

確率的生成モデルに基づく音楽の類似度と ありがち度の推定に関する検討

中野 倫靖^{1,a)} 吉井 和佳^{2,b)} 後藤 真孝^{1,c)}

概要：本稿では、歌声と伴奏を含む音楽音響信号を対象として、「ボーカルの歌声」、「楽曲中の音色」、「リズム」、「和音進行」の確率的生成モデルを構築し、モデルからの生成確率を計算することで、「楽曲間の類似度」や「楽曲のありがち度」を推定する手法を提案する。歌声、音色、リズムに関しては、LPMCC、MFCC、Fluctuation Pattern に基づく音響特徴量を抽出し、それぞれに関して潜在的ディリクレ配分法 (LDA) を用いたトピック分析を行う。個々の楽曲毎に学習したモデルと全曲から学習したモデルから、楽曲における各音響特徴量の生成確率を計算することで、それぞれ楽曲間の類似度とありがち度を推定した。和音進行に関しては、能動的音楽鑑賞サービス Songle のコード認識結果に対し、可変長 Pitman-Yor 言語モデル (VPYLM) でモデル化した。ここでは、個々の楽曲毎に学習したモデルと全曲で学習したモデルから、各曲のパープレキシティの逆数 (各和音の平均的な生成確率) を計算することで、それぞれ類似度とありがち度として推定した。本稿では、ポピュラー音楽 3278 曲を対象として分析した結果を報告する。

1. はじめに

音楽はデジタル化され、今後もアクセス可能な楽曲が単調増加し続けていく。そうした状況において、ある楽曲と別の楽曲の類似度が適切に推定できれば、音楽聴取者が自分の好みの楽曲に出会うための検索手段として活用できる。従来、そのような技術に関しては、音楽情報検索 (MIR: Music Information Retrieval) [1-5] や音楽情報推薦 [6-8] に関する活発な研究開発が国内外で進められてきた。

ここで、楽曲と楽曲との類似度だけではなく、楽曲と楽曲集合 (ジャンル、例えばポピュラー音楽) との類似度を推定できれば、その集合を代表するような楽曲に出会う手段として利用できる。本研究では、このような類似度を「ありがち度」と呼ぶ。これは新たなジャンルを好きになるための「入門曲」を発見する指標として活用できる可能性がある。しかも、このように楽曲集合に基づいた音楽要素のありがち度を推定できれば、楽曲制作者にとっても有用である。なぜなら、そのような音楽要素はジャンル内で確立した表現であり、「ありふれている」ため、制作者にとって

の「共有の知」として利用しやすいからである。

このように、専門家だけでなく一般の人々も「何が似ているのか」「どれくらいありふれているのか」を、知ろうと思った時に容易に知ることができれば、過去の楽曲と共存共栄し、敬意を払う文化を築くことに貢献できる。例えば、論文のように引用され再利用されたら喜びを感じられる音楽文化が考えられる。そのような引用は、カバー楽曲の制作やニコニコ動画^{*1}でのN次創作^{*2}[9]においては通常、敬意を伴ってなされるが、新規楽曲を発表する際に他の楽曲への引用を記述することは稀である。

ここで問題なのは、人々が自分の作品を発表する際に、楽曲数の単調増加に伴って盗作疑惑を招く事例が増えてしまう懸念があることである。あらゆる楽曲は既存曲の影響を受けていると考えられるため、「無自覚に」「何らかの意味で」「部分的に」類似してしまうのは自然であり、本来問題ではないことが多い。しかし、もし自分の作品が何かに似ていると糾弾されるリスクが高いと、誰もが創作と発表を楽しむ「一億総クリエイター時代」が到来しても、安心して楽曲の制作や発表をしにくい社会になりかねない。

このような問題は、類似度に関する人間の能力の限界に起因して発生すると考えられる。人は目の前の二つを比較して類似度を判断することはできても、ある事象が全体の中でどれくらい確率的に起こりえるのか、という「ありが

¹ 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

² 京都大学 大学院情報学研究所
Kyoto University

a) t.nakano [at] aist.go.jp

b) yoshii [at] i.kyoto-u.ac.jp

c) m.goto [at] aist.go.jp

^{*1} <http://www.nicovideo.jp/>

^{*2} 「歌ってみた」動画や「踊ってみた」動画、合唱 (複数の歌ってみた動画のマッシュアップ作品) 等。

ち度」の判断には能力的・物理的に限界がある。例えば、最近よく起きた事象を起きやすいと思い込んだり、多数起きている事象でも遭遇しなければ滅多にないと誤解する。また、アクセス可能な楽曲を全て聞くことは量的に不可能なため、全体を俯瞰した適切な判断を行うことができない。

そこで我々は、音楽における「類似度」の判断を本質的と捉え、その判断能力の拡張や活用に基づく豊かな社会の実現に資するための「OngaCREST プロジェクト」を進めている [10]。本研究プロジェクトでは、膨大な音楽コンテンツ間の類似度を人々が知ることができる（可視化する）情報環境を実現する技術基盤の構築を目的とする。楽曲は「他にいかに類似していないか」という新規性のみならず、人々をいかに感動させ幸せにするかに価値があると、我々は考えている。新規性だけを追求するのではなく、過去の楽曲と共存共栄しながら、感動させる魅力や完成度の高さ等を重視する「過去の楽曲に敬意を払う文化」「感動体験重視型の音楽文化」の実現を目指す。

