

自動作曲を用いた進化実験による 音楽スタイルのクラスター形成過程の分析

中村 栄太^{1,a)} 金子 仁美² 伊藤 貴之³ 金子 邦彦⁴

概要: 本発表では、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の初期結果について報告する。音楽文化の発展には、創作者と鑑賞者の両方の役割が重要だと考えられる。音楽の進化における鑑賞者の働きを定量的に調べるため、世代交代を続ける創作者集団を模した自動作曲エンジンの集合と、それにより作られた曲をその都度評価する多数の鑑賞者からなる実験系を構成し、作られる音楽の特徴および自動作曲エンジンの系統構造の時間変化を調べた。生成曲の評価値と音楽スタイルのクラスター構造の進化、および鑑賞者による評価の傾向の違いについての分析結果を報告する。音楽スタイルの空間においてクラスター構造の形成とクラスターの変形が観察された。メロディーの生成モデルよりも和声に関するモデルに対してより強い選択が働き、そこには鑑賞者の音楽経験の度合いによる評価の傾向の違いが現れることもわかった。鑑賞者の評価のみにより新しいスタイルの音楽が生まれ続ける可能性およびスタイルのクラスターが共存する構造が鑑賞者の多様な嗜好性によるものである可能性が示唆された。

1. はじめに

世界のほぼ全ての民族・社会には音楽文化があり、それぞれの音楽は多様でかつ共通性を持っている [1]。音楽の時代変化の仕組みの理解は、こうした音楽文化の性質の起源を説明し、また文化を発展させる人間の一般的な能力の解明に重要である [2]。クラシック音楽やポピュラー音楽など公共性を持つ音楽の発展は、創作者に焦点を当てて議論されることも多いが [3]、音楽を鑑賞し創作者を経済的に支援する鑑賞者（あるいは消費者）の役割も重要である [4]。世代を超えて受け継がれる文化のモデルとして、遺伝子や生態系の進化と同様の neo-Darwin 的なモデルを考えれば [5]、創作者は音楽創作スタイルという形質を伝達する個体、鑑賞者は自然選択に寄与する環境に対応させて考えることができる。モデルに基づく進化過程の理解には、この両者の働きの定量的な分析が欠かせない。

音楽進化における鑑賞者の役割を調べるため、MacCallum ら [6] は自己複製する作曲アルゴリズムにより創作者集団を模倣して、生成された音楽を多数の鑑賞者が評価する音楽進化実験を行った。作曲アルゴリズムはループ音源を生成するプログラミングコードにより表され、遺伝的ア

ルゴリズム [7] と同様な方法で、ランダムな変異を含みながら世代交代をする。鑑賞者による曲の評価値を適応度として、世代交代時に親となる作曲アルゴリズムが選択される。実験では、ノイズに近い初期状態から比較的音乐性が高いループ音源を作る作曲アルゴリズムが現れることが確認された。この結果は、鑑賞者の評価があれば、ランダムに変異する創作者集団において音楽性を向上させる進化が起こることを示している。また、この進化実験の方法では、進化の再現性の検証や、選択係数などのパラメーターを制御して実験を行えることが利点である。

一方で、大規模なデータ解析により、クラシック音楽やポピュラー音楽の特徴の時代変化が定量的に調べられている [8–11]。その中で、比較的時間スケールが小さいこれらの音楽の進化では、ジャンルなどに関連する音楽スタイルのクラスター構造の変化がその大きな要因になっていることが分かってきている [8,9]。つまり、一つの社会の音楽にはスタイルの異なるクラスターが複数共存しており、クラスターの形成や消滅が音楽の特徴に大きな影響を与える。これは、音楽スタイルの空間には局所的に最適な点を多数持つ構造があることを示唆している。また、これらの音楽では、創作者間で伝達されるのは曲そのものではなく、創作に関する知識あるいは創作スタイルである。自動作曲および認知科学の研究で、音楽創作知識の学習や音楽スタイルの認識は統計学習を通して行われることが示唆されている [12–14]。ここで、創作者の統計学習を通じた音楽スタ

¹ 京都大学 Kyoto 606-8501, Japan

² 桐朋学園大学 Tokyo 182-0021, Japan

³ お茶の水女子大学 Tokyo 112-8610, Japan

⁴ 東京大学 Tokyo 153-8902, Japan

^{a)} eita.nakamura@i.kyoto-u.ac.jp

イルの伝達過程と鑑賞者の評価がいかにクラスター構造に影響を与えるかが疑問となる。

本研究では、統計学習に基づく自動作曲を用いた進化実験により、現実の社会に近い音楽の進化を調べる方法を提案する。枠組みとしては上の MacCallum らの実験に近く、作曲アルゴリズムとして現実の音楽に近いものを生成できる作曲モデルを用いる。この作曲モデルは現実の創作者と同じ様に、多数の曲を状況に応じて生成できる。本稿では、この実験方法について説明して、実験の初期結果を報告し、また今後の展望について議論する。今回は、与えられたコード進行と伴奏に対してメロディーを生成する作曲モデルの集団の進化を分析する。特に、メロディーのスタイルと和音との関係性のスタイルのクラスター構造、生成曲の評価値の進化曲線、そして傾向の異なる鑑賞者がクラスター形成に与える影響について調べる。進化実験は現在進行中で継続的に実施する予定である。Web サイト <https://creevomusic.github.io/exp/> にて一般公開しており、誰でも鑑賞者として参加可能である。

