

# 自動作曲を用いた音楽スタイルの進化実験における混合継承の効果

Effect of blending inheritance in experimental evolution of music styles using automatic composition

中村 栄太 \*1      金子 仁美 \*2      伊藤 貴之 \*3      金子 邦彦 \*4  
Eita Nakamura      Hitomi Kaneko      Takayuki Itoh      Kunihiko Kaneko

\*1 京都大学      \*2 東京芸術大学      \*3 お茶の水女子大学      \*4 Niels Bohr 研究所  
Kyoto University      Tokyo University of the Arts      Ochanomizu University      Niels Bohr Institute

We report results of an experimental evolution to study the mechanism of music style evolution. Creative cultures including music are developed by descent-with-modification processes of knowledge about creating complex artifacts, but how new creation styles emerge and the influence of listeners/consumers through a social evaluation process are not well understood. To study these problems, we conducted an experiment where a population of automatic composition models, which simulates a population of human creators, evolves while being evaluated by many listeners. As a result, we found that adaptive evolutions of music styles can occur when blending inheritance of high-dimensional statistics representing composition styles is incorporated in the generation update process. A significant difference in musical preference depending on musical experience was also found. The results suggest a possibility of constructing a system of automatic composition in a new and preferred music style by the experimental framework.

## 1. はじめに

人の創造的な知能を理解するため、音楽の自動生成に関する研究が行われている。自動作曲研究の歴史は1950年代にまで遡ることができ [Nierhaus 09], これまでに論理規則に基づく方法 [Ebcioglu 90], 統計モデルに基づく方法 [Pachet 11, Fukayama 10], 遺伝的アルゴリズムを用いる方法 [Biles 94] など、様々なアプローチによる研究がある。近年では、深層ニューラルネットワークなどを用いた機械学習に基づく手法が数多く提案されている [Hadjeres 17, Sturm 16, Mittal 21, Agostinelli 23]。一般的に、機械学習に基づく手法では高品質な曲の生成が可能となる一方で、生成される楽曲のスタイルは学習データに依存しているため、人の創作者のように新しくかつ人々に好まれる作曲スタイルを生み出すことは難しい。本研究では、音楽スタイルの進化の仕組みを調べるための、自動作曲を用いた進化実験を考える。

音楽の作曲には、音階やリズム、和声に関する高度な知識が必要であり、音楽文化の発展は作曲知識の継承と変更の繰り返しにより起こっている。また、音楽スタイルの時代変化では方向性を持った作曲スタイルの変化が観測されており [Mauch 15, Interiano 18, Singh 22], 継承の際に選択や変異の効果が重要であることが示唆されている。DarwinTunes 実験 [MacCallum 12] では、進化計算の枠組みを用いて、自己複製するループ音生成アルゴリズムと多数の鑑賞者の評価に基づく選択により、比較的音楽性が高いアルゴリズムが出現することが確かめられた。この実験における音生成アルゴリズムを無制限に曲を生成できる自動作曲モデルに置き換えることで、新規の作曲スタイルが出現する過程を調べられると考えられる。この際に、多数のパラメータにより表される作曲スタイルの空間で、どのような継承と変異のモードにより適応的進化が起こるかを理解する必要がある。

本研究では、異なる作曲スタイルの融合による新しい音楽スタイルの出現を模した過程として、作曲スタイルを表す統計量の混合継承を組み込んだ進化過程の可能性について調べる。

連絡先: 中村栄太, 京都大学, 606-8501 京都府京都市左京区吉田本町, eita.nakamura@i.kyoto-u.ac.jp

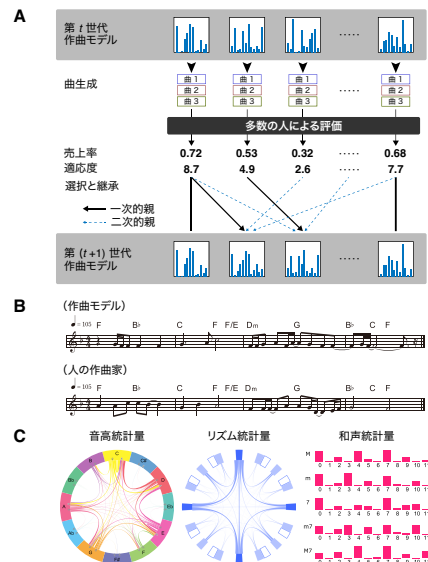


図 1: A: 進化実験システムの概要. B: 作曲モデルが生成したメロディーの例. 同じ条件で人の作曲家が作曲したメロディーも示している. C: 作曲スタイルを表す統計量の一部.

