

# 作曲スタイル分布の動的クラスター構造の分析と予測

中村 栄太<sup>1,a)</sup> Rajsuryan Singh<sup>2</sup>

**概要:** 本稿では、音楽文化の発展の性質を調べるため、作曲スタイル分布の動態について分析する。日本のポピュラー音楽のデータから抽出したメロディーの統計量と、統計モデルに基づく手法を用いて、クラスター構造の動態を特徴付けるための分析を行い、作曲スタイル分布の時間発展を予測するための適応度に基づく進化モデルを構築する。分析により、クラスターの相対頻度に現れるクラスター間のダイナミクスに加えて、クラスターの収縮や中心の移動など、クラスター内のダイナミクスが顕著な動的モードとして確認された。また、これらの動的モードを組み込んだ進化モデルが未来の作曲スタイル分布を予測するために有効であり、クラスターの頻度と分散の予測がしばしば同程度に重要であることが示された。これらの結果は、従来研究では見過ごされることが多かったクラスター内ダイナミクスが、音楽文化の発展を理解する上で重要であることを示唆している。

## 1. はじめに

音楽の特徴がどのように時代変化しているかという問題は、多くの人の興味を引き付け、様々な観点から研究されてきた。伝統的な研究では、音楽学者が作曲技法に関する詳細や社会的文脈との関連について調べている [1–4]。ほとんどの場合、これらの研究は、定性的な分析に留まり、予測や仮説の検証につながる定量的な定式化は行われていない。近年では、音楽電子データの入手が容易になり、コンピューターを用いた分析手法が発達したことから、大規模な定量分析を行う研究が増えている [5–14]。西洋のクラシック音楽やポピュラー音楽では、様々な音楽特徴量の平均値が方向性を持って、徐々に変化する傾向があることが示されている [5–7, 9–11]。また、こうした変化は、急激な変化を含む期間（革命期）を断続的に含むことも分かっており [6, 7, 9]、音楽史研究において広く受け入れられている見方である、時代区分（ロマン派 [2] やロック時代 [3] など）の遷移としての音楽進化の記述を正当化する結果も得られている。

一方で、各社会の中にはジャンルなどに対応する音楽スタイルのクラスターが複数同時に存在することが多いため、各時代における音楽特徴量の分布を平均値のみで表すことによる限界も示されている [7, 9]。例えば、クラシック音楽の作曲家は長音階と短音階を両方用いたが、これらを平均した音階を用いることはほぼなかった。音楽特徴量の

分布のこのような内部構造を扱うため、クラスターリング手法を用いて、得られたクラスターの頻度の時間変化を分析した研究がある [7, 9]。得られたクラスターは、既知のジャンルや解釈可能な作曲スタイルに対応することが多く、音楽スタイルの革命期は、しばしば特定のクラスターの流行や衰退に関連付けられることが分かった。

これらの結果は、作曲スタイルの分布は共時的かつ過渡的なクラスター構造を持ち、分布の時間発展はクラスターの相対頻度のダイナミクス（クラスター間ダイナミクス）により生じることを示している。では、これらのクラスターは内部にダイナミクスを持つであろうか？これは、音楽進化の過程においてクラスターレベルの選択に加えて、作品レベルの特徴に対する選択圧の影響を知る上で重要な問題である。また、クラスターのダイナミクスを組み込むことで作曲スタイルの未来の分布をどの程度予測できるであろうか？これらの問題を調べるためには、クラスター間およびクラスター内のダイナミクスをデータから推定し、予測に用いるためのモデルの構築が必要である。

本稿では、統計モデル手法を用いた音楽データの進化分析を行う。分析には、日本のポピュラー音楽の過去約 70 年間のヒット曲のメロディーのデータを用いる。分析する特徴量としては、自動作曲や音楽ジャンル分類に関する研究で作曲スタイルの知識を表すために用いられる、音楽の統計量を用いる。データから抽出した統計量に対してクラスターリングを行い、クラスターの平均や分散の動態を分析する。また、分析で見つかったクラスター構造の重要な動的モードが、動的ディリクレ混合モデル (Dynamic Dirichlet mixture model; DDMM) により表されることを示す。さ

<sup>1</sup> 京都大学 Kyoto 606-8501, Japan

<sup>2</sup> Universitat Pompeu Fabra, 08002 Barcelona, Spain

<sup>a)</sup> eita.nakamura@i.kyoto-u.ac.jp

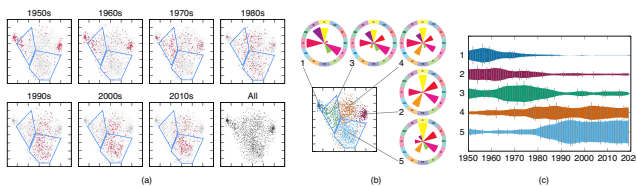


図 1 JPOP データの音高統計量の分布とクラスター構造の時間変化. (a) データ分布の 2 次元可視化結果. (b) クラスター分析の結果 (円の中の図は音高クラス分布を表す). (c) クラスターの相対頻度の変化.

