σ は終端記号の集合、 R はあるルールを適用する確率の集合、 S は開始記号である。 Σ には和音の種類を示すコード記号が対応する。 S は木の根に相当する終端記号である。ルールを適用する確率としては、非終端記号 X を終端記号 F に変更する確率 P と、非終端記号 X を 2 つの非終端記号 $F[X][X]$ に分岐させる確

入力1：参照和音系列



- ✓ ハイブリッドな和音系列を多数生成
- ✓ 参照和音系列とターゲット和音系列の間を内挿

入力2：ターゲット和音系列



図 1 本稿における和音モーフィングの概念。

率 $1 - P$ を考える。このプロセスは、分枝要素 2 の木構造を生成することに対応する。

次に、本研究で取り扱うリズムの階層性について定める。まず、和音を鳴らす最小のリズムを考える。最小のリズムは、生成する木の深さの上限に対応する。和音を鳴らすリズムを細かくするほど、すなわち深い構造の木ほど、複雑で多様なリズムを表現することができる。その反面、推定すべきパラメータ数が多くなってしまふ。本研究では簡単のために、4/4 拍子における 1 拍分を和音を鳴らす最小単位とする。

リズムの階層構造に加え、コード進行では横方向の構造についても考慮する必要がある。音楽では一般に A メロ、B メロ、サビなどと呼ばれる、複数小節にまたがる構造が存在する。それにより、起承転結といったストーリー性が表現されることが多い。このコード進行の大局的構造を木構造に反映させることで、実用的なコード進行を派生できる可能性が高くなると思われる。しかし本研究では、多様な和音系列を生み出すという観点から、このような横方向の構造は極力仮定しないことにする。

図 2 に、提案モデルにより生成される文字列とリズムの関係を示す。本研究では、和音は必ず 1 拍ごとに鳴らすと仮定し、1 小節を部分木の単位とする。従って、非終端記号を 2 つの非終端記号に分岐する回数は 2 回が上限である。ルートノードのみであれば 4 拍同じコードを連続で鳴らすことを意味する。同様に、深さ 1 の葉ノードは 2 拍分の和音、深さ 2 の葉ノードは 1 拍分の和音に対応する。

2.2 和音系列生成のためのベイズ推論

前節で述べた確率的文脈自由文法 G によりユーザが所望する和音系列を導出するために、ベイズ推論のアイデアを導入する。いま、ユーザが求めるコード進行の性質が、プロパティ I として与えられているとする。このとき我々の目的は、 G により導出され得る和音系列の空間 $\Delta(G)$ から、

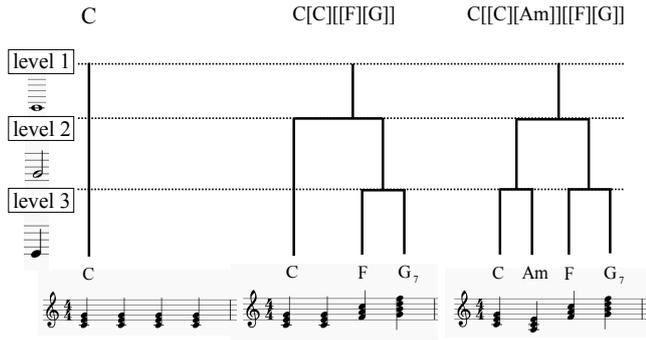


図 2 生成される文字列とリズムの関係

プロパティ I に最も近いものを選ぶ最適化問題になる。これは、確率的推論という観点では、次のように事後確率最大化問題として定式化できる：

$$\hat{\delta} = \underset{\delta}{\operatorname{argmax}} p(\delta|I), \quad p(\delta|I) \propto L(I|\delta)\pi(\delta) \quad (2)$$

ここで、 $\delta \in \Delta(G)$ 、 $L(\cdot)$ は導出された和音系列 δ に対する尤度、 $\pi(\cdot)$ は生成モデルの事前分布である。

