

楽曲群の組織化とユーザーの音楽的嗜好の可視化 Organization of Music Archives and Visualization of User's Preference

小林 広司†
Koji Kobayashi

岡 夏樹†
Natsuki Oka

1. 研究背景

近年 iPod などの大容量携帯音楽プレーヤーや iTunes などの web 上のミュージックストアが定着したことにより、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。そのような状況下でユーザーが自分に適した未知の楽曲を発掘できる可能性が高まっていると言える。しかしそのような膨大な量の多種多様な楽曲からユーザーが自分に適した楽曲を見つけるのは容易ではない。そこでユーザーが好みの楽曲を見つけるのを支援するシステムが望まれている。

支援するシステムの一つは自動選曲システムである。自動選曲システムでは、システム内の選曲アルゴリズムに従って自動的に選曲を行う。このようなシステムはユーザーが一々選曲している余裕のない状況下では有効である。しかしユーザーは流れてきた楽曲が選曲された理由を知ることができず、どうしても受身になってしまう。ユーザーが新たに未知の楽曲を気に入ることができる可能性を高めるには、ユーザーが能動的に探索できるシステムが望ましいと考える。つまりユーザーが未知の楽曲を能動的に探索する手掛かりを与えるシステムが望ましい。

ここで楽曲探索システムの先行研究を紹介する。Hoashiらは楽曲を音響的に解析して楽曲の特徴として音色をベクトルで表現し、ユーザーの嗜好をベクトル表現したものと比較して楽曲をランク付けする手法を提案している[1]。しかしユーザーはその結果を見てもランク付けの根拠を理解することはできない。

また Elias らは楽曲のビートの特徴を抽出して楽曲群を2次元空間に組織化する手法を提案している[2]。しかしユーザーの嗜好情報を反映する機構は備わっていない。

2. 本研究の目標

本研究ではユーザーが能動的に楽曲を探索できるシステムを目指し、ユーザーが根拠を理解できる形で楽曲探索の手掛かりとなる情報を提示することを目標とする。

また未知の楽曲を探索するという観点から、あまり有名でない楽曲の探索も行えるシステムを目標とする。

これらの目標を達成するため本研究では以下2点を行う。

- 1: 音響的特徴に基づく楽曲群の組織化
- 2: 組織化した空間内におけるユーザーの音楽的嗜好の可視化

1点目については様々なジャンルから集めた400曲の楽曲を楽曲の音色、ビートという特徴に基づいて3次元空間への組織化を行う。音響的特徴に注目したのは、有名でない楽曲にも対応するためにはソーシャルフィルタリングよりも楽曲自体の音響波形を解析する手法の方が適して

いると考えるからである。

2点目については400曲の中からランダムな順で再生される楽曲を実験協力者に聞いてもらい、実験協力者の楽曲再生の様子を観察して、その観察結果をもとに実験協力者が聞いていない楽曲に対する評価を推測して可視化する。

未聴取の楽曲に対する評価の推測が精度良くなされれば、ユーザーは3次元空間に組織化された楽曲群を実際に見て、その中で自分がどのあたりの楽曲をどの程度好みそうかを知ることができ、これを楽曲探索の手掛かりとすることができる。

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と結果を述べる。3.2節で得られる音色の特徴を表す2次元ベクトルと3.3節で得られるビートの特徴を表す1次元の値によって楽曲群は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次元ベクトルの系列として表現されるようになる。

次に曲の最初の1秒、最後の6秒はリード・イン、リード・アウト部として無音などになっていることが多々あるので、その部分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[3]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: 音楽ジャンル」[4]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ことである。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、違うクラスのものと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。

生成された VQTree に数値化されたテストデータを通すと、その楽曲のフレームのうちいくつかの葉に辿り着い

†京都工芸繊維大学 大学院工芸科学研究科
Graduate School of Science and Technology, Kyoto Institute
of Technology