本稿では、そのための第一歩として、歌声と伴奏を含む音楽音響信号を対象として、4種類の音楽要素「ボーカルの歌声」「楽曲中の音色」「リズム」「和音進行」の確率的生成モデルを構築する。個別の曲および楽曲集合から、それぞれの音楽要素の確率的生成モデルを構築し、そのモデルからの音楽要素の生成確率を計算することで、楽曲間の類似度と楽曲のありがち度を推定する。

2. 確率的生成モデルに基づく類似度とありがち度

本稿で述べる「確率的生成モデル」、「類似度」、「ありがち度」という用語には、いくつかの捉え方が考えられるため、本章で事前に以下のように定義する。また、概要図を図1に示す。

確率的生成モデル 各音楽要素（音響特徴量や和音進行）がどのような形で出現しやすいか、その確率（生成確率）を計算できるモデルを指す。現時点では、そこから音楽要素を生成したり、楽曲を作ったりすることはできないが、将来的には楽曲を生成するモデルに発展できる可能性がある。

類似度 ある楽曲を基準に考えたときにそれに別の楽曲が似ている度合いを指す。基準の楽曲の確率的生成モデルから別の楽曲の音楽要素の生成確率を計算することで推定する。

ありがち度 ある楽曲集合を基準に考えたときにそれに似ている楽曲の度合いを指し、楽曲集合の確率的生成モデルからその楽曲の音楽要素の生成確率を計算することで推定する。

本稿では、類似度とありがち度について、楽曲を総合的に扱うのではなく、音楽要素単位に焦点を当てて推定する。将来的に、総合的な類似度やありがち度を推定するために

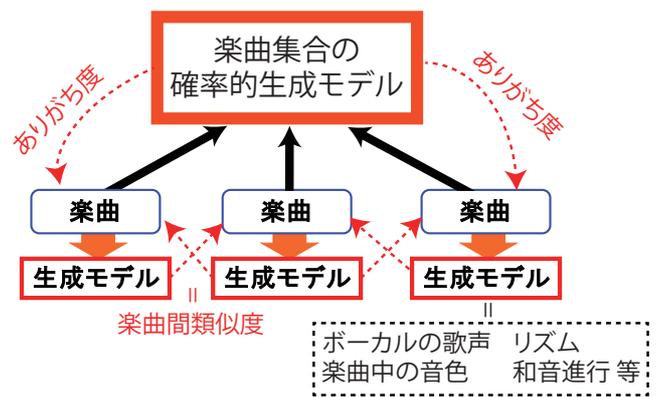


図1 確率的生成モデルに基づく類似度とありがち度の概要図。

は、「音楽要素単位（歌声、リズム等）」、「出現範囲（フレーズ単位、楽曲単位等）」、「出現パターン（順序、Aメロやサビ等の楽曲構造等）」の組み合わせや階層構造等を考慮する必要があると考えられる。そのような展開を見据え、今回提案する確率的生成モデルからの生成確率に基づく類似度は、各音楽要素の類似度を統合しやすい利点がある。また、楽曲全体だけでなく、曲の断片にも適用できる可能性があるため、楽曲制作の支援へも応用しやすい。

3. 音楽要素の確率的生成モデルの構築

ボーカルの歌声、楽曲中の音色、リズムに関しては、それぞれを表現する音響特徴量を時刻毎に推定した後、ベクトル量子化によって離散化して扱う。そのように離散化されたシンボルに基づいて、楽曲毎に潜在的ディリクレ配分法（Latent Dirichlet Allocation: LDA）[11]でモデル化する。

和音進行に関しては、能動的音楽鑑賞サービス Songle [12]のコード認識結果を用いて、可変長 Pitman-Yor 言語モデル（VPYLM）[13]を和音進行解析に発展させて学習した [14,15]。 n グラムモデルの n （和音進行の長さ）が無制限化されていることで、各和音に対して適切な n の事後分布を推定できる。

3.1 歌声の類似度・ありがち度

歌手名同定 [16-18]で従来用いられている分析条件により音響特徴量を推定し、確率的生成モデルを構築する。

3.1.1 歌声特徴量の抽出

伴奏が混在した音楽音響信号からボーカルの特徴を抽出するために、まずは混合音中で最も優勢な音高を推定する PreFEst [19]によってメロディーの F_0 を推定する。その後、20次倍音までを用いた正弦波合成によって再合成し、その信号から分析フレーム長32msでLPMCC（12次元）と ΔF_0 （1次元）を10ms毎に推定する。最後に、歌声・非歌声GMMを用いた高信頼度フレーム選択 [16]によって、歌声らしさが高いフレーム上位15%を選択して用いる。