進化実験システムでは、与えられたコード進行と歌詞に対してメロディー（音符列系列）を生成する自動作曲モデルを用い、生成された曲に対する多数の人の評価を基に作曲モデルの適応度を定義する。本実験の結果を分析して、作曲スタイルの進化における混合継承の効果と鑑賞者の嗜好性の構造について調べる。

## 2. 自動作曲を用いた進化実験の方法

進化実験では、自動作曲モデルの集団を作り、生成された曲を多数の人が評価して、その結果に基づいて作曲モデル集団の世代更新を行う（図 1A）。作曲モデルは、与えられたコード進行と歌詞に対してメロディーを生成する（図 1B）。ここ

で、コード進行は小節数と各コードの長さの情報を含み、歌詞は小節ごとに読みがなで与えられ、各言語音に1音符を対応させるものとする。具体的な作曲モデルとして、本研究では複数の統計モデルを統合した Product of experts モデルを用いる。主要な構成要素のモデルとしては、メロディー音符列の局所的な特徴を記述するマルコフモデル [Nakamura 19] やメロディーとコード進行の関係性をコード記号に条件付いた音高の確率により表す和声作用モデルがあり (図 1C), これらのパラメータを統計的形質と呼ぶ。統計的形質の次元数は約 10 万である。この他、メロディーの一般的な性質を表すために、日本語歌詞に含まれる高低アクセントと音程との関係を表す歌詞作用モデルや反復構造を含むメロディーの概形を確率的に表すモデルなどが用いられているが、本実験ではこれらのモデルパラメータの値は固定している。

各世代において、50 個の作曲モデルからなる集団を作る。世代更新の際には、後述の方法で計算される各作曲モデルの適応度  $w_n$  を用いて、次の世代の作曲モデルが以下の手続きで構成される。まず、適応度が上位の 25 個の作曲モデルは次の世代のメンバーとし、残りの 25 個は次の世代には含めない。新しく導入される残りの 25 個の作曲モデルの各々について、確率  $P(n) \propto w_n$  により (一次的) 親と 2 個の二次的親の合計 3 つの作曲モデルを前の世代の集団から選択する。統計的形質は、これらの親から混合継承により伝達される。つまり、斜行伝達係数  $u$  を  $0 \leq u \leq 1/2$  を満たす定数として、親の形質  $\theta_0$  と 2 個の二次的親の形質  $\theta_k$  ( $k = 1, 2$ ) から、子の形質  $\theta = (1-u)\theta_0 + u(\theta_1 + \theta_2)/2$  と継承される。これは、作曲スタイルが親の作品に加えて、二次的親の作品から学習される状況を表している。なお、この進化過程は、仮に選択圧が働かない場合には、作曲モデルの集団が最終的に、初期作曲モデルの統計的形質を平均して得られる平均モデルに収束する過程になっている。係数  $u$  自体は、一様分布に従う変異を含みながら、親から伝達されるものとする ( $u' \sim u + U(-0.05, 0.05)$ )。以上により、子の統計的形質は親とは少し異なり、新しいスタイルのメロディーが生成できるようになる。

最初の世代の作曲モデルの形質は以下のように初期化する。まず、日本および欧米のポピュラー音楽の計約 1 万曲のデータを用いて混合マルコフモデルを用いてクラスタリングを行い、得られた 25 クラスのデータから最尤推定した統計的形質を持つ作曲モデル (原始作曲モデルと呼ぶ) を 25 個作る。第 1 世代の 50 個の作曲モデルの統計的形質  $\theta$  は、各々に対してランダムに抽出した 3 個の原始作曲モデル  $i, j, k$  の平均  $\theta = (\theta_i + \theta_j + \theta_k)/3$  とおく。斜行伝達係数は、一様分布からサンプルする:  $u \sim U(0.1, 0.3)$ 。