らに、クラスター間の競合および各クラスター内での音楽要素間の競合を記述する適応度に基づくモデルを定式化することで、過去のデータから未来の作曲スタイル分布を予測する手法を構築し、クラスター間およびクラスター内のダイナミクスの各々の貢献度を定量的に調べる。

## 2. 作曲スタイル分布のダイナミクスの分析

本節では日本のポピュラー音楽データのクラスター構造を分析して、クラスター内ダイナミクスの特徴について調べる。また、動的クラスター構造の重要な側面が DDMM により記述されることを説明する。

### 2.1 分析に用いるデータと音楽統計量

日本のポピュラー音楽のデータ (JPOP データ) として、1950 年から 2019 年までのオリコンの年間売上チャートの上位 20 曲のメロディーの電子データを用いる (音源が入手できなかった 1 曲を除いたため計 1399 曲ある)。分析では、各メロディーを音符ごとの音高と発音時刻の系列として表す。音高は半音単位で表し、全ての曲をハ長調またはイ短調に予め移調したものを用いる。発音時刻は、小節内での相対的な位置 (拍節位置)  $b \in \{0, 1, \dots, 47\}$  を示す整数により表す。例えば、 $b = 0$  は小節の頭を表し、4/4 拍子では、 $b = 12$  は第 2 拍の位置を表す (小節を 48 単位に分解して表すものとする)。

各曲の作曲スタイルを表す音楽統計量として、音高とリズムのバイグラム確率を用いる。これらは、音階やリズムパターンを表すために用いられるマルコフモデルの遷移確率と等価である [15]。音高バイグラムを得る際には、拡張音高クラス表現を用いる。即ち、音高  $p'$  から  $p$  への遷移はバイグラム  $(q', q)$  により表す ( $0 \leq q' \leq 11, 0 \leq q \leq 35, q' \equiv p' \pmod{12}, q \equiv q' + p - p' \pmod{36}$ ) を満たす)。これにより、音域には依存せず、 $-17$  半音から  $17$  半音までの音程を記述できる表現になっている。音高バイグラム確率の要素数は、 $12 \times 36$  個である。リズムバイグラム確率は拍節位置のバイグラムの頻度から計算され、その要素数は  $48 \times 48$  である。メロディー生成のためには、音高と拍節位置の確率の両方が必要であるが、解釈を容易にするため、以下ではこれらを別々に分析する。

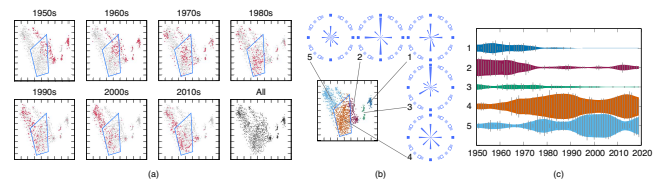


図 2 JPOP データのリズム統計量の分布とクラスター構造の時間変化. (a) データ分布の 2 次元可視化結果. (b) クラスター分析の結果 (円の中の図は拍節位置の分布を表す). (c) クラスターの相対頻度の変化.

各曲  $n$  を音高またはリズム統計量の確率ベクトル  $\theta_n$  と作曲年  $t_n$  の組  $(\theta_n, t_n)$  として表した後、離散分布混合モデルを用いてクラスタリングを行う。  $K$  個のクラスターを EM アルゴリズムで学習して、学習後に各曲を最尤推定により一つのクラスターに割り当てる。ただし、クラスタリングを行う時は時間情報は用いないものとする。

### 2.2 定性的分析

まず、作曲スタイル分布の動態を概観するため、t-SNE [16] を用いて曲の確率ベクトルを 2 次元空間に可視化する。この際、各確率ベクトルの対に対して Jensen-Shannon (JS) ダイバージェンスを計算する。まず、得られた可視化結果におけるクラスター構造を議論する。図 1(a) の音高統計量に対する結果では、左右それぞれに小さなクラスターがあり、中央に大きなクラスターが広がっていることが分かる。大きなクラスターの内部には、密度の変動にいくつかの構造があるが、明確に分離したクラスターはない。図 1(b) に示す 5 クラス ( $K = 5$ ) によるクラスタリングの結果では、2 つの小さなクラスターが捉えられ、大きなクラスターは 3 つのクラスターに分離されていることがわかる。音高クラス確率により表されている各クラスターの平均統計量を見ると、5 個のクラスターがそれぞれ別の音階に結び付けられることが分かる。クラスター 1 はマイナーの五音音階、クラスター 2 はメジャーの五音音階、クラスター 3 はマイナーの六音音階、クラスター 4 は短音階 (ダイアトニックスケール)、クラスター 5 は長音階に対応している。六音音階を除き、これらの音階は音楽学の文献で良く知られるものである [17]。俯瞰的に見れば、クラスター 3, 4, および 5 はダイアトニックスケールの亜種と考えられる。これらのクラスターとそれらを含む大きなクラスターは、階層構造を成している。

リズム統計量の結果についても同様の観察ができる。図 2(a) では、右側に 3 つの明確に分離した小さなクラスターがあり、左側には内部構造を持つ大きなクラスターがある。図 1(b) の 5 クラスによるクラスタリングでは、この大きなクラスター内のデータは 3 つのクラスターに割り当てられ、右側のクラスターは 2 つのクラスターに割り当てられている。5 つのクラスターは拍節位置確率により表された平均統計量により音楽的に解釈できる。クラスター 1 は付

点リズム, クラスター 2 は 4 ビートリズム, クラスター 3 は 3 連符リズム, クラスター 4 は 8 ビートリズム, クラスター 5 は 16 ビートリズムに対応している. この作曲スタイルの分布も階層的構造を持ち, 大きなクラスターは互いに重なり合う小さなクラスターを含んでいる.

次に作曲スタイル分布のダイナミクスについて議論する. 図 1(a) と 2(a) から, 両方の統計量について, 大きな時代変化があることが分かる. 音高統計量では, 1950 年代と 1960 年代には多くの曲が左右の両極端の領域に分布しているのに対して, 1990 年代以降は多くの曲が中央に分布している. この全体的な変化はクラスター頻度の変化に明確に見ることができる (図 1(c)). 初期にはクラスター 1 と 2 が優勢であり, 後期にはクラスター 4 と 5 が優勢である. 同様に, リズム統計量では, 1950 年代と 1960 年代には多くの曲が上部右側の領域に位置しているのに対して, 2000 年代以降では下部左側の領域に位置している. 図 2(c) のクラスター頻度の変化は, この全体の遷移を反映している. これにより, 日本のポピュラー音楽の作曲スタイルの全体的な変化はクラスター間ダイナミクスにより捉えられることが分かる. これは, アメリカのポピュラー音楽 [7] や西洋クラシック音楽 [9] に関する結果と同様である.