2. 方法

進化実験では、自動作曲モデル（以下、バーチャル作曲家と呼ぶ）の集団を作り、それらが作った曲を多数の人が評価して、その結果に基づいてバーチャル作曲家集団の世代更新を行う。以下、作曲モデル、世代更新の方法、評価の方法について順に述べる。

2.1 作曲モデル（バーチャル作曲家）

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。ここで、コード進行には小節数と各コードの長さの情報を含むものとする。歌詞は、小節ごとに平仮名表記の読みがなで与えられるものとして、各言語音に1音符を対応させるものとする。つまり、作曲システムとしては、入力のコード進行に対してメロディー音符列を出力する。

自動作曲の具体的なモデルとして、ここではトーラスマルコフモデル [15] を用いる。このモデルでは、メロディーの各音符 n は半音単位の整数で表される音高 p_n と小節内でのビート位置（拍節位置）で表される発音時刻 b_n のペアで記述され、それらの遷移確率 $P(p_n, b_n | p_{n-1}, b_{n-1})$ がモデルパラメーターである。また、メロディーとコード進行の関係性をコード記号 C_m に条件付いた音高の確率 $P(p_n | C_m)$ により表す（これを和声作用モデルと呼ぶ）。実験では、音高クラスを3オクターブ周期に拡張した拡張音高クラスを音高変数として用いて、拍節位置は1小節を48分割した単位を用いる。これらの確率によりコード進行 $C = (C_m)$ に対するメロディー $X = (p_n, b_n)$ の確率は

$$P(X|C) \propto \prod_n [P(p_n, b_n | p_{n-1}, b_{n-1})]^\beta [P(p_n | C_m)]^\alpha$$

と表される。ここで β と α は、それぞれメロディー遷移確率と和声作用モデルを重み付ける係数である。実際には、メロディーは上式の確率に基づいたサンプリングにより生成するが、 β が大きいと確率がより大きなメロディーがサンプルされる確率がさらに高くなるため、より典型的なメロディーが生成される。一方で α が大きい時は、コード進行との関係をより強く反映したメロディーが生成される。メロディーのスタイルとして同音連打の頻度や跳躍進行の頻度も重要と考えられるため、それぞれの頻度を制御するパラメーターも導入する。具体的には、上記の式で同音連打の確率に対して $e^{-d_{\text{same}}}$ の重みを与える同音連打減衰係数 d_{same} と3半音以上の跳躍音程に対して $e^{-d_{\text{step}}}$ の重みを与える跳躍音程減衰係数 d_{step} を用いる。統計パラメーターであるメロディー遷移確率と和声作用モデルは、コード進行付きのメロディーデータから学習可能である。

このモデルは、メロディーのスタイルを精度良く記述可能であり、かつパラメーターが解釈しやすく、世代更新やクラスター分析も容易に行える利点がある。例えば、作曲システム Orpheus [16] などを用いられている通常のマルコフモデルでは音高とリズムの遷移確率は独立であるため、リズムと音高の関係性を捉えられる上のモデルはメロディーのスタイルをより忠実に表現できる。この他、LSTM などの DNN を用いる手法があるが [17–19]、DNN のパラメーターは音符から計算される統計量と単純には関係付かず、学習ごとにパラメーター値が大きく変わるため、下で考えるようなパラメーターの伝達を単純には行えない。

次節で説明するように、各バーチャル作曲家は統計パラメーターの伝達に関するもう一つのパラメーター u を持つ（斜行伝達係数と呼ぶ）。以上をまとめると、バーチャル作曲家が持つパラメーター（形質）は、以下の通りである。

- 統計パラメーター：遷移確率 $\pi(p', b'; p, b)$ と和声作用モデル $\chi(p; C)$
- 生成制御パラメーター：典型度 β 、和声重要度 α 、同音連打減衰係数 d_{same} 、跳躍音程減衰係数 d_{step}
- 伝達制御パラメーター：斜行伝達係数 u

以降、統計パラメーターのことを統計的形質、生成制御パラメーターと伝達制御パラメーターを合わせて非統計的形質と呼ぶ。統計パラメーターの数は、遷移確率が約 $3 \cdot 10^6$ （ただし多くのものは0で実質的なパラメーター数は 10^5 程度）、和声作用モデルが60である。

2.2 世代更新の方法

各世代において、前節のモデルで表されるバーチャル作曲家 N 人からなる集団を作る（今回の実験では $N = 50$ としている）。世代更新の際には、次節で具体的に定義する適応度 w_n が各バーチャル作曲家について計算される。これを用いて次の世代でのバーチャル作曲家が以下の手続きで構成される。まず、 N 人の中で、適応度が上位の $N/2$ 人

を選び、それらを次の世代のメンバーとする。残りの $N/2$ は次の世代には含めない。次の世代の残りの $N/2$ 人は、新しいバーチャル作曲家として集団に導入される。導入されたバーチャル作曲家は適応度を規格化した確率 $P(n) \propto w_n$ により親（師匠）となるバーチャル作曲家を前の世代の集団から選択する。子（弟子）のバーチャル作曲家はそのパラメーターを親から受け継ぐが、生成制御パラメーターに関しては、一様分布に従う変異を含める。例えば、典型度 β は $\beta' \sim \beta + U(-0.2, 0.2)$ とする。実験では変異の誤差の大きさは、典型度、和声重要度、同音連打減衰係数、跳躍音程減衰係数のいずれも 0.2 であり、値域はこの順にそれぞれ $[0, 5]$, $[0, 5]$, $[0, 5]$, $[-3, 3]$, $[-3, 3]$ とした。