2.3 事前分布

モデルの事前分布 $\pi(\cdot)$ は、文法 G の生成確率により記述することができる：

$$\pi(\delta) \propto \prod_{s \in \delta} P(s|\operatorname{parent}(s)) \prod_{t \in \delta \rightarrow} \Phi_t(\varphi_t)$$

ここで、 $P(\cdot)$ は木の生成確率、 s は δ に含まれる全ての successor である。また、 φ_t は和音名と 1 対 1 に対応する記述パラメータであり、その実現値は確率分布 Φ_t から生成されるとする。確率分布 Φ_t は、音楽理論的な観点から設定される。例えば、音楽理論ではあるキーを構成する音から作られる和音をダイアトニックコードと呼ぶが、ポピュラー音楽においてはそのキーのダイアトニックコードは他のコードよりも頻繁に登場することが多い。また、ダイアトニックコードの中でも、主要三和音と呼ばれる重要な役割をもつ和音は、A メロ B メロといった大局的構造と連動して出現することが多い。さらに、ジャズであればテンションコードと呼ばれる、緊張感のある響きの和音が高い確率で選ばれと仮定できる。このように、 Φ_t の与え方により、ターゲットとする音楽のジャンルやコードの役割に関する事前情報を暗に与えることができる。

2.4 尤度

本研究では、既存のコード進行から新たなコード進行を派生させることを目的としている。いま、目標とする和音系列 δ_I があり、そのプロパティを I とした場合、尤度は確率的文法 G により導出された和音系列 δ の I に対するフィッ

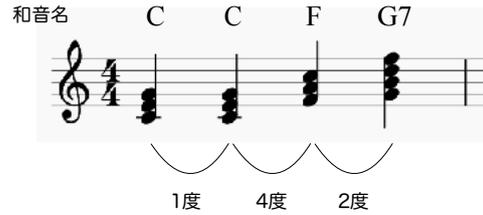


図 3 和音系列に出現する音の跳躍の数え方。コードのルート音の跳躍のみを数える。

ティングの良し悪しを表すと解釈できる。和音系列のプロパティとしては様々な特徴が考えられるが、ここではコードの遷移という特徴に着目する。 $F = \{-8, -7, \dots, 9\}$ を予め想定した音の跳躍の集合としたとき、尤度を次のように与える：

$$L(I|\delta) = \frac{1}{\sqrt{2\pi\sigma_I^2}} \exp\left(-\frac{k(\delta, \delta_I)}{2\sigma_I^2}\right), \quad (3)$$

$$k(\delta, \delta_I) \propto \sum_{f \in F} w(f)\phi_f(\delta)\phi_f(\delta_I),$$

$$\phi_f(\delta) \in \{0, 1, 2, \dots\}, \quad \phi_f(\delta_I) \in \{0, 1, 2, \dots\}$$

ここで、 $w(f) \in \mathbb{R}_+$ は音の跳躍の重要度を示す係数、 $\phi_f(\delta)$ は導出された和音系列 δ 内に出現する跳躍の頻度を数え上げる関数である。本研究では図 3 に示すように、コードのルート音（和音を構成する音の中で最も低い音）の跳躍のみを数え上げる。つまり、尤度はメジャーやマイナー、7th, sus4 といった単体の和音の性質には依存しない。

2.5 和音系列の遷移法則

式 2 の最適化問題を厳密に解く代わりに、事後分布からのランダムサンプリングを考える。我々の目的はユーザが所望するプロパティ I_δ を満たす和音系列 δ を推定ことである。しかし、生成された和音系列がプロパティ I_δ を厳密に満たす必要はない。むしろ、既存の和音系列から新たな和音系列を派生させるという観点では、プロパティ I_δ に近い性質を持つ和音系列をできるだけ多数得ることが重要である。従って本研究では、MCMC (Markov Chain Monte Carlo) の 1 種である Metropolis-Hastings (M-H) 則 [10] [11] に基づき和音系列を局所的に変化させながらサンプリングを行う。