音楽音響信号には16kHzのモノラル音響信号を用い、LPMCCはLPCスペクトルからMFCCを計算することで

推定した。その際、LPCの次数は25、メルフィルタバンク数は15とした。 ΔF_0 は、50msについて算出した。

歌声GMMと非歌声GMMは、RWC研究用音楽データベース(ポピュラー音楽)[20]から100曲(日本語80曲と英語20曲)を用いて推定した。それぞれ混合数を32、混合重みのディリクレ分布のハイパーパラメータの値を1.0としてベイズ推定を行った。それによって、歌声GMMでは混合数12、非歌声GMMは混合数27に縮退して得られた。

3.1.2 確率的生成モデル

歌声におけるLDAの基本的な処理手順は、我々の過去の研究[17,18]と同様である。

音響特徴量について、次元毎に平均を引いて標準偏差で割る正規化を行った後、 k -meansアルゴリズムを用いてベクトル量子化を行う。ここでプロトタイプは、RWC研究用音楽データベース(ポピュラー音楽)100曲から $k=100$ として学習した。LDAのトピック数は100として、周辺化Gibbsサンプラーを用いて学習した[21]。トピック分布のハイパーパラメータはすべて1、シンボル分布のハイパーパラメータはすべて0.1とした。

3.1.3 類似度・ありがち度の計算

LDAによる分析では、各楽曲におけるトピック分布(トピック混合比の分布)と各トピックにおけるシンボル分布(シンボルのユニグラム確率の分布)を得る。これらは、ベイズ推定においてはディリクレ分布を仮定し、その事後分布を推定することになる[11,21]。

ここで、ディリクレ分布のパラメータは、各トピックもしくは各シンボルの仮想的な観測回数とみなすことができる。これを利用して、楽曲毎のトピック分布から楽曲集合のトピック分布を推定する。すなわち、楽曲毎のトピックのディリクレ分布のパラメータを足し合わせる(ハイパーパラメータは1度だけ足す)ことで、楽曲集合のトピックのディリクレ分布のパラメータを得る。このようにすることで、各トピックの意味(シンボル分布)を保存したまま、楽曲集合のトピック分布を得ることができる。

類似度とありがち度は、このようにして得られた楽曲毎のモデルと楽曲集合のモデルから、ある楽曲の生成確率を計算することで推定する。この際、シンボルの観測数(フレーム数)で正規化を行った。

3.2 音色の類似度・ありがち度

音声認識で広く用いられている特徴量と分析条件により音響特徴量を推定し、確率的生成モデルを構築する。

3.2.1 音色特徴量の抽出

音色特徴量は、分析フレーム長25msとしてMFCC(12次元) Δ MFCC(12次元) Δ パワー(1次元)を10ms毎に算出し、フレーム数の15%をランダムに選択した。

16kHzのモノラル音響信号を用い、プリアンファシス係数を0.97として高域強調を行った。MFCCの算出におい

てメルフィルタバンク数は15とし、リフタリング係数を22とした。 Δ 特徴量は、50msについて算出した。

3.2.2 確率的生成モデル

歌声同様、音響特徴量について、平均と標準偏差を用いて正規化を行った後、 $k=64$ とした k -meansアルゴリズムでベクトル量子化する。プロトタイプは、RWC研究用音楽データベース(ポピュラー音楽)100曲から学習した。LDAの分析条件は歌声の場合(3.1.3)と同様である。

3.2.3 類似度・ありがち度の計算

歌声による場合と同様(3.1.3) LDAによってモデル化し、楽曲毎のトピック分布から楽曲集合のトピック分布を得る。このようにして得られた生成モデルからの各楽曲の音楽要素の生成確率を計算して、フレーム数で正規化することで、類似度もしくはありがち度を得る。

3.3 リズムの類似度・ありがち度

音楽類似度[22,23]で従来用いられている分析条件により音響特徴量を推定し、確率的生成モデルを構築する。

3.3.1 リズム特徴量の抽出

リズム特徴量には、分析区間を約6秒とし、Fluctuation Pattern (FP)[22,23]1200次元を約3秒毎に算出した。FPの具体例を図2に示す。また、RWC研究用音楽データベース(ポピュラー音楽)[20]100曲で主成分分析を行い、累積寄与率95%となる79次元までの射影行列を得る。

FP算出時においては、分析フレーム長約23.2msとしてハニング窓を用いたFFTを約11.6msに行い、Barkスケールに基づいたフィルタバンク分析を行う。それぞれのフィルタ出力(フィルタ数20)における振幅変調を、上述した6秒の分析フレーム長のFFTによって求め、変調周波数0~10Hzの60ピンについて計1200次元(=20×60)の特徴量を得る。最後に、これを上記の処理で得られた射影行列を用いて次元削減する。ただし以上の処理では、周波数マスキングや等ラウドネス曲線に基づく補正など、心理音響モデルに基づいた補正処理が伴う(詳細、[22,23])。

11.025kHzのモノラル音響信号に対し、MATLAB用のMA (Music Analysis) toolbox [23]を用いて推定した。

3.3.2 確率的生成モデル

歌声同様、音響特徴量について、平均と標準偏差を用いて正規化を行った後、 $k=64$ とした k -meansアルゴリズムでベクトル量子化する。プロトタイプは、RWC研究用音楽データベース(ポピュラー音楽)100曲から学習した。LDAの分析条件は歌声の場合(3.1.3)と同様である。

