

楽曲群を組織化した空間におけるユーザーの音楽的嗜好の推測

Estimating the User's Preference in an Organized Music Archive Space

小林 広司*¹ 岡 夏樹*²
Koji Kobayashi Natsuki Oka

*¹京都工芸繊維大学 大学院工芸科学研究科 情報工学専攻

Division of Information Science, Graduate School of Science and Technology, Kyoto Institute of Technology

*²京都工芸繊維大学 大学院工芸科学研究科 情報工学部門

Department of Information Science, Graduate School of Science and Technology, Kyoto Institute of Technology

We need information which helps us to find pieces of music that we may like from large digital music archives. In this study, a music collection is organized into a three-dimensional space according to the sound characteristics of each piece. Then the evaluation score of each piece of music is calculated based on the listening history and the proximity in the space. We also conducted an experiment on the effect of listener's feelings on the preference of music.

1. 研究背景

近年 iPod などの大容量携帯音楽プレーヤーや iTunesMusicStore などの Web 上のミュージックストアが出現したことにより、ユーザーは多種多様な楽曲を手軽に入手できるようになった。つまりユーザーはあまり有名でない楽曲や、あるいは自分があまり知らない種類の楽曲も手軽に入手できるようになった。これはユーザーが自分の好みに合った未知の楽曲を発見したり、今まであまり知らなかった種類の音楽を好きになったりできる可能性が高くなったことを意味する。そのような中で膨大な楽曲群の中から自分が気に入る未知の楽曲を見つける手助けをするシステムが求められているのがわかる。

2. 本研究の目標

ユーザーが労力をかけずに楽曲を検索するという意味では自動選曲システムが最適である。ユーザーは何もせずともシステムが自動的にユーザーの好みの曲を推測して再生してくれる。もちろん運転中などユーザーが日々選曲している余裕がない場合は自動選曲は非常に有効である。しかし自動選曲システムによって楽曲を検索する場合、ユーザーはどうしても受身の姿勢になってしまう。そのような性質上、ユーザーが未知の楽曲を検索してその楽曲を好きになる確率が高いという観点では、自動選曲は必ずしも最適とは言えない。ユーザーが未知の楽曲を検索してその楽曲を好きになるには、ある程度ユーザーの能動性を許容し、検索すること自体に楽しみがあるシステムの方が適していると考えられる。つまりユーザーが能動的に楽曲を探す上で、その手掛かりを与えてくれるシステムが望ましい。

そこで本研究ではそのような能動的にユーザーが楽曲を検索できるシステムを意識し、楽曲自体の情報とユーザーが一部の楽曲に与える評価から、ユーザーに提示できる形で楽曲検索の手掛かりとなる情報を得ることを目標とする。

また未知の楽曲を検索するという観点から、有名な曲だけでなく、マイナーな楽曲をも探せる手掛かりであることが望ましい。

このような目標を達成するために本研究では以下の3点を行った。

- 音響的特徴（音色、ビート）に基づく楽曲群の組織化
- 楽曲群を組織化した空間内におけるユーザーの音楽的嗜好の推測と可視化
- 気分による音楽的嗜好の変化の観察

これによりユーザーは楽曲群が組織化された3次元空間を見ることができ、その空間の中で自分の好きそうな曲がどれで、好きではなさそうな曲がどれかを楽曲検索の手掛かりとすることができる。なお本研究では特に1点目、2点目に重点をおいている。

3. 楽曲群の組織化

この章では400曲の楽曲から音色、ビートという音響的特徴を抽出して楽曲群を3次元空間に組織化するまでの手順を述べる。また組織化した結果について定性的に評価する。

この章で得られるビート1次元、音色2次元の特徴量によって楽曲群は3次元空間に組織化される。

3.1 楽曲の数値化などの前処理

楽曲のデータは初め wav 形式の波形データで与えられる。楽曲から音響的特徴を抽出するためにはまず楽曲のデータを数値化する必要がある。本研究においては元の波形データを MFCC に変換する。なお窓長 30ms、シフト幅 10ms、用いた窓はハミング窓である。この処理によって一つの楽曲は数万個の13次元ベクトルの羅列として表現されるようになる。次にリード・イン、リード・アウト部として無音になっていることの多い曲の最初の1秒、最後の6秒はこの段階で切り落としておく。