M-H 法では提案分布に従って遷移先候補点を生成する。ここでは、図 4 に示すような 2 つのタイプの遷移を考える。一つは、任意の終端記号を和音系列全体からランダムに一つ選び、和音を別の和音に置き換える遷移である。もう一つは、1 つの終端記号を 2 つの終端記号に分割する、もしくは 2 つの終端記号を 1 つに統合する遷移である。この場合、ある和音を連続して鳴らす回数も変化することになる。これら複数の遷移を確率的に選択し、コード進行の局所的な置換、複雑化、簡約化を繰り返しながら、新たな和音系列を

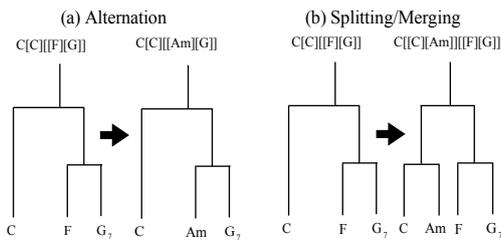


図 4 和音系列の遷移の種類.

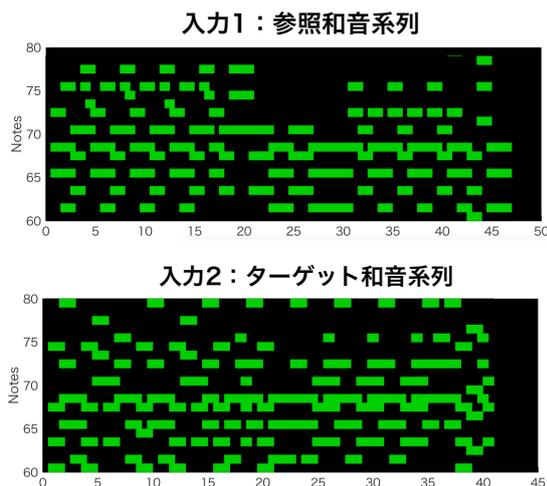


図 5 入力データのピアノロール.

多数生成する.

任意の終端記号のパラメータ φ_t を φ'_t に置換するときは, $\varphi'_t \sim \Phi_t$ のように事前分布からの独立リサンプリングを考える. このとき, M-H 則では遷移先候補を次の確率で採択する:

$$\alpha_{\delta \rightarrow \delta'} = \min \left\{ 1, \frac{p(\delta'|I)\Phi_t(\varphi_t)}{p(\delta|I)\Phi_t(\varphi'_t)} \right\} = \min \left\{ 1, \frac{L(\delta'|I)}{L(\delta|I)} \right\}$$

3. 実験

既存の和音系列を入力としたときの提案手法の評価実験を行う. 図 5 に入力データのピアノロールを示す(ただし, 実際に入力するデータはコード記号の遷移を表す文字列である). 入力データは Ab キーのコード進行のうち, 曲の構成が比較的近いものを選んだ. 上図が δ の初期状態となる参照和音系列, 下図はターゲット和音系列である.

前章で示した M-H 則に基づき, 遷移を 5000 回繰り返した際の尤度の推移を図 6 に示す. 式 3 で示したガウス関数の分散は $\sigma_l = 0.01$ とした. また, 同じく式 3 で示した w については, 全ての f に対して同じウェイトとした.

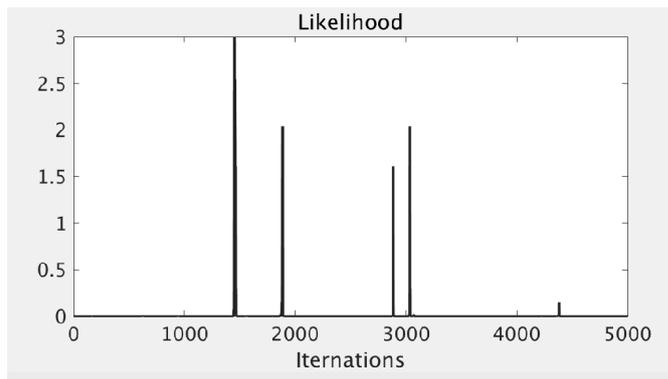


図 6 ランダムサンプリングの過程における尤度の遷移.