以下の方法で、作曲モデルが生成した曲を人が聴いて評価することで、世代更新の際に用いる適応度を求める。各世代において歌詞付きのコード進行を 20 曲分用意して、各作曲モデルはこれらのコード進行に対してメロディーを生成する。歌詞は全て「ららら」で歌い、伴奏音は全ての作曲モデルに対して同一のものを用いており、比較評価はメロディー音符列に基づいて行われるように設計している。鑑賞者による評価はランダムに提示した 2 個の作曲モデルの曲に対して、良いと思う方を選択することで行う。これは現実の社会における比較選択に基づく消費・鑑賞活動を模している。これを複数の鑑賞者で多数回行った結果、作曲モデル  $n$  の曲が選ばれた回数を  $s_n$ 、その作曲モデルの曲が聴き比べされた回数を  $c_n$  と記し、規格化された数  $s_n/c_n$  を売上率と呼ぶ。売上率を用いて、作曲モデルの適応度は  $w_n = \exp(\beta_s s_n/c_n)$  と定める。今回の実験では、謝金を支払って協力してもらった約 70 名の大学生・大学院生

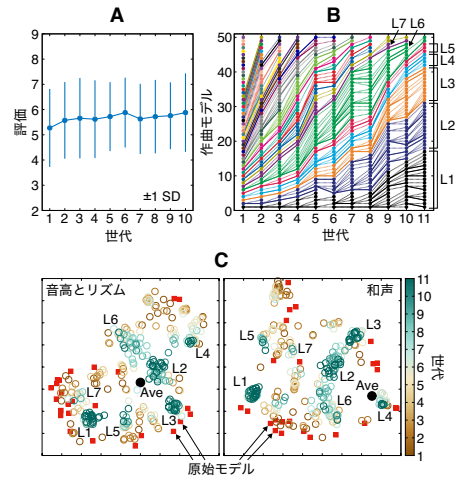


図 2: 進化実験の結果. A: 平均評価値の変化. B: 作曲モデルの系統図. 世代を超えて残った個体を実線, 親子関係を点線で示している. C: 統計的形質空間における進化の様子を t-SNE による 2 次元可視化空間上で示した図.

とその他数名が鑑賞者として参加した。鑑賞者には音楽の鑑賞時間、創作 (演奏および作曲) の経験、作曲法 (和声法) の学習経験についてのアンケートに答えてもらい、聴き比べの結果に加えて、各曲に対して 10 段階の採点評価もしてもらった。

### 3. 結果

進化実験では数日毎に 10 回の世代更新を行い、第 11 世代までに 6 つの系統にまで淘汰された (図 2A)。以下では、第 10 世代まで残った主要 7 系統のラベルを L1 から L7 と表す。採点による平均評価値は増加したことから、選択による効果が確認された (図 2B)。音高とリズムのスタイルおよび和声作用のスタイルを表す統計的形質を t-SNE を用いて 2 次元可視化した結果では、各系統に属する作曲モデルがクラスターを成しており、実験過程でスタイルが大きく変化した L2 などの系統が確認できる (図 2C)。

作曲スタイルの進化の効果を確認するために、主要 7 系統の第 1 世代と第 10 世代の最も高評価だった作曲モデル、25 の原始モデル、全データから学習した平均モデル、日本のポピュラー音楽データから学習した 19 の作曲モデルを比較する追加実験を行った。この追加実験は、Web 上の公開サイト crevo-music.com で一般の利用者がこれらの作曲モデルを自動作曲に使用して聴き比べすることにより行った。この結果、L2 と L7 の系統で有意な売上率の向上が見られ、混合継承を組み入れた進化実験で作曲スタイルの適応進化が起きたことが確認された。特に高評価だった L2 の第 10 世代のモデルは、比較した全てのモデルの中で最も売上率が高く、特に、平均モデルよりも有意に売上率が高かった。これは、平均的な作曲スタイルが必ずしも高評価ではないことを示唆している点で興味深い [Langlois 90]。これにより、統計的形質の空間内の適応度地形は非自明な構造を持ち、その中で多数の人の評価による選択圧により適応進化が起こったことが示された。また、混合継承の過程をトラックする分析では、L2 系統では多くの混合継承が行われ、進化により作曲スタイルが大きく変化したのに対して、L7 系統では、混合継承の効果は小さく、作曲スタイルの変化も比較的小さいことがわかった。