3.2 音色の特徴抽出

音色の特徴抽出においては TreeQ アルゴリズム [1] を用いる。TreeQ アルゴリズムを MFCC 化した楽曲に適用すると時系列で変化していく楽曲の音色において、どれくらいの時間どんな音が鳴っていたかという情報を得ることができる。TreeQ アルゴリズムではまず学習データから特徴空間を区切る VectorQuantizationTree を生成する必要がある。本研究では学習用データとして「RWC 研究用音楽データベース：音楽ジャン

連絡先: 小林広司 京都工芸繊維大学 〒606-8585 京都市左京区松ヶ崎橋上町, Tel 075-724-7016, Fax 075-724-7029, m7622017@edu.kit.ac.jp

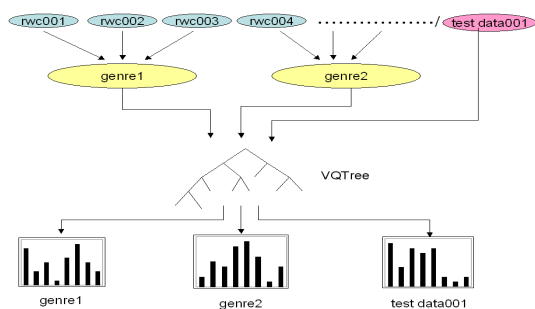


図 1: 学習データとテストデータのヒストグラム化

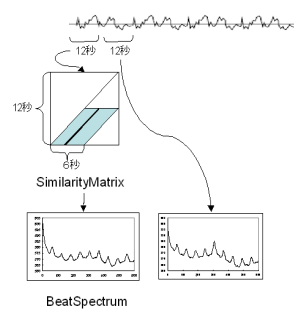


図 3: ビートの特徴抽出

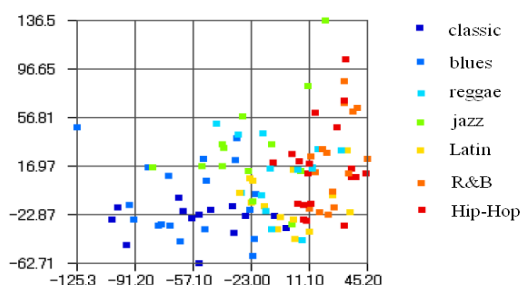


図 2: ジャンル別に色分けされた特徴抽出結果

ル」[2]の中から99曲(33ジャンル,各3曲ずつ)を選んで用い、学習データの正解カテゴリとしてジャンルを用いた。すなわちAというジャンルの3曲から得られるフレーム(個々の13次元ベクトル)はAというクラスに属するというのである。よってクラス数は33となる。学習データのフレームが、できるだけ同じクラスのものと同じ葉に、違うクラスのは違う葉に辿り着くようにVQTreeが生成される。本研究では葉の数が100になるまでVQTreeを生成した。これは13次元の特徴空間を100個に区切ることを意味する。

次に得られたVQTreeを用いて図1のようにして学習データとテストデータを100次元のヒストグラムとして表す。生成されたVQTreeに数値化されたテストデータ(選んだ400曲)を通すと、その楽曲のフレームのうちいくつかの葉にたどり着いたのかという100次元のヒストグラムが得られる。さらにこのヒストグラムを正規化する。

次にテストデータの正規化されたヒストグラムと各ジャンルの正規化されたヒストグラムの距離を求めることによって、テストデータを33次元のベクトルとして表現する。最後に得られた33次元のベクトルを主成分分析し、第一主成分スコアと第二主成分スコアをその楽曲の音色の特徴量として採用した。