統計パラメーターの伝達に関しては、親から子に十分の量の生成例が示されたと仮定し、統計学習による統計誤差は 0 とおく。一方で、統計パラメーターの学習には親世代の親以外のバーチャル作曲家（二次的親）の作品も学習データに使えるとする。具体的には、斜行伝達係数 u を $0 \leq u \leq 1/2$ を満たす定数として、親のパラメーター θ_0 と K 人の二次的親のパラメーター θ_k ($k = 1, \dots, K$) から、 $\theta = (1-u)\theta_0 + u\theta_k/K$ とパラメーターが学習される。実験では $K = 2$ としている。一方、 u 自体は生成制御パラメーターと同様に、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 ($u' \sim u + U(-0.05, 0.05)$)。これにより、子の統計パラメーター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。

最初の世代のバーチャル作曲家の形質は以下のように初期化する。統計パラメーターは既存曲のメロディー・コード進行データからの学習を用いて設定する。初期学習には、日本のポピュラー音楽および欧米ポピュラー音楽やジャズの計約 1 万曲のデータを用いた（この中には Wikifonia データ [20] も含まれる）。まず、このデータに対して混合マルコフモデルを用いてクラスタリングを行う。トーラスマルコフモデルはパラメーター数が非常に大きく、大きな初期値依存性により直接クラスタリングを行うことが難しいため、拡張音高クラスの遷移確率と拍節位置の遷移確率についてそれぞれ 10 クラスのクラスタリングを行った後に、その組み合わせを初期値として両方の遷移確率の積についてクラスタリングを行い、最終的に混合確率の大きな 25 クラスを選択する。得られた 25 クラスのデータからメロディー遷移確率と和声作用モデルのパラメーターを学習することで、原始バーチャル作曲家 25 人を作る。第 1 世代の 50 人のバーチャル作曲家の統計パラメーター θ は、各々に対してランダムに抽出した 3 人の原始バーチャル作曲家 i, j, k の平均 $\theta = (\theta_i + \theta_j + \theta_k)/3$ とおく。この際、全ての原始バーチャル作曲家が最低 2 回抽出されるようにする。生成制御パラメーターと伝達制御パラメーターは、以下の様に一様分布からサンプルする： $\beta \sim U(0.7, 1.2)$, $\alpha \sim U(0.7, 1.2)$, $d_{\text{same}} \sim U(0, 1)$, $d_{\text{step}} \sim U(0, 1)$, $u \sim U(0.1, 0.3)$ 。

2.3 評価の方法

以下の方法で、バーチャル作曲家が生成した曲を人が聞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。各世代においてバーチャル作曲家に対する課題となる、歌詞付きのコード進行を M 個用意する。今回の実験では、予め用意した約 3 千の 4 小節のコード進行から各世代で 20 個をランダムに抽出する（各曲の時間長は 10 秒から 15 秒である）。各バーチャル作曲家はこれらのコード進行に対してメロディーを生成する。得られた $NM = 1000$ 個の曲を評価に用いる。この際、コード進行から自動で作ったギター・ベース・ドラムの伴奏音を付加して、メロディーは歌声合成システム Sinsy [21] により生成した音源を用いる。歌詞は全て「ららら」で歌い、伴奏音はコード進行ごとに異なるが、全てのバーチャル作曲家に対して同一のものをを用いており、バーチャル作曲家同士の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聴き比べの形式で行う。毎回、ランダムに抽出したコード進行に対して、ランダムに抽出した 2 人のバーチャル作曲家の曲を対にして提示する。これに対して、鑑賞者自身が良いと思う方を選択してもらう。この評価方法（A/B テストとも呼ばれる）は、MacCallum らの実験 [6] で用いた点数付けによる評価に比べて、評価が容易であり人による評価スケール尺度依存性を取り除ける点で有利である。また、現実の社会における曲の評価も売り上げ額や動画再生回数など、比較選択に基づく消費・鑑賞活動の結果として観測されることが多いことから、採点評価よりも比較評価の方が現実に近い評価法だと考えられる。以下では、各曲および各バーチャル作曲家に対して、良いと選択された回数を「売上数」と呼ぶことにする。つまり 1 回の聴き比べの結果では、良いと選択された方のバーチャル作曲家は売上数が 1 だけ増え、もう一方のバーチャル作曲家の売上数は変化しない。

1 世代の評価を終えた時のバーチャル作曲家 n の総売上数を s_n 、そのバーチャル作曲家の曲が聴き比べされた総回数を c_n と記す。バーチャル作曲家の適応度は $w_n = \exp(\beta_s s_n / c_n)$ と定める。ここで、 β_s は選択の強さを決める係数であり、規格化された売上数 s_n / c_n を以下では売上率と呼ぶ。適応度の定義に売上数 s_n ではなく売上率を用いるのは、聴き比べに参加した回数 c_n に関する統計揺らぎの効果を取り消すためである。定義より、売上率の平均は $1/2$ であり、これよりも大きな売上率を持つバーチャル作曲家は総合的に見て平均より高く評価されたことになる。前節の世代更新の方法では、これらのバーチャル作曲家は次世代に残ることになる。

評価はインターネット上のサイトで行っている。聴き比べをする際には、鑑賞者には年代、音楽を聴く時間、創作（演奏および作曲）の経験、和声法の学習経験についてのアンケートに答えてもらう。評価は匿名で行い、鑑賞者に関