一方で, クラスター内ダイナミクスも観察することができる. まず, 1990 年代から 2010 年代にかけて音高統計量のクラスター 5 の分散が小さくなっていることが見て取れる. このクラスター収縮はこの時期の市場集中の動きに起因すると考えられる. 実際, 1990 年代のチャートには 103 の異なるアーティストが現れているが, 2000 年代と 2010 年代にはそれぞれ 76 と 32 のアーティストしか現れていない. また, 音高統計量の分布のクラスター 3 は, 1950 年代から 1980 年代にかけて, クラスターの中心部分が大きく移動している. これは, チャートに現れた曲における傾向が, マイナーの五音音階からダイアトニックスケールを好むように変化したことを反映している. 類似のクラスターの移動はリズム統計量のクラスター 4 で 1960 年代から 1980 年代にかけても観察できる. この移動はシンコペーションをより好むようになった動きを表したものである.

以上の結果により, 作曲スタイル分布の変化では, クラスター内ダイナミクスの特徴的な動的モードとしてクラスターの収縮や中心の移動が含まれることが分かった. よって, 音楽進化の有効なモデルには, クラスター間とクラスター内のダイナミクスの両方を組み込む必要があることが示唆された.

### 2.3 ディリクレ分布を用いたクラスター内ダイナミクスの定量分析

前節の定性的な分析は直感的ではあるが, 2 次元空間での可視化は, 実際の分布の粗い近似と見なすべきである. ここではクラスター内ダイナミクスを定量的に分析するた

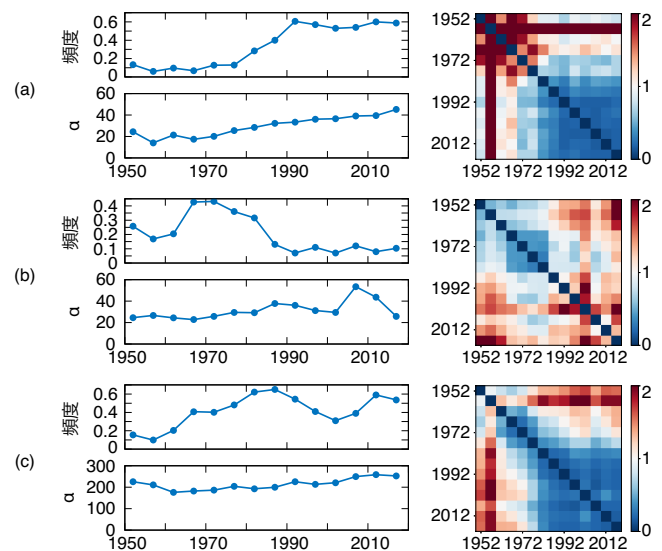


図 3 クラスターの相対頻度, 集中度, および平均分布の変化. (a) と (b) は図 1 のクラスター 5 とクラスター 3, (c) は図 2 のクラスター 4 に対する結果. 平均分布の変化は右側の, 時間スライス間での対称 KL ダイバージェンスにより表されている.

め, 確率ベクトル空間におけるクラスターを定式化する. ユークリッド空間におけるクラスター分析では, ガウス分布を用いる定式化が一般的である. 同様に, 確率ベクトルの空間では, 以下に説明するディリクレ分布を用いることができる.  $D$  次元の確率ベクトル  $\theta = (\theta_i)_{i=1}^D$  ( $\sum_i \theta_i = 1$ ) に対して, ディリクレ分布は次の通り定義される.

$$\text{Dir}(\theta; \alpha, \mu) = \Gamma(\alpha) \prod_{i=1}^D \frac{\theta_i^{\alpha\mu_i - 1}}{\Gamma(\alpha\mu_i)}. \quad (1)$$

ここで, パラメーター  $\alpha > 0$  は集中度, 確率ベクトル  $\mu = (\mu_i)_{i=1}^D$  は平均分布 (または基底分布) と呼ばれる.

ディリクレ分布は確率ベクトルを生成する確率分布であり, 要素ごとの平均と分布は,  $E(\theta_i) = \mu_i$  と  $V(\theta_i) = \mu_i(1 - \mu_i)/(\alpha + 1)$  で与えられる. これにより, 平均  $\mu$  はクラスターの中心として解釈でき, 集中度  $\alpha$  はクラスターのサイズに関するパラメーターである (分散は  $\alpha + 1$  に反比例する). あるクラスターに割り当てられた確率ベクトルの集合が与えられた時, これらのパラメーターは最尤推定法を用いて推定できる [18].

図 3 は, いくつかのクラスターについて時間スライスごとに分割されたデータサンプルから推定されたディリクレ分布のパラメーターを示している. 平均分布の時間変化を定量化するため, 一般の確率分布に対する距離尺度である対称 Kullback-Leibler (KL) ダイバージェンスを用いている. 図 3(a) は音高統計量のクラスター 5 の変化を示したもので, 1990 年代から 2010 年代にかけてのクラスター収縮は集中度パラメーターの増加により表されている. この期間ではどの時間スライス間での平均分布の対称 KL ダイバージェンスは 1 よりも小さく, クラスター中心の移動は比較的小さいことがわかる. 図 3(b) は音高統計量のクラ