学習データの正解カテゴリとしてclassicやjazzといったジャンル情報を用いたため、結果はジャンルを反映したものになった。そこでテストデータのうちジャンルがはっきりしているもの(classicやHip-Hopなど.Rockやpopは定義が曖昧である。)のみ取り出してジャンル別に色分けした特徴抽出結果を図2に示す。図2において横軸は第一主成分スコア、縦軸は第二主成分スコアである。

図2を見るとbluesとclassicの組とR&BとHip-Hopの組は座標にはっきりとした差があることがわかる。また,Latinやreggaeは特徴空間の右下あたりに集まっている.jazzは特徴空間の中央あたりに比較的たくさん集まっているように見え

るが、かなり離れた部分にも存在する。全体的に見てジャンル別に配置が分かれてはいるわけではないが、それぞれのジャンルの楽曲は特徴空間のうちある程度限られた領域に集まっている。

また同時に図2は各ジャンルとも特徴空間中にある程度広がりをもった配置となっているとも見れる。これは、例えば classic といってもいろいろな曲があるので当然のことである。各ジャンル内での楽曲の配置が音響的特徴を反映したものであるかを考える。Latinの領域の近くにある classic の曲は,classic 中においては比較的打楽器が多用されており、打楽器の音色が目立つ曲となっている。Latinは打楽器が多用されるジャンルであり、このような楽曲がLatinの近くにあるのは納得のいく話である。またjazzのうち縦軸の座標が大きいものはサクソなどのjazzらしい管楽器の音が目立つ曲になっている。反面クラシック領域の近くにあるジャズの曲はストリングス系の楽器が使われていてclassicalな雰囲気をもっているものもある。

このように同じジャンルの中でも楽曲ごとに違った特徴を持っていて、その個々の楽曲の音色の特徴を反映する傾向にある座標が得られている。

3.3 ビートの特徴抽出

ビートの特徴抽出においてはBeatSpectrumの技術[3]を用いる。BeatSpectrumとはMFCCの連続として表された楽曲の自己相関に注目してビートの特徴を抽出する方法である。数値化された波形データからBeatSpectrumの技術を用いてビートの特徴を抽出する様子を図3に示す。まずは楽曲を12秒毎に切り分けていく。切り分けられた12秒の部分から図3のようにして、その0秒から6秒までの繰り返しの強さがBeatSpectrumとして得られる。シフト幅10msで数値化しているため、10msの繰り返しの強さ、20msの繰り返しの強さ、...、6000msの繰り返しの強さの600次元のベクトルとして表されることになる。各12秒毎にこのBeatSpectrumが得られるので、1つの楽曲にBeatSpectrumが複数得られることになる。

1つの楽曲を1つのBeatSpectrumで表現するためにこの複数の600次元ベクトルを要素ごとに中央値をとる。中央値をとることによりその楽曲の全体的な特徴が現れる。このようにして1つの600次元ベクトルとして表されるようになったBeatSpectrumに主成分分析をかけ、その第一主成分スコアをビートの特徴量とした。

ビートの特徴抽出結果について述べる。400曲それぞれに対して得られたBeatSpectrumのうち、代表的な4パターンのグラフを図4に示す。図4を見るとBeatSpectrumの値が全体的に高い楽曲と低い楽曲、目立ったピーク値のある楽曲とない楽曲などにわかれることがわかる。

全体的にBeatSpectrumの値が高い楽曲は、エレキギターな

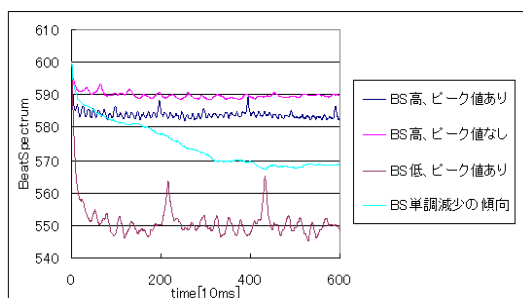


図 4: BeatSpectrum の代表的な 4 パターン

どのコード楽器やドラムなどの打楽器で同じような音が細かく刻まれている楽曲に多い。さらにその中でも目立ったピーク値がある楽曲は短いギターリフなどのある決まったフレーズが繰り返し使われている楽曲に多い。また BeatSpectrum の値が単調減少するような楽曲は、フルートのような管楽器やテノール歌手のソロのような、ビートを刻まない連続的な音のみで構成されている楽曲が多い。これは連続的に音を鳴らしているために時間的に近い部分ほど音が似ている傾向にあるからだと思う。