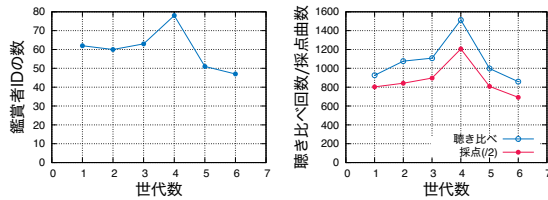


図 1 世代ごとの鑑賞者 ID の数と聴き比べ回数および採点された曲数。

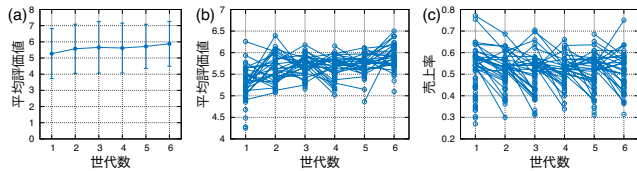


図 2 (a) 平均評価値の進化と各世代での標準偏差。 (b) バーチャル作曲家ごとの平均評価値。 (c) バーチャル作曲家ごとの売上率。 (b) と (c) では同一のバーチャル作曲家が 2 世代続いて現れる場合に線で結んで表示している。

するこれら以外の情報は収集しないが、一連の聴き比べの結果が同じ鑑賞者かどうかの判定はできるように ID を付与してデータを収集する。各鑑賞者についての聴き比べ回数は任意としているが、適応度の計算および分析に用いるデータでは最大 50 回までの結果を用いて、特定の鑑賞者の評価が全体に大きく影響することを避けている。

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・大学院生が鑑賞者に含まれている。これらの中には音楽大学で作曲や演奏の専門教育を受けている学生も含まれている。これらの参加者には、聞き比べの際に適応度の計算に用いる A/B テストの結果に加えて、各曲に対して 10 段階の採点評価も行うように指示している。

3. 結果

以下では 2021 年 8 月 1 日から行っている進化実験の初期分析結果について報告する。現在までに 7 世代（評価は 6 世代）に渡るデータが得られている。実験では選択係数の値は $\beta_s = 3$ とした。

3.1 評価値と非統計的形質の進化

図 1 に示すように、鑑賞者 ID の数は平均 60、聴き比べ回数は平均 1,080、採点された曲数は平均 1,750 で遷移している。これから各世代の各バーチャル作曲者は平均約 40 回聴き比べされ、仮に良い方の曲がランダムに選ばれるとすると、売上率の標準偏差は 0.08 程度と見積られる。

図 2(a) に平均評価値の進化を示す。初めの世代更新で平均値が大きく上昇した後、第 2 世代から第 4 世代では大きく変化せず、第 5 世代以降は再び上昇していることがわかる。また標準偏差は世代を経るごとに小さくなっている。図 2(b) に示すバーチャル作曲家ごとの平均評価値でも同様の傾向が見られる。図 2(c) では、バーチャル作曲家ご

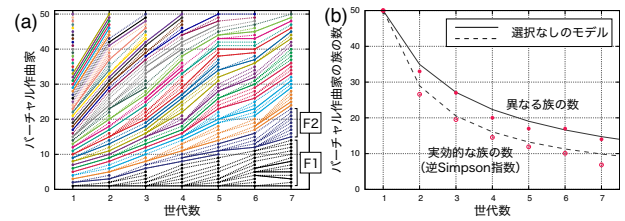


図 3 (a) バーチャル作曲家の系統樹。同一のバーチャル作曲家は実線で、親子関係は点線で示す。先祖を共有するバーチャル作曲家は同じ色で表わされている（これらを族と呼ぶ）。(b) バーチャル作曲家の族の変化。実線と点線は選択圧がないと仮定した場合の Galton-Watson モデルの期待値の遷移を示す。

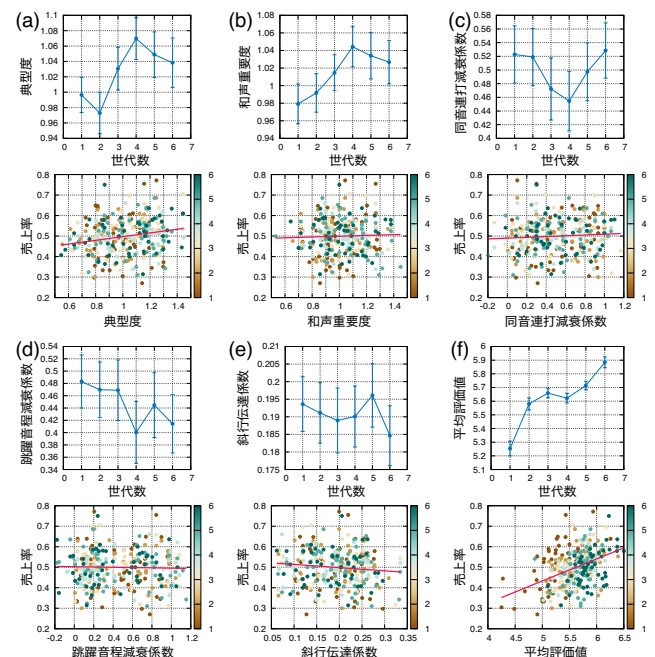


図 4 非統計的形質の進化と売上率との相関。誤差棒は標準誤差を表す。下側のパネルでは各点はバーチャル作曲者を表し、色は世代を表している。(f) は平均評価値と売上率の相関を示す。相関係数の値と p 値は、(a) $\rho = 0.20$ ($p = 5 \cdot 10^{-3}$), (b) $\rho = 0.036$ ($p = 0.53$), (c) $\rho = 0.066$ ($p = 0.25$), (d) $\rho = -0.027$ ($p = 0.64$), (e) $\rho = -0.10$ ($p = 0.08$), (f) $\rho = 0.42$ ($p < 10^{-5}$).