進化実験における鑑賞者の嗜好性の構造について調べるた

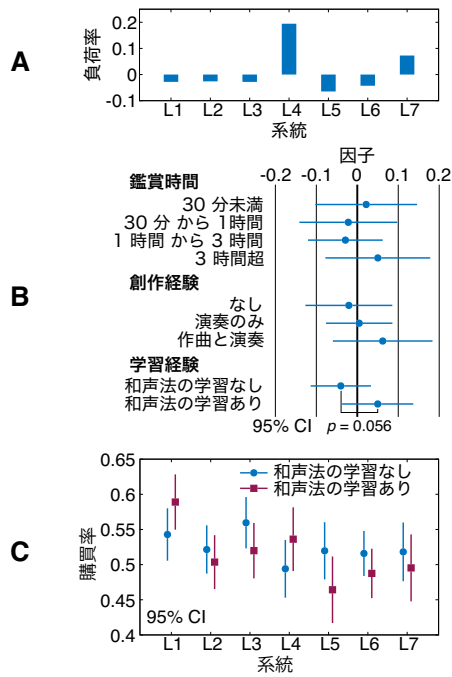


図 3: 鑑賞者の嗜好性の分析. A: 購買率の因子分析における第 1 因子の負荷率. B: 因子と鑑賞者の属性の関係. C: 作曲法の学習経験の有無に対する各系統の購買率の差.

め、各鑑賞者ごとに主要 7 系統の作曲モデルが生成した曲を選択した回数を規格化した購買率を分析した。因子分析による結果では、作曲法の学習経験の有無により因子の平均値にやや有意な差異が見られ、鑑賞時間と創作経験に対しては有意な差異は見られなかった (図 3A,B)。作曲法の学習経験で分割した鑑賞者のグループ毎の平均購買率の比較では、L1 と L4 系統は作曲法の学習経験有りのグループで高評価になる傾向があり、L3 と L5 系統は逆の傾向が見られた (図 3C)。また、線形判別分析 (LDA) を用いた結果では、これら 2 群の平均購買率には有意な差異 ( $p = 1 \times 10^{-4}$ ) が見られた。

主要 7 系統の作曲モデルのスタイルの特徴を分析するため、音高とリズム、和声のスタイルについて定義した 31 の特徴量を各系統の第 10 世代の最高評価の作曲モデルについて比較分析した。この結果、平均モデルとの比較における顕著な特徴は、7 系統の全てで少なくとも部分的に異なっていた。中でも、ダイアトニック音階以外の音高を生成する確率は、L1 と L5 が平均よりも高く、L3 と L6 が平均よりも低く、シンクペーションを含むリズムを生成する確率は、L1 と L2、L6、L7 のいずれも平均よりも低かった。また、作曲法の学習経験が有る鑑賞者に高評価だった L1 と L4 系統に共通して、他には見られない特徴としては、短三和音における非和声音の確率が平均よりも高いという特徴があった。

#### 4. 結論

本研究では、自動作曲モデルの集団を多数の人の評価のもとで進化させる実験を行い、約 10 万次元の作曲スタイルの空間において混合継承を組み込んだ世代交代により適応進化が起こることを示した。これは、作曲スタイルの進化において、複数の作曲スタイルの創作者あるいは楽曲から学習を行う過程の重要性を示唆している。一方で、この方法では、既存の作曲スタイルから大きく外れた全く新しい作曲スタイルが出現させ

ることは困難であり、この他の変異モードについても今後調査を行う必要がある。

また、鑑賞者の嗜好性の構造分析を行い、作曲法の学習経験の有無が嗜好性の差異を生む要因の一つと判明した。これは現実社会では創作者と鑑賞者の間での評価の違いを示唆しており、今後、自動作曲技術が社会の大多数である鑑賞者の嗜好性に合わせて適応するとすると、従来の音楽文化の進化とは異なる方向に向かう可能性も示している。こうした多様な嗜好性を持つ鑑賞者の存在は、現実の社会で見られる様な複数の音楽スタイルが共存する構造 [Mauch 15, Singh 22] を理解するためにも重要だと考えられる。