3.3.3 類似度・ありがち度の計算

歌声による場合と同様(3.1.3) LDAによってモデル化し、楽曲毎のトピック分布から楽曲集合のトピック分布を得る。このようにして得られた生成モデルからの各楽曲の音楽要素の生成確率を計算して、フレーム数で正規化することで、類似度もしくはありがち度を得る。

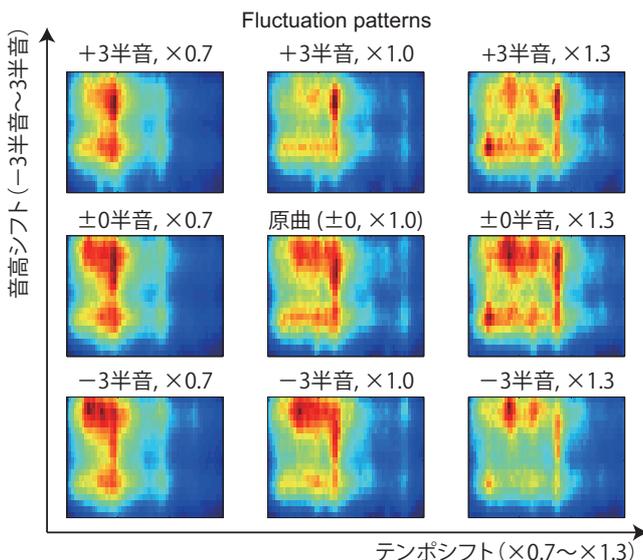


図 2 Fluctuation Pattern (FP) の例。一曲中の FP のメディアンを取った値。同一楽曲の音高をシフト (周波数軸伸縮) させた場合とテンポをシフト (WSOLA による速度変化) させた場合の FP の違い。

3.4 和音進行の類似度・ありがち度

和音進行認識 [12] で従来用いられている分析条件により音響特徴量を推定し、確率的生成モデルを構築する。

3.4.1 音楽音響信号に対する和音進行認識

9 種類の代表的な和音の種類 (major, major 6th, major 7th, dominant 7th, minor, minor 7th, half-diminished, diminished, augmented) に加え、ベース音が異なる major 和音の変種 5 種類 ($/2$, $/3$, $/5$, $/b7$, $/7$) の、計 14 種類 ($= 9 + 5$) の和音に対応し、それぞれに根音は 12 種類存在するので、全部で 168 種類 ($= 14 \times 12$) の和音を推定できる。また、和音が存在しない区間も推定できる [12]。

本手法では、クロマ特徴量に基づく和音推定用 HMM と、3 種類のスケール (major, natural minor, harmonic minor) に対応した調 (キー) 推定用 HMM を組み合わせた。その際、調ごとに使用される和音が大きく異なることを文献 [24] のように考慮し、一旦、調推定用 HMM に基づいて調の事後分布を推定し、それを和音推定用 HMM による各和音の確率の重みとして用いた。和音変化は拍の場所のみで起き、かつ、小節の先頭で起きる可能性が高いように Viterbi 探索で考慮した。

3.4.2 確率的生成モデル

和音進行の推定結果の正規化及び整形を行い、それに基づいてモデルパラメータを学習する。まず調 (キー) が C になるようにルート音をシフトさせて、 b を \sharp に統一する正規化を行う。続いて、ベース音の違いは考慮せず同一の和音として扱い、推定結果で同じ和音が連続した場合は一つにまとめる整形を行った。

学習に利用した和音は、ベース音の違いを除いた 8 種類 (major, major 6th, major 7th, dominant 7th, minor,

minor 7th, diminished, augmented) であり、それぞれに根音は 12 種類存在し、和音がない区間も扱うため、全部で 97 種類 ($= 8 \times 12 + 1$) の語彙数で学習した。

3.4.3 類似度・ありがち度の計算

楽曲集合の生成モデルは、全曲分の和音進行を用いて、前述のとおり VPYLM を用いて学習した。しかし、楽曲毎の生成モデルについては、楽曲毎の和音数が少ないために、ベイズモデルのみでは適切な学習が行えなかった。ベイズモデルは、観測データの裏にある真の生成過程を推定しようとするため、ある楽曲のデータのみを用いた場合でも、全体の楽曲をカバーするような生成過程の可能性も捨てずに保持している。そのため、学習データ数が少ないと、楽曲の特徴を十分に反映したモデルが学習されにくい。つまり、非常になだらかな事後分布が得られるため、他の楽曲との類似度計算に支障が出る。

そこで、楽曲毎に tri-gram ($n = 3$) を最尤推定し、楽曲集合から学習した VPYLM と統合することで対処した。具体的には、それぞれのモデルから計算される生成確率を、最尤推定した tri-gram を $1.0 - 10^{-5}$ 、楽曲集合から学習した VPYLM を 10^{-5} の比率で混合した。これは、より楽曲の特徴を十分に反映したモデルが学習できるように、あえてオーバーフィットするような処理となっている。