具体的な例を一つ挙げると、Hip-Hop の音楽は細かく刻まれるコード和音などがなく、ビートの要所のみで打楽器がならされるような楽曲が多いので、BeatSpectrum の値が全体的に低く、かつ目立ったピーク値のあるパターンであることが多い。

このような BeatSpectrum を主成分分析した結果、全体的に BeatSpectrum の値が高い楽曲ほど主成分スコアが低く、BeatSpectrum の値が低い楽曲ほど主成分スコアが高くなった。したがって主成分分析で得られた 1 次元の数値は打楽器音などが細かく刻まれている度合いであると理解することができる。しかし管楽器の音の目立つ楽曲は細かくビートが刻まれないにもかかわらず、BeatSpectrum の値がそれほど低くならないという事実もあり、そのように理解するうえでの問題点も残っている。またこの値はビートというもののごく一部の特徴を反映したものである。

4. ユーザーの音楽的嗜好推測実験

前章で得られた特徴空間におけるユーザーの音楽的嗜好の推測実験を行った。また気分によるその変化も観測した。この章では実験方法、音楽的嗜好の推測方法、及び推測結果について述べる。

4.1 実験方法

実験を行うために 400 曲の楽曲の中からランダムな順で選曲して再生し、その楽曲に対するユーザーの評価を観察するシステムを PDA 上に実装した。ユーザーは楽曲聴取時の気分を「楽しい気分」「憂鬱な気分」「悲しい気分」「イライラする」「集中したい」「元気を出したい」「気合を入りたい」「リラックスしたい」という 8 つの中から選択する（複数選択可）。楽曲に対するユーザーの評価としてはその楽曲の何%まで聞いて曲を飛ばしたか、あるいは最後まで楽曲を聞いたかを観察している。このようにして得られる「聴取した楽曲」「その時の気分」「楽曲に対するユーザーの評価」という 3 つで 1 セットの情報を基にシステムはユーザーの嗜好とその気分による変化を推測する。

今回は 20 代前半男性の被験者 5 人に PDA を渡してランダ

ムな順で再生される音楽を聞いてもらい、楽曲が気に入ったら最後まで聞き、気に入らなかつたら曲を飛ばすということを経験する。曲を飛ばす頻度にもよるが 4 時間のうちに被験者は楽曲を 70 曲から 180 曲ほど聞くことになる。

4.2 ユーザーの楽曲に対する評価スコア

400 曲の各楽曲はそれぞれ 9 つのスコアを持つ。8 つは前述した 8 つの気分に対応するスコアであり、このスコアが高いことはその楽曲がその気分合っている、もしくはシステムがそう推測していることを示す。また残りの一つはユーザーの嗜好に対応するスコアであり、これが高いことはその楽曲がユーザーの好みである、あるいはシステムがそう推測していることを示す。ユーザーが例えば「楽しい気分」「元気を出したい」気分を選択している時に楽曲 A を聞いた場合、楽曲 A、及びその近くの座標の楽曲の「楽しい気分」「元気を出したい」「ユーザーの嗜好」スコアが更新される。「ユーザーの嗜好」スコアは選ばれている気分に関係なく常に更新される。

次に具体的なスコアの更新方法を述べる。なおスコアの初期値は全て 0 である。ユーザーがある楽曲を x % 聞いて「曲を飛ばす」ボタンを押したとき、その楽曲のスコアのうち選ばれている気分のスコアと「ユーザーの嗜好スコア」の更新値 r_0 は次の式 (1) で与えられる。

$$r_0 = \frac{2(x - 50)}{100} \quad (1)$$

次に聞いた楽曲の近傍の楽曲のスコアを更新する。規則的に近傍の楽曲が 40 曲程度選ばれ、その楽曲のスコアに式 (2) で得られる更新値を加えて更新する。

この工程においてユーザーが評価していない楽曲のスコアが得られることになる。楽曲を 3 次元空間に組織化したのはこのように未評価の楽曲に対するスコアを得るためである。