今後、作曲スタイルの進化実験を繰り返すことにより、作曲スタイル空間の適応度地形の構造がより詳細に理解でき、その静的部分と動的部分 (頻度依存バイアス [Nakamura 19, Youngblood 19] など) について調べられると考えられる。また、従来の音楽文化では発見されていない作曲スタイルで人々に好まれるものが発見できる可能性についても調査を続ける予定である。

#### 謝辞

本研究は、日本学術振興会科研費 19K20340, 21K02846, 21K12187, 22H03661 および Novo Nordisk Foundation から支援を受けた。実験システム構築に協力頂いた柴田剛氏、錦見亮氏、Du Yicheng 氏、齋藤遼太郎氏、大山偉永氏、および実験に協力頂いた皆様に感謝する。

#### 参考文献

[Agostinelli 23] Agostinelli, A., et al.: MusicLM: Generating music from text (2023), arXiv:2301.11325

[Biles 94] Biles, J.: GenJam: A genetic algorithm for generating jazz solos, in *Proc. International Computer Music Conference (ICMC)*, pp. 131–137 (1994)

[Ebcioglu 90] Ebcioglu, K.: An expert system for harmonizing chorales in the style of J. S. Bach, *Journal of Logic Programming*, Vol. 8, No. 1, pp. 145–185 (1990)

[Fukayama 10] Fukayama, S., et al.: Automatic song composition from the lyrics exploiting prosody of the Japanese language, in *Proc. Sound and Music Computing Conference (SMC)*, pp. 299–302 (2010)

[Hadjeres 17] Hadjeres, G., Pachet, F., and Nielsen, F.: DeepBach: a steerable model for Bach chorales generation, in *Proc. International Conference on Machine Learning (ICML)*, pp. 1362–1371 (2017)

[Interiano 18] Interiano, M., Kazemi, K., Wang, L., Yang, J., Yu, Z., and Komarova, N. L.: Musical trends and predictability of success in contemporary songs in and out of the top charts, *Royal Society Open Science*, Vol. 5, p. 171274 (2018)

[Langlois 90] Langlois, J. H. and Roggman, L. A.: Attractive faces are only average, *Psychological Science*, Vol. 1, No. 2, pp. 115–121 (1990)

- 
- [MacCallum 12] MacCallum, R. M., et al.: Evolution of music by public choice, *Proceedings of the National Academy of Sciences of the USA*, Vol. 109, No. 30, pp. 12081–12086 (2012)
- [Mauch 15] Mauch, M., MacCallum, R. M., Levy, M., and Leroi, A. M.: The evolution of popular music: USA 1960–2010, *Royal Society Open Science*, Vol. 2, No. 150081 (2015)
- [Mittal 21] Mittal, G., et al.: Symbolic music generation with diffusion models (2021), arXiv:2103.16091
- [Nakamura 19] Nakamura, E., et al.: Unsupervised melody style conversion, in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 196–200 (2019)
- [Nierhaus 09] Nierhaus, G.: *Algorithmic Composition: Paradigms of Automated Music Generation*, Springer Science & Business Media (2009)
- [Pachet 11] Pachet, F. and Roy, P.: Markov constraints: steerable generation of Markov sequences, *Constraints*, Vol. 16, No. 2, pp. 148–172 (2011)
- [Singh 22] Singh, R. and Nakamura, E.: Dynamic cluster structure and predictive modelling of music creation style distributions, *Royal Society Open Science*, Vol. 9, No. 220516 (2022)
- [Sturm 16] Sturm, B. L., et al.: Music transcription modelling and composition using deep learning, in *Proc. Conference on Computer Simulation of Musical Creativity* (2016)
- [Youngblood 19] Youngblood, M.: Conformity bias in the cultural transmission of music sampling traditions, *Royal Society Open Science*, Vol. 6, No. 191149 (2019)