個々の楽曲毎に学習したモデルと全曲で学習したモデルから、各曲のパープレキシティを計算し、その逆数から楽曲中のそれぞれの和音に対する平均的な生成確率を算出し、類似度とありがち度を推定する。

4. 実験および結果

日本の音楽チャートであるオリコン^{*3}で 2000~2008 年までの上位 20 位以内に登場した楽曲、3278 曲を対象として、生成モデルの構築を行った。本稿では、各音楽要素の生成モデルに基づく類似度の推定に関する実験 A と、ありがち度の推定に関する実験 B の 2 種類を実施した。

実験 A では、対象楽曲中、楽曲数の多い上位 20 アーティスト (表 1) の楽曲 463 曲を対象として、それぞれの類似度を推定した。また、実験 B では、RWC 研究用音楽データベース (ポピュラー音楽) [20] を対象に、各音楽要素におけるありがち度の推定を行った。

4.1 実験 A: 楽曲間類似度の推定

図 3 から図 6 までに、それぞれの音楽要素に関して、楽曲間類似度を算出した結果を示す。横軸は各楽曲の生成モデルを意味し、縦軸は各楽曲の音楽要素を意味する。ここで左図は類似度を直接描画したものであり、右図は結果の見易さのために、各生成モデルから推定した類似度が高い上位 10% (46 曲) だけを黒く塗りつぶした描画である。

*3 <http://www.oricon.co.jp/>

表 1 実験 A で対象とした楽曲数の多い上位 20 アーティスト

ID	歌手名	性別	曲数
A	浜崎 あゆみ	女	33
B	B'z	男	28
C	モーニング娘。	女 (複数)	28
D	倉木 麻衣	女	27
E	倅田 来未	女	25
F	BoA	女	24
G	EXILE	男 (複数)	24
H	L'Arc ~ en ~ Ciel	男	24
I	愛内 里菜	女	24
J	w-inds.	男 (複数)	23
K	SOPHIA	男	22
L	中島 美嘉	女	22
M	CHEMISTRY	男 (複数)	21
N	Gackt	男	21
O	GARNET CROW	女	20
P	TOKIO	男 (複数)	20
Q	ポルノグラフィティ	男	20
R	平井 堅	男	20
S	Every Little Thing	女	19
T	GLAY	男	19
合計		男 11 - 女 9	463

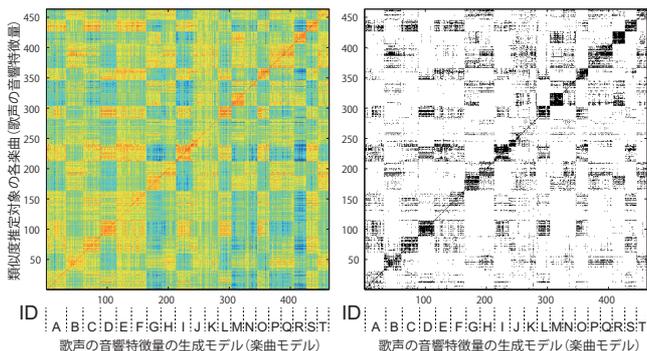


図 3 歌声特徴に基づく類似度 (左) とその上位 10% (右)

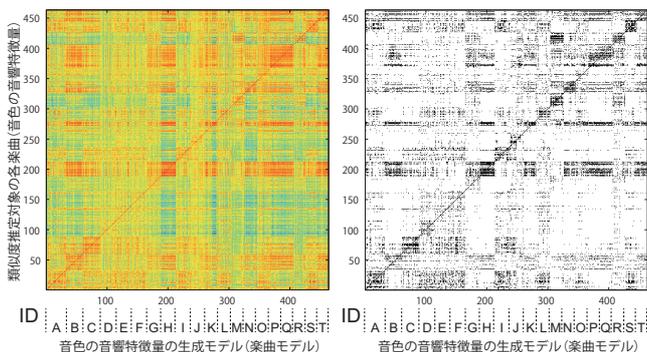


図 4 音色特徴に基づく類似度 (左) とその上位 10% (右)

結果からは、歌声や楽曲中の音色ではアーティスト内での類似度が高くなる傾向があり、リズムや和音進行では同一アーティストであっても類似度が高くはならない傾向があることが読み取れる。しかし、図 5 と図 6 の矢印で示した箇所のように、リズムや和音進行に関しても、アーティ