スター 3 の変化を示したもので、1950 年代から 1980 年代にかけてのクラスター中心の移動は対称 KL ダイバージェンスの行列の非対角成分の比較的大きな値に反映されている。同様に、図 3(c) に示されたリズム統計量のクラスター 4 の変化では、対称 KL ダイバージェンスの行列の非対角成分が 1960 年代から 1980 年代において、1990 年代から 2010 年代における値よりもかなり大きくなっていることから、初期の期間におけるクラスター中心の移動が定量的に示されていることがわかる。

これらの結果より、作曲スタイルのクラスター内ダイナミクスの定量分析にディリクレ分布が用いられることが示された。特に、クラスターの収縮（または膨張）は集中度の変化、クラスター中心の移動は平均分布の変化により表される。次節で説明するように、ディリクレ分布は未来の分布を予測する際に、予測精度を尤度に基づいて評価するためにも用いることができる。

### 3. 予測モデルの構築

#### 3.1 動的ディリクレ混合モデル (DDMM)

本節では、クラスター構造のダイナミクスを組み込んだモデルを構築し、未来の作曲スタイル分布の予測と、この予測におけるクラスター内ダイナミクスの重要性の検証を行う。2.3 節で述べた通り、確率ベクトルの空間における作曲スタイルのクラスターはディリクレ分布により記述できる。よって、複数のクラスターから構成される作曲スタイルの分布は、次で定義するディリクレ混合モデル (Dirichlet mixture model; DMM) により記述できる。

$$P(\boldsymbol{\theta}, t) = \sum_{k=1}^K \pi_k(t) \text{Dir}(\boldsymbol{\theta}; \alpha_k(t), \boldsymbol{\mu}_k(t)). \quad (2)$$

式 (1) と同様に  $\boldsymbol{\theta}$  は確率ベクトルを表し、 $k$  はクラスターの添字、 $K$  はクラスター数を表す。混合確率と呼ばれる  $\pi_k$  はクラスター  $k$  の相対頻度を表し、 $\sum_k \pi_k = 1$  を満たす。

式 (2) では、作曲スタイル分布のダイナミクスは右辺に現れるパラメータの時間発展に分解されている。クラスター間ダイナミクスは、クラスター頻度の時間変化を表す時間依存の混合確率  $\pi_k(t)$  に組み込まれている。クラスター内ダイナミクスは、クラスター中心の移動を表す時間依存の平均分布  $\boldsymbol{\mu}_k(t)$  と、クラスターの収縮または膨張を表す時間依存の集中度  $\alpha_k(t)$  に組み込まれている。式 (2) のモデルを動的ディリクレ混合モデル (dynamic Dirichlet mixture model; DDMM) と呼ぶ。

与えられた現在までのデータから未来の分布の予測するために DDMM を使うには、二つの問題を扱う必要がある。一つ目はパラメータ推定問題で、与えられたデータから現在までの各時刻におけるパラメータ  $\pi_k(t)$ 、 $\alpha_k(t)$ 、 $\boldsymbol{\mu}_k(t)$  の値を推定するものである。二つ目はパラメータ予測問題で、この推定パラメータに基づいて未来のパラ

メータ値を予測するものである。以下、一つ目の問題を 3.2 節で、二つ目の問題を 3.3 節で議論する。問題設定として、2 節と同様に、各曲のデータはどれか一つのクラスターに割り当てられている状況を考える。

#### 3.2 パラメータ推定問題

パラメータ推定の単純な方法は、時間スライスごとにデータを分割してパラメータを推定することである。しかし、この方法はデータ量が限られているときには効率的でない。そこで、データを時間に関して指数関数的に減衰する重みを付けた上で、パラメータを推定する方法を考える。未来のデータの予測に、現在までのデータしか使えない状況を考えているため、各時刻のパラメータ推定にはその時刻までのデータのみを用いることにする。

混合確率と平均分布の推定は、相対頻度の計算により行えるため、重み付きのデータへの拡張は簡単である。集中度に関しては、各時間スライスのデータに関する対数尤度を重み付けした関数を最大化するように推定する。時間に関する重み付けの減衰率（時定数）は、推定された DMM の尤度を最大化することで最適化する。実験的に、この値は音高統計量に対して  $\tau = 10$ 、リズム統計量に対して  $\tau = 8$ （どちらも単位は年）と求まった。

#### 3.3 パラメータ予測問題

DDMM のパラメータを予測するため、クラスター間とクラスター内ダイナミクスのモデルを構築する。これらのダイナミクスは、特定のスタイルで曲を作る創作者の好みや特定のスタイルの曲を好む鑑賞者の好みに起因すると考えられる。そこで、進化生物学 [19, 20] の考え方を参考にすると、競合を含む増殖過程に基づくモデルが考えられる。クラスター間ダイナミクスは、クラスターごとに異なる適応度（複製率）を持つ個体の増殖過程として捉えられ、適応度はクラスターに対する創作者と鑑賞者の好みにより決定される。クラスター内ダイナミクスは、曲を構成する音楽要素間の競合を含む同様の過程により記述され、適応度は曲の内容に対する創作者と鑑賞者の好みを表す。

これらの進化モデルを予測問題に用いるためには、適応度を過去のデータから推定する必要がある。進化の過程は確率的である一方で、観測できるのは有限のデータであるため、適応度の推定においては統計ノイズを考慮に入れる必要がある。また、音楽の好みはそれ自体、時代により変化すると考えられるため [10]、適応度の時間変化も組み入れたモデルが望ましい。以下では、統計的な枠組みに基づき、適応度の頑健な推定と時間変化する適応度への拡張が可能なモデルの定式化を行う。