との売上率の標準偏差が第 2 世代以降は統計揺らぎの大きさ 0.08 程度であること、そして世代が変わる際に同じバーチャル作曲家の売上率が大きく変わることがある様子が確認できる。第 2 世代以降で売上率の分散が最も大きい第 3 世代でも、F 検定の p 値は 0.21 と大きい。売上率と平均評価値に相関があることは、図 4(f) で確認できる。これから、聴き比べによる売上率には比較的大きな統計揺らぎが含まれ、世代単位では弱い選択圧が働いているが、この選択の効果で平均評価値が徐々に上昇したことがわかる。

バーチャル作曲家の系統樹を図 3(a) に、族単位の多様性の変化を図 3(b) に示す。ここで、先祖を共有するバーチャル作曲家の集合を族と定義している。族ごとの大きさには偏りが生まれ、第 7 世代では、最大の族だけで集団の

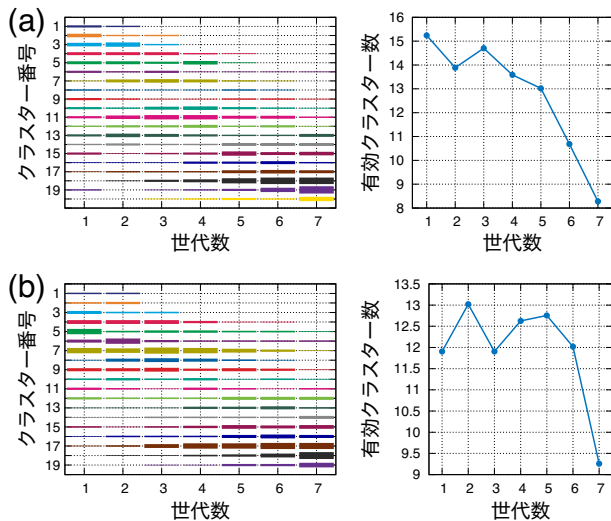


図5 クラスタ頻度の変化。(a) はメロディー遷移確率, (b) は和声作用モデルに関する結果。左側は各世代における各クラスターに属するバーチャル作曲家の数を箱の高さで表している。右側は逆 Simpson 指数による有効クラスター数を示している。

約 1/3, 上位 3 つの族で集団の半数以上を占めていることがわかる。図 3(b) の逆 Simpson 指数においても同様の傾向が見られる。逆 Simpson 指数は多様性の尺度で, 各族 i の割合を p_i ($\sum_i p_i = 1$) とする時, $(\sum_i p_i^2)^{-1}$ で定義される実効的な族の数を表す。この様な族の多様性の減少は, 有限サイズの集団では選択圧がない場合でも起こるものである。族の多様性の進化のモデルである Galton-Watson 過程 [5] を今回の世代更新過程に当てはめたモデルにより, 選択圧がない場合の族の多様性の期待値の変化を計算した結果が 3(b) に示されている。逆 Simpson 指数の減少の度合いはこのモデルの予測値よりも小さいが, その差は小さくなく, ここでも選択圧の効果は比較的小さいことが確認できる。

図 4 に各非統計的形質の進化と売上率との相関を示す。今回の実験の範囲で売上率と比較的大きな相関が見られた非統計的形質は, 典型度と斜行伝達係数でこれらは相関がある方向に平均値が変化したことがわかる。典型度が小さい場合は, メロディーに音階から外れた音符が含まれる確率が増すため, 負の選択が働いたと考えられる。ただし, これまでのデータではこれらの形質の分布はそれほど大きくは変化しておらず, 今後より長い時間での変化を観測する必要がある。

3.2 音楽スタイルのクラスター構造の進化

バーチャル作曲家の統計的形質を調べ, 音楽スタイルがどう進化したかを分析する。まず, メロディー遷移確率に注目して, クラスタ構造を調べる。第 7 世代までに現れたバーチャル作曲家のメロディー遷移確率を混合マルコフモデルに基づいてクラスタリングを行った結果を図 5(a) に示す。ここでは, 細かい構造まで捉えられるようにクラス

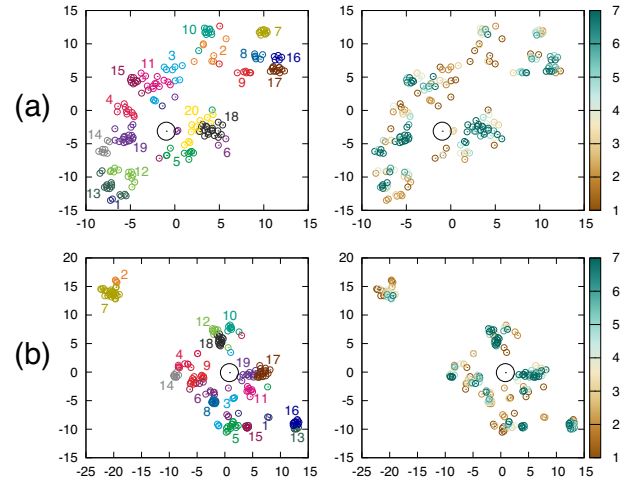


図6 対称 KL ダイバージェンスを距離として t-SNE により得られた 2 次元空間の中でのバーチャル作曲家の統計的形質の分布。(a) はメロディー遷移確率, (b) は和声作用モデルに関する結果。左側の色と数字は図 5 のクラスター番号と対応している。右側では色は各バーチャル作曲家が現れた一番最近の世代数を表す。各パネルで中心付近にある大きな円は, 進化実験の初期化に用いた全ての学習データで学習した平均モデルを示している。