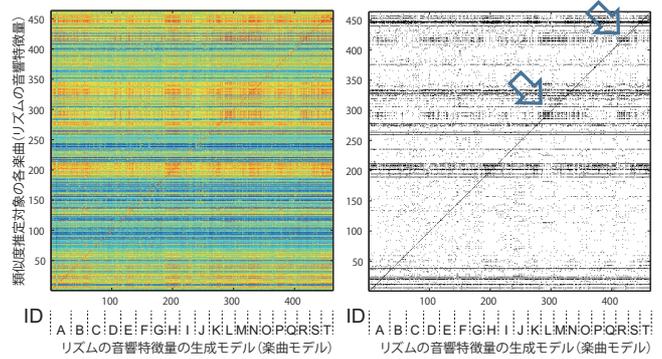


図 5 リズム特徴に基づく類似度 (左) とその上位 10% (右)。矢印はアーティスト内での類似度が高い箇所。

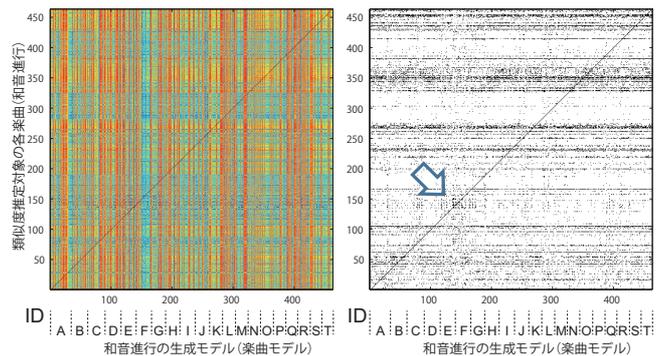


図 6 和音進行に基づく類似度 (左) とその上位 10% (右)。矢印はアーティスト内での類似度が高い箇所。

スト内での類似度が高くなる場合があった。

上記の結果が直観的に理解しやすいことから、構築した生成モデル及び推定された楽曲間類似度が適切であることを示唆している。

4.2 実験 B: ありがち度の推定

図 7 に、歌声・音色・リズムに関する楽曲集合のトピック混合比を示す。音楽要素毎に、各トピックの背後にあるシンボル分布が異なるため、トピック混合比の直接的な比較に意味はないが、全体的になだらかなトピック分布となっていることが分かる。これらのトピックモデルに基づいた、RWC 研究用音楽データベース (ポピュラー音楽) 100 曲分に関する各音楽要素のありがち度を図 8、図 9、図 10 に示す。また、表 2 にその詳細を示す。実際にこれらを聴取したところ、歌声特徴においては女性歌手 (No.60, 70, 20) の歌声、音色特徴においては楽器の音色、リズム特徴においてはドラムの音やそのリズムが類似している印象を受けた。これらは定性的な印象であるが、ありがち度の高い事象を明らかにする定量的な方法を含めて、手法の性能評価が今後の検討課題である。

また図 11 に、RWC 研究用音楽データベース (ポピュラー音楽) 100 曲分に関する和音進行のありがち度を示す。表 3 は、その上位 5 曲において出現する、一般的に使用されやすい和音進行の例である。表 3 からは、ありがち度が

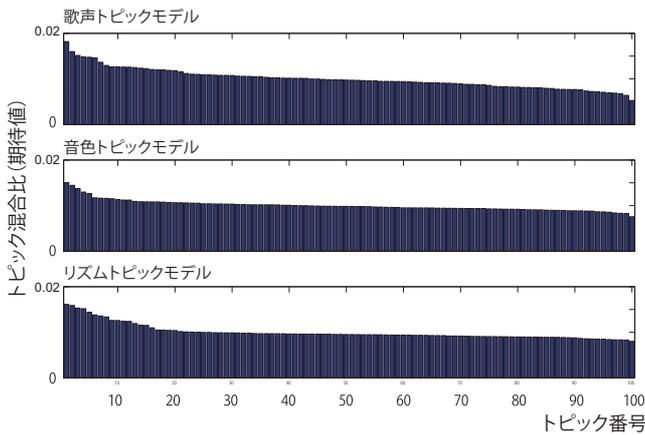


図 7 ポピュラー音楽 3728 曲から学習した歌声・音色・リズムのトピック混合比。

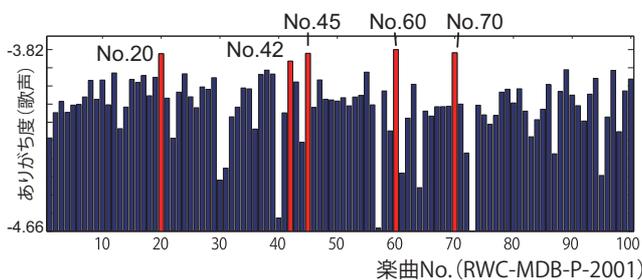


図 8 RWC 研究用音楽データベース (ポピュラー音楽) 100 曲における歌声に関するありがち度。生成モデルはポピュラー音楽 3728 曲から学習。

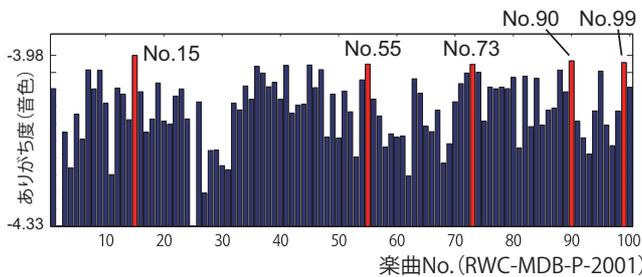


図 9 RWC 研究用音楽データベース (ポピュラー音楽) 100 曲における音色に関するありがち度。生成モデルはポピュラー音楽 3728 曲から学習。

高い楽曲は「C F G C」や「Am F G C」をはじめとして、よく使用される和音進行を含むことがわかる。

5. おわりに

本稿では、膨大な音楽コンテンツ間の類似度を人々が知ることができる (可視化する) 情報環境を実現することを目指して、楽曲間の「類似度」と、楽曲と楽曲集合の類似度としての「ありがち度」を推定する手法を提案した。個々の楽曲毎に学習したモデルと全曲から学習したモデルから、楽曲における各音楽要素の生成確率を計算することで、それぞれ楽曲間の類似度とありがち度を推定した。