40 曲のうち選ばれた楽曲 A と聞いた楽曲の距離が d_A の時、楽曲 A のスコアの更新値 r_A は以下の式で与えられる。

$$r_A = \frac{r_0}{1 + d_A} \quad (2)$$

このようにスコアを更新し 4 時間後に PDA を被験者から回収する。最後にそれぞれのスコアをそのスコアが更新された回数で割る。これはスコアの更新値 r の平均値を求めることにあたる。このようにして得られるスコア更新値の平均値をユーザーの楽曲に対する最終的な評価スコアとする。この最終的な評価スコアの分布をもって、ユーザーの音楽的嗜好の推測結果とする。

4.3 音楽的嗜好の推測結果

図 5 に被験者 5 人それぞれの嗜好スコアの分布を示す。図 5 において色のついた点一つ一つが楽曲を示している。赤い点で表される楽曲ほどユーザーの好みである、あるいはシステムがそのように推測した楽曲であることを示す。

被験者 1~3 に共通して言えるのは、 x の値が低く、 y の値が高い部分のスコアが高いということである。この領域に多く含まれる楽曲はロックやポップスなどの現代一般的に良く聞かれている種類の音楽である。

しかし被験者 4 は被験者 1~3 と違い、この領域の楽曲のスコアが低く、逆に y の値が低い領域の楽曲のスコアが比較的高い。 y の値が低い領域には classic 音楽や blues 音楽が多く含まれる領域である。被験者 4 にアンケートしたところ、日頃から classic 音楽を好んで聞くということであるので、それが嗜好スコアの分布に反映されていると言える。

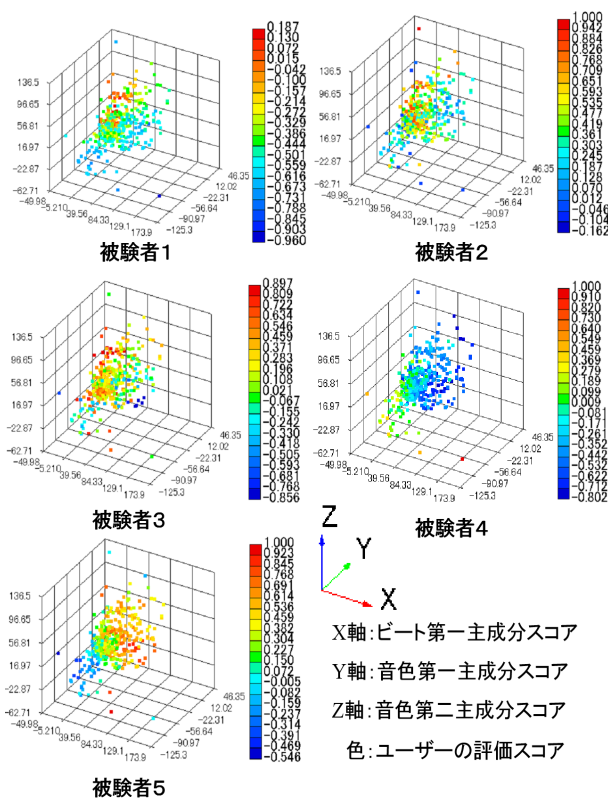


図 5: 音楽的嗜好の推測実験結果

また被験者 5 は x の値が高く, y の値が高い部分のスコアが高い。

4.4 音楽的嗜好の気分による変化

被験者の音楽的嗜好の気分による変化は, 気分スコアを比べることによって観察した。図 6 に被験者 1 の「楽しい気分」「集中したい」「イライラする」気分スコア及び嗜好スコアの分布を示す。図 6 を見ると「イライラする」という気分を選択している時は classic や blues がたくさんある領域 (y 値低, z 値低) のスコアが比較的高くなっているのがわかる。これは他の気分スコア, 及び嗜好スコアとは異なった傾向である。しかしそれ以外の気分スコア間ではあまりはっきりした差は見られない。