ター数を 20 としている。またクラスター番号は, クラスタに属するバーチャル作曲家の世代の平均に対する順序を表している。初期にはクラスターの頻度分布が広がっていたのが, 世代を経るに従って分布が偏り, 実効的なクラスター数が減少していることがわかる。また, 世代ごとに頻度が高いクラスターが変化していることも確認できる。図 6(a) はメロディー遷移確率を 2 次元特徴量空間に埋め込んだ際の, バーチャル作曲家の距離関係を表している。メロディー遷移確率の対称 KL ダイバージェンスを距離として t-SNE を用いた結果であり, バーチャル作曲家間の距離をなるべく保つように 2 次元に配置している。初期には遷移確率の空間で広がって分布していたが, 徐々に少数のクラスターに分布が偏り, クラスタ構造が明確になっている様子が確認できる。

第 7 世代において大きな頻度を持つ第 18, 19, 20 番目のクラスターを詳しく見る。クラスター 18 と 20 は, いずれも初期には存在せず, 第 2 世代以降に発生して成長している。また, 図 6(a) ではこれらのクラスターは近接あるいはオーバーラップしており, 見方によっては 1 つのクラスターと見なせる。系統を見ると, クラスタ 18 の初めのバーチャル作曲家 (ID 105) は, クラスタ 6 に属する族 F1 の祖先のバーチャル作曲家 (ID 91) から派生しており, 大きく離れたクラスター 1 と 7 にそれぞれ属するバーチャル作曲家を二次的親に持つことがわかった。クラスター 20 の初めのバーチャル作曲家 (ID 134) は, この ID 105 のバーチャル作曲家を親として, クラスタ 15 とクラスター 17 に属するバーチャル作曲家を二次的親として持つ。ID

105 および ID 134 のバーチャル作曲家はいずれも $u > 0.2$ の大きな斜行伝達係数を持ち、その統計的形質を親から大きく変化させて新しいクラスターの初期メンバーとなっている。クラスター 18 と 20 のメンバーは、主に族 F1 に属する。この様にクラスター 18 と 20 は、族 F1 の祖先から斜行伝達によりその中心位置を移動させながら広がって形成された。

クラスター 19 は、第 2 世代で一旦消えて第 3 世代から再び成長しているが、系統を見ると、このクラスターに属する第 1 世代と第 3 世代のメンバーは直接的な系統でつながっていない。つまり、最近まで残っているこのクラスターは第 3 世代で発生している。このクラスターの初めのメンバー (ID 135) はクラスター 14 に属する族 F2 の祖先のバーチャル作曲家 (ID 58) を親として持ち、クラスター 2 と 6 に属するバーチャル作曲家を二次的親として持つ。このバーチャル作曲家の斜行伝達係数も $u = 0.21$ と大きい。クラスター 19 のメンバーは主に族 F2 に属する。以上の様に、第 7 世代で最も発達しているクラスターはいずれも、その祖先から大きく移動しながら形成されていることがわかった。

図 5(b) と 6(b) に示すように、和声作用モデルについても世代を経るにつれてクラスター構造がより明確になっている。和声作用モデルについては、空のクラスターが 1 つあったため、クラスター数は 19 になっている。クラスター 17 と 19 は、メロディー遷移確率のクラスター 18 と 20 と類似した関係にある。

統計的形質に関するクラスターと売上率 (複数世代存在したバーチャル作曲家に対しては時間平均をとったもの) との関係調べるため、分散分析を行った。メロディー遷移確率に対しては $F = 1.28$ ($p = 0.20$)、和声作用モデルに対しては $F = 1.65$ ($p = 0.05$) であった。統計的な有意度は高くないが、和声作用モデルに比較的強く選択が働いたことが示唆され、メロディー遷移確率に関しては和声モデルの形質から誘発されたバイアス [22] によってクラスター化が促進された可能性も考えられる。

これまでのところ、メロディー遷移確率と和声作用モデルの両方について、平均モデルに近いところが適応度が高いといった単調な傾向は見られず、互いに大きく離れたクラスターが共存する構造が見えている。こうした複数クラスターが共存する構造が長期的に維持されるのか、また各クラスターの位置が方向性を持って移動を続けるのかなど、クラスター構造の長期的な振る舞いが今後の観測により明らかになると期待される。