まず、混合確率  $\pi_k(t)$  の進化を予測するモデルについて説明する。適応度に基づく進化ダイナミクスの基本方程式 [19] は以下で表される。

$$\pi_k(t+1) = w_k \pi_k(t) / \bar{w}, \quad (3)$$

または

$$\ln \pi_k(t+1) = \ln \pi_k(t) + \ln(w_k / \bar{w}). \quad (4)$$

ここで、 $w_k$  はクラスター  $k$  の適応度を表し、平均適応度は  $\bar{w} = \sum_{\ell} w_{\ell} \pi_{\ell}(t)$  で定義される。二つ目の方程式は、一つ目の方程式の単純な変形であるが、対数確率に対する運動方程式として見る事ができる。相対適応度  $w_k / \bar{w}$  の対数が速度に対応する。

いくつか注意点を述べる。まず、相対適応度は  $\bar{w}$  を通して全ての  $\pi_{\ell}(t)$  に依存するが、進化速度が小さい場合は、 $\pi_k(t)$  のダイナミクスに対する他の  $\pi_{\ell}(t)$  の影響は無視できる。以下ではこの近似を用いて、 $\pi_k(t)$  のダイナミクスはそれぞれ切り離して考える。また、上の方程式は無限集団に対するものであり、有限集団における統計ノイズの効果は含んでいない。  $x(t) = \ln \pi_k(t)$  および  $v(t) = \ln(w_k / \bar{w})$  と書いて、観測される  $x(t+1)$  に含まれるノイズを  $\epsilon(t)$  と表すと、ノイズを含むモデルは以下で表される。

$$x(t+1) = x(t) + v(t) + \epsilon(t). \quad (5)$$

具体的には、 $\epsilon(t)$  はガウスノイズで表し、 $\langle \epsilon(t) \rangle = 0$  および  $\langle \epsilon(t) \epsilon(s) \rangle = \sigma^2 \delta_{ts}$  を満たすものとする。時間不変の適応度の場合は、 $v(t)$  は定数となる。

このモデルを拡張して適応度のダイナミクスを取り入れるには、 $v(t)$  を時間依存変数としてその力学方程式を追加すればよい。 $v(t)$  の時間変化を表す変数  $\eta(t)$  を導入することで、自然なモデルとして次のものが得られる。

$$v(t+1) = v(t) + \eta(t). \quad (6)$$

ただし、 $\eta(t)$  もガウスノイズで記述されるものとする。一般には、高次の時間微分を表す変数を導入して、モデルをさらに拡張することもできる。式 (5) と (6) は統計学で状態空間モデルとして知られている [21]。よく知られた状態空間モデルの理論を用いることで、データから適応度（あるいは速度  $v(t)$ ）を推定して未来の  $x(t)$  の値を予測できる。

同様のモデルは平均分布  $\mu_{ki}(t)$  および集中度  $\alpha_k(t)$  の進化の予測にも適用できる。平均分布の場合は、適応度は添字  $i$  に対応する音楽要素に対する創作者と鑑賞者の好みを表す。集中度は頻度を表すパラメーターではないため、対応する適応度について直感的な解釈を与えることは難しい。しかし、その進化が滑らかな変化を伴うダイナミクスによるものであれば、近似モデルとして状態空間モデルを用いることができる。以下では、これらのモデルを用いる。

実際のデータを用いた解析により、データの予測には、モデルで予測された適応度の値よりも保守的な値を用いる方が予測精度が高くなる事が分かった。即ち、モデルで速度を予測した後、データ予測を行う際に、一定の削減率（適応度削減係数と呼ぶ）をかけた値を用いるというもの

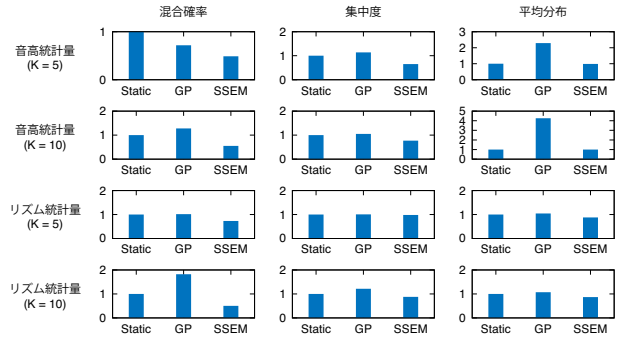


図 4 静的モデル (Static), ガウス過程 (GP), 状態空間進化モデル (SSEM) による DDMM パラメーターの予測の誤差. 全ての誤差は静的モデルの値で規格化されている。

である。以下の解析では、削減係数はパラメーターそれぞれに対して  $[0.01, 1]$  の範囲で最適化されたものを用いる。

### 3.4 予測精度の評価

予測精度の評価のため、二つの実験を行う。一つ目の実験では、DDMM パラメーターの予測精度を要素ごとに評価して、他の手法と比較することで、適応度に基づく進化モデルの有効性を検証する。二つ目の実験では、予測分布の尤度により予測精度の全体的な評価を行い、パラメーター要素ごとの予測モデルの重要性や予測時間の依存性について調べる。

DDMM パラメーター予測の実験では、まず、3.2 節の方法により、参照年  $t_0$  以前に作られた音楽データから DDMM パラメーターを推定する。次に、それ以降の年におけるモデルパラメーターを 3.3 節の方法で予測する。以降、この方法を状態空間進化モデル (state-space evolutionary model; SSEM) と呼ぶ。比較のため、参照年に対する推定モデルパラメーターそのまま未来における予測値として使う方法を以降、静的モデル (static model) と呼ぶ。また、回帰問題の一般的な手法であるガウス過程 (Gaussian process; GP) [22] による予測結果との比較も行う。ガウス過程は、線形回帰モデルを一般化したものと考えられる。実験では、線形カーネル、ガウスカーネル、指数  $\nu = 5/2$  の Matérn カーネル、およびバイアス項からなる多重カーネル GP を用いた。SSEM と GP を平均分布の予測に用いる際には、過去データにおける平均確率が 0.01 よりも大きい要素に関してのみこれらの手法を適用し、それ以外の低頻度の要素に対しては静的モデルを用いた。これは低頻度の要素は統計ノイズの影響が大きく、信頼できる予測が難しいからである。GP のカーネルの重みとハイパーパラメーターは、GPy ライブラリー [23] により最適化したものを用いた。予測能力の評価では、DDMM パラメーターの誤差を要素ごとに計算した。混合確率と平均分布に対しては、予測誤差として対称 KL ダイバージェンスを用いた。集中度に対しては対数自乗誤差を用いた。