図 7 に, 尤度が比較的大きな値をとる和音系列のルート音跳躍のヒストグラムを示す. 横軸が跳躍の音程を表し, 縦軸が和音系列内でのその跳躍の出現回数である(跳躍がない場合が 1 度である点に注意されたい). 青の棒グラフが各反復で生成された和音系列のヒストグラム, 赤の棒グラフはターゲット和音系列のヒストグラムに対応する. これらのヒストグラムを比較すると, 尤度が大きな和音系列ほど, ターゲット和音系列と近いコード進行パターンが出現してしていることがわかる. 図 8 は, 生成した和音系列のうち尤度が比較的高い和音系列の 16 小節分の楽譜を, 尤度の値により順序づけしたものである. このように尤度の値によって順序づけすることにより, 生成されたコード進行がどの程度ターゲット和音系列の特徴に近いかを俯瞰ことができる.

4. おわりに

本稿では, computer graphics 分野で研究されている metropolis procedural modeling と同様のコンセプトに基づき, 既存のコード進行間を内挿する和音モーフィング法を提案した. 画像生成分野ではディープラーニングなどの教師あり機械学習の応用が目覚ましい成果を上げつつある. 音楽生成においても同様の潮流があるが [12], 画像生成分野と比べるとベンチマークとして利用できる訓練データセットが少ない. 和音モーフィング技術は, 既存の訓練データから新しいデータを派生させることで訓練データの「水増し」を行うデータ・オグメンテーション技術にも応用できる. 従って, 音楽制作の現場のみではなく, 教師あり機械学習に基づく楽曲生成という観点からも有用である.

本研究では, 和音系列に出現するルート音の跳躍のヒストグラムをプロパティとして用いた. 実験では和音のルート音の跳躍というプロパティに基づきモーフィングを行なった. コード進行のプロパティにはそれ以外にも, コード進行の複雑さ, 雰囲気, 緊張感などが考えられる. それら

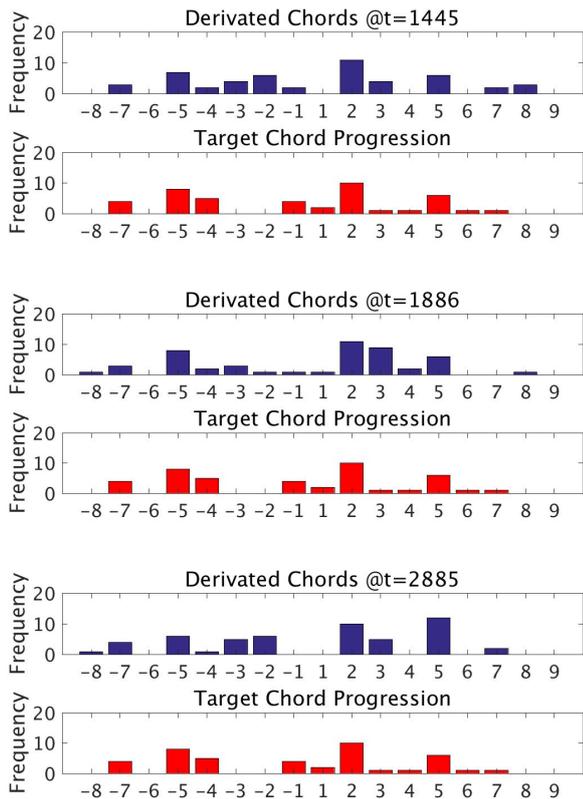


図 7 尤度が比較的大きな値をとる和音系列のルート音跳躍の頻度分布。



図 8 生成した和音系列のうち尤度が比較的高いものを尤度の値により順序づけしたものを。

を複合的に用いることで、より効果的な和音モーフィング技術が実現できるはずである。実験では示さなかったが、尤度の分散を調整することで、生成される和音系列がどの程度ユーザが指定したプロパティに拘束されるかを制御することができる。このように、意外性や多様性をどの程度許容するかを明示的に制御できることは、音楽創作支援によって重要な要素であると我々は考える。