提案手法により、「ボーカルの歌声」「楽曲中の音色」「リ

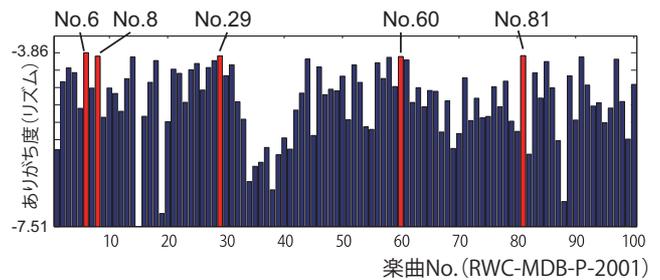


図 10 RWC 研究用音楽データベース (ポピュラー音楽) 100 曲におけるリズムに関するありがち度。生成モデルはポピュラー音楽 3728 曲から学習。

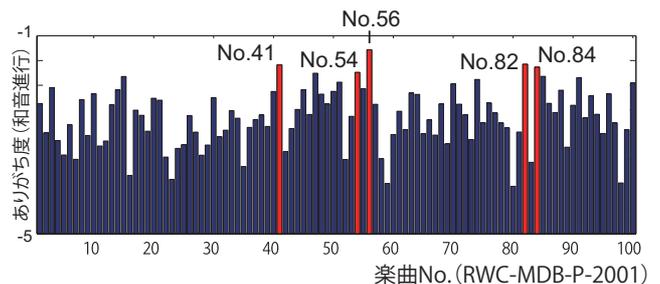


図 11 RWC 研究用音楽データベース (ポピュラー音楽) 100 曲における和音進行に関するありがち度。生成モデルはポピュラー音楽 3728 曲から学習。

ズム」「和音進行」の類似度を個別に推定することができ、歌声と音色がアーティスト内での類似度が高くなる傾向があること、リズムや和音進行では同一アーティストであっても類似度が低くなる傾向があることを示した。

自動推定のみでは推定誤りが発生し、モデルパラメータの推定精度を下げる可能性もあるが、本稿ではそのような誤りを含めた結果を評価した。将来的には、技術的な推定性能向上や、能動的音楽鑑賞サービス Songle [12] におけるユーザによる誤り訂正を活用する発展が考えられ、推定誤りの減少はモデル化の精度向上につながる可能性がある。

今後は、個々の音楽要素の類似度やありがち度の統合に関する検討が課題である。また、類似度に基づいて新たなコンテンツと出会える鑑賞支援技術を実現していくことで、主体的にコンテンツと出会って鑑賞できるようにする。さらに、「ありがち度」の高い事象 (例えばコード進行やジャンルごとの慣例的な事象) を制作における共有の知として活用できる創作支援技術を実現していくことで、誰もが気軽にコンテンツ創作を楽しめるようにすることも今後の展望として見据えている。

謝辞 本研究の一部は JST CREST「OngaCREST プロジェクト」の支援を受けました。また本研究では RWC 研究用音楽データベースを使用しました。

表 2 実験 B において各音楽要素のありがち度が高い上位 5 曲

音楽要素 (順位)	No.	歌手名	性別
歌声 (1)	60	飯島 柚子	女
歌声 (2)	70	松坂 珠子	女
歌声 (3)	45	森元 康介	男
歌声 (4)	20	市川 えり	女
歌声 (5)	42	森元 康介	男
音色 (1)	15	小澤 克之	男
音色 (2)	90	井口 慎也	男
音色 (3)	99	井口 慎也	男
音色 (4)	55	鍋木 朗子	女
音色 (5)	73	西 一男	男
リズム (1)	6	オリケン	男
リズム (2)	81	ドナ・パーク	女
リズム (3)	29	西 一男	男
リズム (4)	8	フィーバース	女 (2 名)
リズム (5)	60	M&Y	女 (2 名)
和音進行 (1)	56	橋本 まさし	男
和音進行 (2)	82	井口 慎也	男
和音進行 (3)	41	小澤 克之	男
和音進行 (4)	84	井口 慎也	男
和音進行 (5)	54	凜	女

表 3 和音進行のありがち度が高い上位 5 曲の和音進行の一部 (実験 B)。下線は一般的に使われやすいと考えられる和音進行。「...」は前後に和音が続くことを表す。

No.	正規化・整形後の和音進行の推定結果 (部分的に表示)
56	... F:maj C:maj G:maj F:maj C:maj G:maj ...
82	... G:maj C:maj F:maj G:maj C:maj F:maj E:maj A:min F:maj G:maj C:maj F:maj ...
41	F:maj C:maj F:maj C:maj F:maj ...
84	... G:maj C:maj F:maj G:maj C:maj F:maj ...
54	G:maj F:maj G:maj F:maj G:maj ...