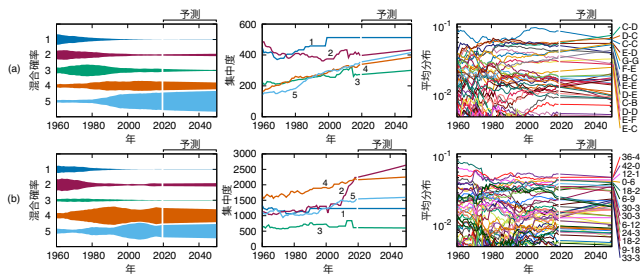


図5 (a) 音高統計量と(b) リズム統計量の分布変化の予測。一番右のパネルは5番目のクラスターの平均分布を示しており、右側のラベルは対応するバイグラムを示している。

図4に結果を示す。参照年  $t_0 \in [1969, 2009]$  に対して、音高統計量とリズム統計量の両方に対する結果を上記の3手法について示してある。混合確率に対しては、全ての場合においてSSEMは静的モデルに比べて誤差が有意に小さかった。GPは、音高統計量でクラスター数  $K=5$  の場合を除いて、静的モデルより誤差が大きかった。集中度では、全体的にSSEMは静的モデルより誤差が小さく、音高統計量でクラスター数  $K=10$  の場合は両者の誤差は同程度であった。GPの誤差は全ての場合において静的モデルの誤差と同じか、より大きかった。平均分布のパラメータでは、SSEMと静的モデルは音高統計量では同程度の誤差であり、リズム統計量ではSSEMの方が少し誤差が小さかった。GPは全ての場合で誤差が最大であった。

全体的に見ると、GPは静的モデルよりも誤差が大きいたく、DDMMパラメータの予測は非自明な問題と言える。多くの場合において、SSEMは静的モデルよりも正確にパラメータを予測することができた。特に、混合確率の予測における誤差の低減の度合いが大きく、クラスター間ダイナミクスの予測における適応度に基づくモデルの有効性が示された。一方で、平均分布における誤差の低減の度合いは小さく、これらのパラメータはより複雑なダイナミクスに従っている可能性を示唆している。

SSEMによるDDMMパラメータの予測値を図5に示す(クラスター数  $K=5$  の場合)。ここではトレンドの観察を容易にするため、平均分布に対する適応度削減係数は1とした時の結果を示している。音高統計量とリズム統計量の両方において、今後、5番目のクラスターが発達を続け、優勢になると予測された。リズム統計量では、2番目のクラスターも近い将来には流行することが予測されている。流行するクラスターの全てにおいて、集中度が増加すること、即ち、各クラスター内での多様性は減少することが予測された。

音高統計量の5番目のクラスターに注目すると、高頻度のバイグラム確率について変化が予測されていることが分かる。特に、(C,D), (F,E), (B,C)などの順次進行の確率は増加し、(C,C), (E,E), (D,D)などの同音連打の確率は減少すると予測されている。リズム統計量の5番目のクラス

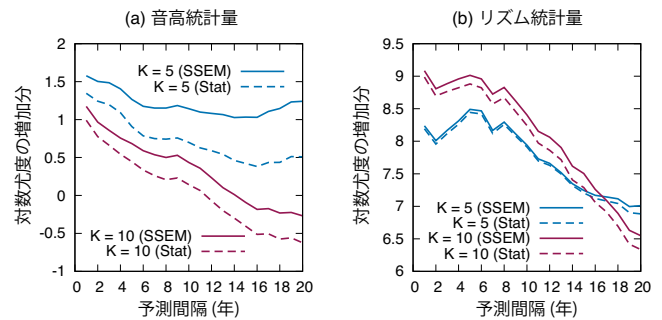


図6 (a) 音高統計量と(b) リズム統計量に対する、SSEMと静的モデル(Static)の対数尤度の増加分。

ターでは、現在、高頻度のバイグラムの確率が徐々に減少することが予測され、バイグラム確率は全体的に一様化する傾向が示されている。注目すべき例外は、(6,9), (30,33), (33,36)などの16分音符長に対応するバイグラムで、これらの確率は急激に増加することが予測されている。音高とリズムの統計量に見られるこれらのトレンドは、同音連打や8ビートリズムなどの単調な進行の割合が減少するという点で共通している。

作曲スタイル分布の予測を行う二つ目の実験では、SSEMにより予測されたパラメータを用いて、参照年  $t_0$  から20年間の期間における確率ベクトルの尤度を測定する。ここで、 $t_p (> t_0)$  年における分布の予測には、 $t_p + \Delta_p$  年に対して予測されたパラメータを用いた。 $\Delta_p$ は先読みをするためのパラメータであり、重み付けによるスムージングの効果を相殺するために導入している。実験では、最適化された値として、音高統計量に対して  $\Delta_p = 8$ 、リズム統計量に対して  $\Delta_p = 15$  を用いた。同様の理由で、分布の予測時には、集中度は増加させた値を用いた。これは各年における作曲スタイル分布の集中度は、スムージングされた分布の集中度よりも大きくなっていると考えられるからである。実験では、全ての場合において、集中度を5倍したものをを用いた。比較のため、同じ条件で静的モデルを用いて予測した場合の尤度も測定した。クラスター数が  $K=5$  および  $10$  の場合を評価し、クラスターが1つの場合 ( $K=1$ ) の静的モデルをベースラインとして用いた。