本研究では、楽曲のもつ起承転結といった横方向の構造を仮定しなかった。従って、現状では曲の長さや構成が大きく異なる和音系列間を内挿することは困難である。これに対しては、例えば確率的言語モデルに楽曲の構成を潜在変数として組み込むなどの方法が考えられる。このように楽曲の構成を prior として取り入れることは、実用的な和音モーフィング技術を実現する上で不可欠な課題になるだろう。

本研究では簡単な評価実験を行い初期結果を報告したが、生成された和音系列が音楽としてのどの程度自然なものであるか、十分な検証ができていない。モーフィングプロセスにおいてどの程度音楽の定石から逸脱したコード進行を許容するかについては、model prior に依存する。極端に言えば、音楽理論的に少しでも不可解なコード進行には 0 に近い事前確率を設定すれば、ほとんどのケースで自然なコード進行が得られるはずである。しかしながら音楽制作支援という観点では、自然なコード進行を提示するだけのシステムが最良とは限らない。画像生成分野では、作品としての自然さと創造性の両立に切り込んだ研究も注目されつつある [13]。和音モーフィングでも、新規性と自然さのバランスを制御できる仕組みの実現は挑戦的課題である。

参考文献

- [1] Takeda, H., Saito, N., Otsuki, T., Nakai, M., Shimodaira, H. and Sagayama, S.: Hidden Markov Model for Automatic Transcription of MIDI Signals, *MMSP*, pp. 428–431 (2002).
- [2] Simon, I., Morris, D. and Basu, S.: MySong: Automatic Accompaniment Generation for Vocal Melodies, *CHI 2008 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 725–734 (2008).
- [3] Lerdahl, F. and Jackendof, R.: A Generative Theory of Tonal Music, *MIT Press, Cambridge* (1983).
- [4] Hamanaka, M. and Hirata, K.: Melody Morphing Method based on Time-span Tree, 情報処理学会研究報告音声言語情報処理 (SLP) (2008).
- [5] 津島啓晃, 中村栄太, 糸山克寿 and 吉井和佳: ベイズ文脈自由文法に基づく和音系列の教師なし構文解析と自動生成, 情報処理学会第 79 回全国大会 (2017).
- [6] Mann, S. and Picard, R.: Virtual bellows: constructing high quality stills from video, *Image Processing, 1994. Proceedings. ICIP-94, IEEE International Conference* (2002).
- [7] Kasat, D. R., Jain, S. and Thakare, V. M.: A survey of face morphing techniques, *IJCA Special Issue on Recent Trends in Information Security* (2014).
- [8] Mani, V. and Arivazhagan, S.: Survey of Medical Image Registration, *Journal of Biomedical Engineering and Technology*, Vol. 1, No. 2, pp. 8–25 (2013).
- [9] Talton, J. O., Lou, Y., Lesser, S. and Duke, J.: Metropolis Procedural Modeling, *ACM Transactions on Graphics*, Vol. 30, No. 2 (2011).
- [10] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E.: Equations of state calculations by fast computing machines, *J. Chem. Phys.*, Vol. 21, No. 6

(1953).

- [11] Hastings, W.: Monte Carlo sampling methods using Markov chains and their application, *Biometrika*, Vol. 57, pp. 97–109 (1970).
- [12] Briot, J.-P., Hadjeres, G. and Pachet, F.: Deep Learning Techniques for Music Generation – A Survey, *arXiv:1709.0162* (2017).
- [13] Elgammal, A., Liu, B., Elhoseiny, M. and Mazzone, M.: CAN: Creative Adversarial Networks, Generating Art by Learning About Styles and Deviating from Style Norms, *arXiv:1706.07068* (2017).