参考文献

[1] 後藤真孝, 平田圭二: 音楽情報処理の最近の研究, 日本音響学会誌, Vol. 60, No. 11, pp. 675-681 (2004).
 [2] Pardo, B.(ed.): *Special issue: Music information retrieval*, Communications of the ACM, Vol. 49, No. 8, pp. 28-58 (2006).
 [3] Casey, M., Veltkamp, R., Goto, M., Leman, M., Rhodes, C. and Slaney, M.: Content-Based Music Information Retrieval: Current Directions and Future Challenges, *Proceedings of the IEEE*, Vol. 96, No. 4, pp. 668-696 (2008).
 [4] Downie, J. S.: The music information retrieval evaluation exchange (2005-2007): A window into music information retrieval research, *Acoust. Sci. & Tech.*, Vol. 29, pp. 247-255 (2008).
 [5] Downie, J. S., Byrd, D. and Crawford, T.: Ten Years of ISMIR: Reflections on Challenges and Opportunities, *Proc. ISMIR 2009* (2009).
 [6] 吉井和佳, 後藤真孝: 音楽推薦システム, 情報処理, pp. 751-755 (2009).
 [7] Song, Y., Dixon, S. and Pearce, M.: Survey of Music Recommendation Systems and Future Perspectives, *Proc. CMMR 2012*, pp. 395-410 (2012).

[8] Knees, P. and Schedl, M.: A Survey of Music Similarity and Recommendation from Music Context Data, *ACM Trans. on Multimedia Computing, Communications and Applications*, Vol. 10, No. 1, pp. 1-21 (2013).
 [9] Hamasaki, M., Goto, M. and Nakano, T.: Songrium: A Music Browsing Assistance Service with Interactive Visualization and Exploration of a Web of Music, *Proc. WWW 2014* (2014).
 [10] 後藤真孝: 未来を切り拓く音楽情報処理, 情報処理学会音楽情報科学研究会研究報告, 2013-MUS-99, No. 33, pp. 1-9 (2013).
 [11] Blei, D. M., Ng, A. Y. and Jordan, M. I.: Latent Dirichlet Allocation, *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022 (2003).
 [12] 後藤真孝, 吉井和佳, 藤原弘将, Mauch, M., 中野倫靖: Songle: 音楽音響信号理解技術とユーザによる誤り訂正に基づく能動的音楽鑑賞サービス, 情報処理学会論文誌, Vol. 54, pp. 1363-1372 (2013).
 [13] 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, pp. 4023-4032 (2007).
 [14] 吉井和佳, 後藤真孝: 和音進行解析のための語彙フリー無限グラムモデル, 情報処理学会研究報告 音楽情報科学研究会, 2011-MUS-91, pp. 1-10 (2013).
 [15] Yoshii, K. and Goto, M.: A Vocabulary-Free Infinity-Gram Model for Nonparametric Bayesian Chord Progression Analysis, *Proc. ISMIR 2011*, pp. 645-650 (2014).
 [16] Fujihara, H., Goto, M., Kitahara, T. and Okuno, H. G.: A Modeling of Singing Voice Robust to Accompaniment Sounds and Its Application to Singer Identification and Vocal-Timbre-SimilarityBased Music Information Retrieval, *IEEE Trans. on ASLP*, Vol. 18, No. 3, pp. 638-648 (2010).
 [17] 中野倫靖, 吉井和佳, 後藤真孝: トピックモデルを用いた歌声特徴量の分析, 情報処理学会研究報告 音楽情報科学研究会, 2013-MUS-100, pp. 1-7 (2013).
 [18] Nakano, T., Yoshii, K. and Goto, M.: Vocal Timbre Analysis Using Latent Dirichlet Allocation and Cross-Gender Vocal Timbre Similarity, *Proc. ICASSP 2014* (2014).
 [19] Goto, M.: A Real-time Music Scene Description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals, *Speech Communication*, Vol. 43, No. 4, pp. 311-329 (2004).
 [20] 後藤真孝, 橋口博樹, 西村拓一, 岡 隆一: RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, 情報処理学会論文誌, Vol. 45, No. 3, pp. 728-738 (2004).
 [21] Griffiths, T. L. and Steyvers, M.: Finding scientific topics, *Proc. of the National Academy of Sciences of the United States of America*, Vol. 1, pp. 5228-5235 (2004).
 [22] Pampalk, E., Rauber, A. and Merkl, D.: Contentbased Organization and Visualization of Music Archives, *Proc. ACM MM '02*, pp. 570-579 (2002).
 [23] Pampalk, E.: *Computational Models of Music Similarity and Their Application to Music Information Retrieval*, Ph.D. Dissertation, Vienna Inst. of Tech. (2006).
 [24] Mauch, M. and Dixon, S.: Simultaneous Estimation of Chords and Musical Context from Audio, *IEEE Trans. on ASLP*, Vol. 18, pp. 1280-1289 (2010).