図6では、( $K=1$ の静的モデルに対する)対数尤度の増加分を、予測年と参照年の差で定義される予測間隔に対して示している(参照年については  $t_0 \in [1969, 2009]$  の範囲で平均してある)。まず、全ての場合において、SSEMは静的モデルよりも尤度が大きく、少なくとも20年間までの予測間隔では、SSEMによるDDMMパラメータの予測が作曲スタイル分布の予測に役立つことが示された。次に、音高統計量とリズム統計量の両方について、 $K=10$ のモデルによる尤度の増加分は予測間隔が長くなると減少しており、予測間隔が20年の付近では  $K=1$  (音高統計量) または  $K=5$  (リズム統計量) のモデルによる増加分よりも小さくなっている。この結果は、クラスター数が多

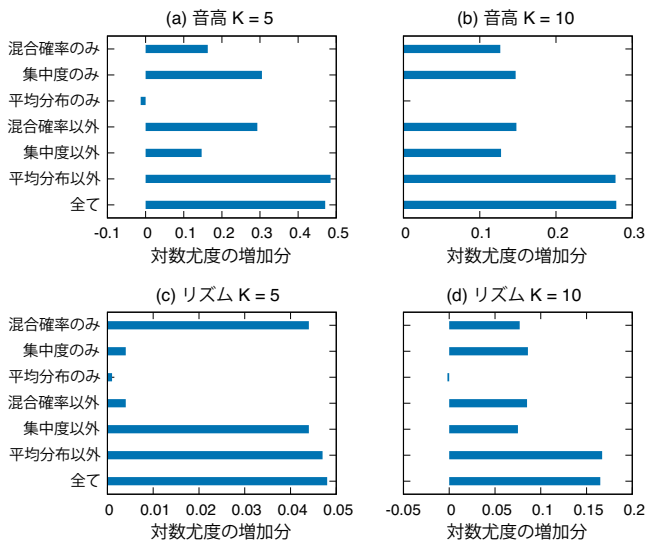


図 7 対数尤度の増加分に対する DDMM パラメータの各要素の貢献度. (a) 音高統計量 ( $K = 5$ ), (b) 音高統計量 ( $K = 10$ ), (c) リズム統計量 ( $K = 5$ ), (d) リズム統計量 ( $K = 10$ ).

いモデルでは分布の予測がより難しいことを示している。過学習やデータスパースネスの問題に加えて、予測が難しいモデルパラメータ数の増加が尤度の増加分の悪化を招いた可能性が考えられる。

最後に、尤度の増加分に対する DDMM パラメータの各要素の貢献度を調べるため、これらのパラメータの部分集合に対して SSEM の予測を行い、その他については静的モデルを用いた場合の増加分を比較する。図 7 の結果では、混合確率と集中度に対する SSEM の予測は尤度の増加分に寄与しており、その程度はしばしば同等であることがわかる。これはクラスター間ダイナミクスとクラスター内ダイナミクスの予測の両方が尤度の向上に重要であることを示唆している。しかし、平均分布の予測の貢献度は小さく、時には悪影響を与えることもわかる。これは、先程と同様、これらのパラメータの予測が難しいことが主な原因である。

#### 4. 議論

本研究では、統計モデルの手法を用いて作曲スタイル分布の時間発展を調べた。JPOP データから抽出された音楽統計量から得られたクラスターを分析することで、クラスターの収縮や中心の移動などのクラスター内ダイナミクスと、クラスターの相対頻度におけるクラスター間ダイナミクスの両方が確認された。クラスターのこれらの動的モードを統一的に組み込んだモデルとして動的ディリクレ混合モデル (DDMM) を構築し、これを用いた未来の作曲スタイル分布の予測について調べた。特に、作曲スタイル分布の時間発展を特徴づけ、予測する上で、クラスターの分散の分析の重要性が明らかになった。これまでの文化進化の定量分析の研究では、クラスターの頻度や平均のみに注目

するものが多く [7, 10, 24, 25], クラスターの分散に対するさらなる調査が望まれる。

本研究の結果から文化進化の量的および理論的な理解のための示唆が得られた。第一に、作曲スタイルの空間における異なるクラスターは異なるクラスター内ダイナミクスのモードを持っていることが示されたことから、このダイナミクスの違いの原因を調べる必要がある。文化データのクラスターは、社会の中で特定のコミュニティーが関わるようなジャンルに対応することが多いため、コミュニティーごとの嗜好性のタイプを調べることも重要と考えられる。

第二に、図 1 や図 2 から、創作スタイルは階層的なクラスター構造を持ち、クラスターは互いに重複することが多いことが示された。このような状況では、クラスター数を定めることは難しく、クラスター間ダイナミクスとクラスター内ダイナミクスの違いも相対的な意味しか持たない。本研究のように分析時には固定したクラスター数を用いる方法はダイナミクスの解釈を容易にする利点もあるが、ノンパラメトリックベイズの枠組み [26, 27] を用いて全てのクラスター数の可能性を考慮した分析も重要だと考えられる。

第三に、状態空間進化モデル (SSEM) を用いた予測結果では、混合確率と集中度に比べて、平均分布の予測が難しいことが分かった。作曲スタイルのデータ空間の高次元性がこの原因の一つとして考えられる。また、図 5 の分析では、音楽要素の進化は音楽的な関係性に従っており、要素ごとに独立ではないことが示された。これは音楽の表面上の特徴量次元よりも、進化の有効次元が小さく、生物の表現型の進化での観察 [28, 29] と類似する現象が起こっている可能性を示している。よって、データから進化の有効空間を推定して、その低次元空間での進化モデルを構築することが重要だと考えられる。