3.3 鑑賞者の傾向

鑑賞者による評価の傾向の違いを調べるため、鑑賞者 ID ごとに一連の評価結果における各クラスターでの平均評価値と平均購買率を求め、因子分析を行った。この結果とア

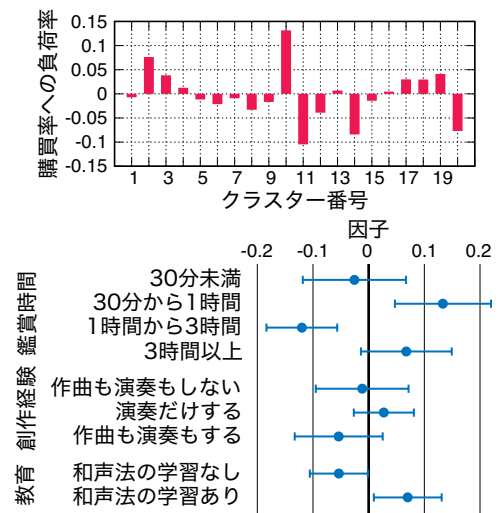


図 7 メロディー遷移確率のクラスターごとの購買率に対する因子分析の結果。下の図は、各グループに関する因子の平均と標準誤差を示している。

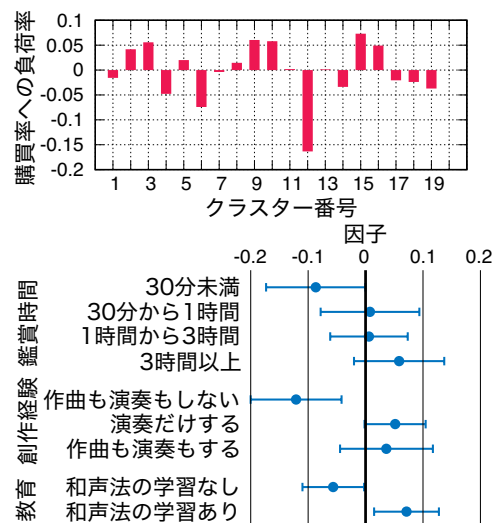


図 8 和声作用モデルのクラスターごとの購買率に対する因子分析の結果。下の図は、各グループに関する因子の平均と標準誤差を示している。

ンケート調査により得られる鑑賞者の属性との関連を調べた。なお、今回の実験では鑑賞者の年代は 10 から 20 才台に集中していたため、年齢についての解析は行っていない。アンケート調査では、1 日当たりの平均音楽鑑賞時間 (一般的な音楽に関するもの) に関して「30 分未満」「30 分から 1 時間」「1 時間から 3 時間」「3 時間以上」の 4 項目、創作経験に関して「作曲も演奏もしない」「演奏だけする」「作曲だけする」「作曲も演奏もする」の 4 項目、和声法の学習経験に関しては「(授業などで和声法の) 学習なし」「学習あり」の 2 項目で答えてもらった。創作経験に関しては、「作曲だけする」と答えた人はいなかったため、以下この項目は除いて考える。

まず、クラスターごとの平均評価値に対する因子分析の結果では、第 1 因子に全てのクラスターに同じ方向に働く

バイアスが見られた。属性では、音楽鑑賞時間が「1 時間から 3 時間」と答えた人のグループ、創作経験では「作曲も演奏もする」と答えた人のグループに関してそれぞれ、全体的な平均評価値の値を低くする顕著なバイアスが見られた。第 2 因子では、クラスターごとに正負の異なる負荷率が見られたが、鑑賞者の属性に関して大きな差異は見られなかった。

次に、購買率に対する因子分析の結果を図 7 と 8 に示す。メロディー遷移確率と和声作用モデルのクラスターに関する結果の両方で、第 1 因子にクラスターごとに正負の異なる負荷率が見られる。なお、BIC による基準では、両方のケースにおいて 1 つだけの因子を持つモデルのフィットが最も良かったため、第 1 因子に注目する。図 7 のメロディー遷移確率に関する結果では、鑑賞時間と学習経験の中で優位な差があり、鑑賞者の属性によって平均的に高い評価を与えるクラスターが異なることがわかる。ただし、鑑賞時間に関しては平均値の分布が単調ではなく、解釈は難しい。図 8 の和声作用モデルに関する結果では、創作経験と学習経験の中で優位な差があり、また鑑賞時間についても単調な傾向があり、和声に関する評価の傾向には音楽経験の度合いが影響することがわかった。両方の統計的形質に関して、第 7 世代で残っているクラスターの中で、鑑賞者によって平均的な評価が割れるものが混ざっていることも確認できる。

以上の結果より、鑑賞者の音楽経験の度合いにより評価の傾向に差が見られることが確かめられた。これはクラスターごとにその音楽を高く評価する鑑賞者の集団の特徴が異なることを意味し、複数のクラスターの共存の背後には鑑賞者ごとの音楽の嗜好性の違いが影響している可能性がある。一方で、この結果はそれぞれの属性を持つグループの平均に関する有意性を見ており、因子空間でこれらのグループの分布が明確なクラスター構造を持っている訳ではない。つまり、鑑賞者個人レベルでの評価の傾向のばらつきの方が、属性ごとのばらつきよりも大きいことも確認できている。また実験方法に関する考察として、採点評価と比較評価を比べると、前者は鑑賞者による全体的なバイアスが大きいため、適応度の計算などに用いる際には注意が必要であることも確かめられた。

4. 議論

本研究では、自動作曲を用いた進化実験の初期データの分析を行い、多くの鑑賞者の評価による選択と斜行伝達を含むバーチャル作曲家の複製過程の繰り返しにより、音楽スタイルの空間においてクラスター構造の形成とクラスターの変形が観察され、そしてその背後には鑑賞者による評価の傾向の差異が影響する兆候が見られた。また、バーチャル作曲家の世代交代が繰り返される中で、そのスタイルは祖先のものとは変化しており、平均的な評価値は上昇