被験者 5 人全体を通して, 選ばれた気分によってスコアの分布がはっきり変わることはあまり見受けられなかった。

このようにはっきりした差が見られなかったことから, 被験者に自分で気分を選ばせる手法はあまり有効でないことがわかる。被験者は自分で明確に気分を把握しているわけではなく, 仮に明確に気分を把握していたとしてもチェックボックスという表現形式では表現力に乏しいからである。

5. おわりに

2 章で述べた目標に基づき, ユーザーに対して提示する楽曲検索の手掛かりとして図 5 のような情報が得られた。図 5 において各被験者ごとにスコアの高い部分と低い部分に分かれ, この図を見て自分の好みそうな楽曲がどこにあるかなどを知ることができる。しかし手掛かりとしての有効度にはまだまだ課題を残している。

まず課題として挙げられるのは, 各軸が何を表しているのが理解しがたいことである。3.2 節で述べたように音色の特徴を

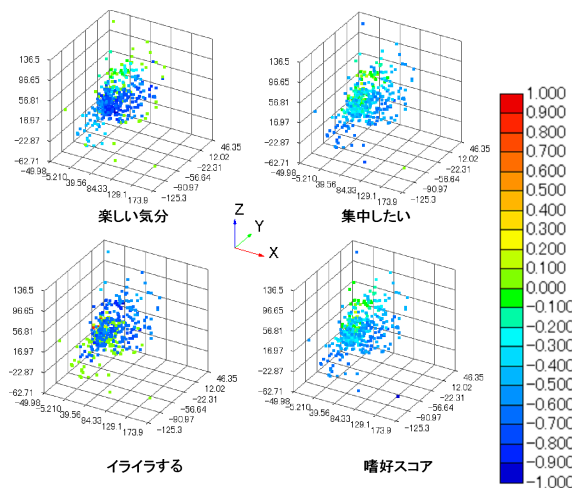


図 6: 被験者 1 の気分スコア

ある程度反映した楽曲群の組織化がなされてはいるが, 図 4 において横軸が何を表し, 縦軸が何を表しているかを明言することはできない。それぞれの軸が楽曲のどのような特徴を表しているのかを明言できる形で特徴抽出することができれば, よりユーザーにとって有効な手掛かりとなる。

また特徴量の選び方と抽出方法を改良する必要がある。現段階では音色, ビートという特徴量に従って楽曲群を組織化しているとは言うものの, 例えば 3.3 節で述べたように, ビートに関してはその一面を捉える特徴抽出をしたに過ぎない。図 3 の BeatSpectrum のグラフにはその一面以外の情報もあるので, その情報を抽出して数値化することができればより多面的にビートと言うものを捉えられるようになる。

また今回は被験者に対して未聴取の楽曲を 4 時間聞いてもらって図 5 の情報を得たが, 時間短縮のために予めユーザーが好きな楽曲を何曲か入力してもらい, それによって得られる分布を初期値とするなどの方法も考えられる。

謝辞

本研究は RWC 研究用音楽データベース (音楽ジャンル) [2] を利用しました。

参考文献

- [1] T.Foote:Content-Based Retrieval of Music and Audio,Multimedia Storage and Archiving Systems II, Proceedings of SPIE 3229, pp.138-147,1997.
- [2] 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: 音楽ジャンルデータベースと楽器音データベース”, 情報処理学会 音楽情報科学研究会 研究報告 2002-MUS-45-4, pp.19-26, 2002.
- [3] T.Foote,Shingo Uchihashi:The Beat Spectrum:A New Approach To Rythum Analysis,Proceedings of the International Conference on Multimedia and Expo 2001 (ICME), Tokyo, Japan. August, pp.22-25, 2001.