最後に、確率的生成モデルの統計量の進化を分析する本研究の方法論は、創作物の特徴の分析や予測を可能とするのみならず、過去・現在・未来のスタイルにおける創作物の生成も可能である。合成による分析 (Analysis-by-Synthesis [30]) の枠組みと合致するこのアプローチは、現在広く研究されている機械学習に基づく自動音楽生成技術 [31] の課題である、既存の音楽の模倣を超えた、創造的な音楽生成ができる知能システムの実現にも応用できると期待される。

#### Acknowledgment

謝辞 本研究は、日本学術振興会科研費 16J05486, 22H03661, 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。

## 参考文献

- [1] R. L. Crocker, *A History of Musical Style*, McGraw-Hill, 1966.
- [2] J. P. Burkholder, D. J. Grout, and C. V. Palisca, *A History of Western Music (8th ed.)*, W. W. Norton & Company, 2010.
- [3] K. Negus, *Popular Music in Theory*, Polity Press, 1996.
- [4] F. Holt, *Genre in Popular Music*, The University of Chicago Press, 2007.
- [5] J. Serrà, Á. Corral, M. Boguñá, M. Haro, and J. L. Arcos, “Measuring the Evolution of Contemporary Western Popular Music,” *Scientific Reports*, **2(521)**, pp. 1–6, 2012.
- [6] P. H. R. Zivic, F. Shifres, and G. A. Cecchi, “Perceptual Basis of Evolving Western Musical Styles,” *Proc. Natl. Acad. Sci.*, **110(24)**, pp. 10034–10038, 2013.
- [7] M. Mauch, R. M. MacCallum, M. Levy, and A. M. Leroi, “The Evolution of Popular Music: USA 1960–2010,” *Royal Society Open Science*, **2**, pp. 150081, 2015.
- [8] D. Park, A. Bae, M. Schich, and J. Park, “Topology and evolution of the network of western classical music composers,” *EPJ Data Science*, **4**, pp. 1–15, 2015.
- [9] C. Weiß, M. Mauch, S. Dixon, and M. Müller, “Investigating Style Evolution of Western Classical Music: A Computational Approach,” *Musicae Scientiae*, **23(4)**, pp. 486–507, 2019, DOI: 10.1177/1029864918757595.
- [10] M. Interiano, K. Kazemi, L. Wang, J. Yang, Z. Yu, and N. L. Komarova, “Musical trends and predictability of success in contemporary songs in and out of the top charts,” *Royal Society Open Science*, **5**, pp. 171274, 2018.
- [11] E. Nakamura and K. Kaneko, “Statistical evolutionary laws in music styles,” *Scientific Reports*, **9(15993)**, pp. 15993, 2019.
- [12] M. Youngblood, “Conformity bias in the cultural transmission of music sampling traditions,” *Royal Society Open Science*, **6(9)**, pp. 191149, 2019.
- [13] F. C. Moss, M. Neuwirth, and M. Rohrmeier, “The line of fifths and the co-evolution of tonal pitch-classes,” *Journal of Mathematics and Music*, pp. 1–25, 2022.
- [14] P. E. Savage, S. Passmore, G. Chiba, T. E. Currie, H. Suzuki, and Q. D. Atkinson, “Sequence alignment of folk song melodies reveals cross-cultural regularities of musical evolution,” *Current Biology*, **32(6)**, pp. 1395–1402.e8, 2022.
- [15] C. Ames, “Automated composition in retrospect: 1956–1986,” *Leonardo*, **20(2)**, pp. 169–185, 1987.
- [16] L. Van der Maaten and G. Hinton, “Visualizing data using t-SNE,” *Journal of Machine Learning Research*, **9(11)**, 2008.
- [17] C. S. Stevens, *Japanese Popular Music*, Routledge, 2008.
- [18] T. P. Minka. Estimating a Dirichlet distribution. <https://tminka.github.io/papers/dirichlet/>.
- [19] J. F. Crow and M. Kimura, *An Introduction to Population Genetics Theory*, Harper and Row, 1970.
- [20] M. A. Nowak, *Evolutionary dynamics: exploring the equations of life*, Harvard University Press, 2006.
- [21] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [22] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2006.
- [23] GPY: A Gaussian process framework in python. <http://github.com/SheffieldML/GPy>.
- [24] T. L. Griffiths and M. Steyvers, “Finding scientific topics,” *Proc. Natl. Acad. Sci.*, **101(suppl. 1)**, pp. 5228–5235, 2004.
- [25] J.-B. Michel et al., “Quantitative Analysis of Culture Using Millions Of Digitized Books,” *Science*, **331**, pp. 176–182, 2011.
- [26] C. E. Antoniak, “Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems,” *The Annals of Statistics*, **2(6)**, pp. 1152–1174, 1974.
- [27] C. Rasmussen, “The infinite Gaussian mixture model,” in *Proc. Advances in Neural Information Processing Systems*, pp. 554–560, 1999.
- [28] Z. Frentz, S. Kuehn, and S. Leibler, “Strongly deterministic population dynamics in closed microbial communities,” *Physical Review X*, **5(4)**, pp. 041014, 2015.
- [29] C. Furusawa and K. Kaneko, “Formation of dominant mode by evolution in biological systems,” *Physical Review E*, **97**, pp. 042410, 2018.
- [30] D. M. MacKay, “Mindlike behaviour in artefacts,” *The British Journal of the Philosophy of Science*, **2(6)**, pp. 105–121, 1951.
- [31] J. D. Fernández and F. Vico, “AI methods in algorithmic composition: A comprehensive survey,” *Journal of Artificial Intelligence Research*, **48**, pp. 513–582, 2013.