することも確認された。詳しい解析にはより長期間のデータが必要であるが、これらの結果は音楽進化における創作者と鑑賞者/消費者の働きを定量的に調べる上で実験の有用性を示唆している。以下、結果の考察と今後の展望について述べる。

まず、特に和声作用モデルのクラスターと売上率の間に有意な相関が見られたことと、成長しているクラスターの中にはスタイルの変異により位置が移動しているものがあることから、音楽スタイルの進化には鑑賞者の評価が大きく影響を与える可能性が示された。これは、ポピュラー音楽など現代社会の音楽の進化において、音楽が多様化しながら発達する過程の中でも鑑賞者が重要な役割を持つことを示唆する。また、鑑賞者による評価の傾向の違いの存在は、現代の多くの社会で見られる異なる音楽スタイルのクラスターの共存が鑑賞者による嗜好傾向の違いに起因する可能性も示している。今回の実験では評価の傾向の違いの要因として音楽経験の度合いがあることがわかったが、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する。現実社会では音楽創作者は少数であるため、売上高などの音楽の評価値としては鑑賞・消費する側の影響より重要であると考えられる一方で、創作に関わる形質を受け継ぐのは創作者であるため、形質の伝達に関する適応度は売上高の単純な関数にはなっていない可能性も考えられる。これは今後、自動作曲技術の普及などにより音楽がいかに進化するかを考える上で重要な問題である。

今回は統計的形質について、正確な伝達と混合継承タイプの斜行伝達のみを考えたが、より現実的なタイプの変異過程を考える必要がある。精密な自動作曲モデルは多数のパラメータを持つため、この問題は高次元空間における探索の側面も持つ。VAE などの次元圧縮法を組み合わせる方法の利用や、音高とリズムなど、異なる音楽要素に関する形質の部分空間上での組み替えも重要だと考えられる。本稿では、バーチャル作曲家の形質の進化に注目したが、生成曲の特徴に関してもクラスター構造の解析もできる。現実の音楽では作曲家の形質は観測が難しいため、これは現実データとの照合の面においても重要である。

この進化実験を繰り返すことにより、鑑賞者による評価としての音楽の適応度地形がどれだけ複雑な構造を持つか、そしてこの構造の静的部分と動的部分（頻度依存バイアスなど）について調べられるであろう。これは文化における進化の再現性の問題、そして芸術における創作スタイルの探索空間の広さを調べる上で有用な知見を与えられる。また、新しい音楽スタイルの自動発見などの技術につながる可能性もある。

謝辞 本研究は、日本学術振興会科研費 19K20340 および 2021 年度京都大学リサーチ・ディベロップメントプログラム【いしづえ】から支援を受けた。実験に協力頂いた皆様に感謝する。

参考文献

- [1] P. E. Savage, S. Brown, E. Sakai, and T. E. Currie, PNAS 112, 8987 (2015).
- [2] N. L. Wallin, B. Merker, and S. Brown (eds), The Origins of Music (MIT press, 2000).
- [3] H. Goodall, The Story of Music (Chatto & Windus, 2013).
- [4] K. Negus, Popular Music Theory (Polity, 1996).
- [5] L. L. Cavalli-Sforza and M. W. Feldman, Cultural Transmission and Evolution (Princeton University Press, 1981).
- [6] R. M. MacCallum, M. Mauch, A. Burt, and A. M. Leroi, PNAS 109(30), 12081 (2012).
- [7] J. H. Holland, Scientific American 267(1), 66 (1992).
- [8] P. H. R. Zivic, F. Shifres, and G. A. Cecchi, Proc. Natl. Acad. Sci. 110(24), 10034 (2013).
- [9] M. Mauch, R. M. MacCallum, M. Levy, and A. M. Leroi, Royal Society Open Science 2, 150081 (2015).
- [10] C. Weiß, M. Mauch, S. Dixon, and M. Müller, Musicae Scientiae 23(4), 486 (2019).
- [11] E. Nakamura and K. Kaneko, Scientific Reports 9, 15993 (2019).
- [12] J. R. Saffran, R. N. Aslin, and E. L. Newport, Science 274, 1926 (1996).
- [13] D. Huron, Sweet Anticipation: Music and the Psychology of Expectation, MIT press (2008).
- [14] J. D. Fernández and F. Vico, Journal of Artificial Intelligence Research 48, 513 (2013).
- [15] E. Nakamura, K. Shibata, R. Nishikimi, K. Yoshii, Proc. ICASSP, 196 (2019).
- [16] 深山覚, 中妻啓, 酒向慎司, 西本卓也, 小野順貴, 嵯峨山茂樹, 情報処理学会論文誌 54(5), 1709 (2013).
- [17] B. L. Sturm, J. F. Santos, O. Ben-Tal, and I. Korshunova, in Proceedings of the 1st Conference on Computer Simulation of Musical Creativity (2016).
- [18] H. Tsushima, E. Nakamura, K. Itoyama, K. Yoshii, Proc. ISMIR, 145 (2018).
- [19] K. Chen, G. Xia, S. Dubnov, Proc. International Conference on Semantic Computing, 128 (2020).
- [20] F. Simonetta, F. Carnovalini, N. Orio, and A. Rodà, Proc. Audio Mostly 2018 on Sound in Immersion and Emotion, 26 (2018).
- [21] Y. Hono et al., Proc. APSIPA ASC, 1003 (2018).
- [22] R. Boyd and P. J. Richerson, Culture and the Evolutionary Process (The University of Chicago Press, 1985).