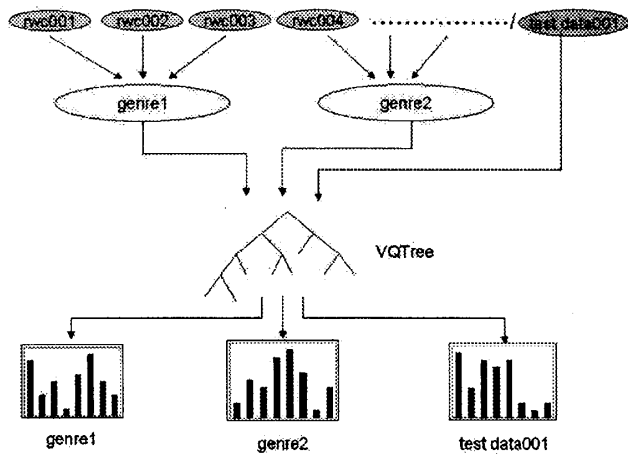


図1：学習データとテストデータのヒストグラム化

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。

次に正規化されたテストデータのヒストグラムと各ジャンルの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（寄与率 31.6%）と第二主成分（寄与率 26.5%）をその楽曲の音色の特徴量とした。

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。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。）に注目したところこれらの特徴量は classic, blues の組と Hip-Hop, R&B の組の間で違った傾向を見せるなどの結果が得られた。

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[5]を用いる。BeatSpectrum とは MFCC の系列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。BeatSpectrum を用いて数値化された波形データからビートの特徴を抽出する様子を図2に示す。

まずは楽曲を 12 秒毎に切り分けていく。BeatSpectrum の技術を用いて、切り分けられた 12 秒の部分から 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ビートの特徴が抽出される。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 (median) をとる。図3に 1 つの楽曲における複数の BeatSpectrum の中央値をとる例を示す。図3を見ると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とがわかる。

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（寄与率 93.9%）をビートの特徴量とした。この結果ビートの特徴量は 600 次元ベクトルの要素の全体的な高低を表す尺度となった。これは 600 次元の要素間の値の差が、複数のベクトル間の全体的な大きさの差に比べて微小

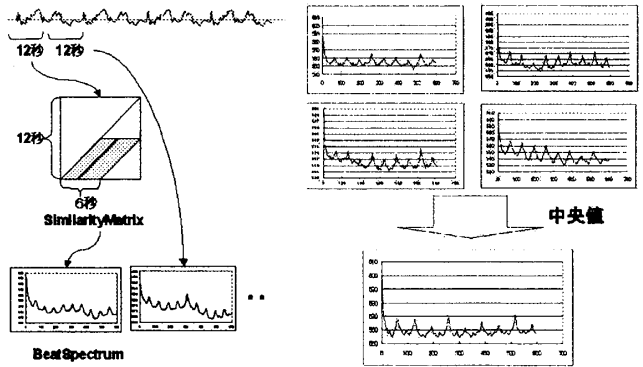


図2：ビートの特徴抽出 図3：BeatSpectrumの中央値

だったことに起因する。

これによってビートの特徴量は小刻みに打楽器やコード楽器が鳴っている度合いとなったが、この過程でビートに関する多くの情報が失われてしまった。システムの性能を上げるには特徴抽出法を検討しなおす必要がある。

4. ユーザーの音楽的嗜好の可視化実験

前章で得られた 400 曲の楽曲を組織化した 3 次元空間を用いてユーザーの音楽的嗜好の可視化実験を行った。またユーザーの気分による嗜好の変化についても観察したが、本稿においては気分による変化については言及しない。

この章ではユーザーの音楽的嗜好をシステムが推測して可視化する方法とその実験結果を示す。

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。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た。

実験方法は実験協力者 5 人に PDA を渡して音楽を聞いてもらい、楽曲が気に入ったら最後まで聞き、気に入らなかったら曲を飛ばすということを数日間のうちに 4 時間ほど行ってもらおうというものである。4 時間のうちに実験協力者が聞く楽曲数は楽曲を飛ばす頻度によって異なってくる。図4の可視化結果とともにそれぞれの実験協力者が 4 時間のうちに聴取した楽曲数を示す。

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ユーザーの嗜好に対応するスコアをもつ。このスコアが高ければその楽曲がユーザーの好みの楽曲であるということを示す。

このスコアはユーザーがその楽曲、あるいは組織化された空間においてその楽曲の近くにある楽曲を気に入ったと判断した時に上がる。逆に気に入らないと判断した時はスコアが下がる。

次に評価スコアの計算方法について述べる。まずは 1 曲聞くごとに得られるスコア r について説明する。

ユーザーがある楽曲を x % 聞いて曲を飛ばした時その楽曲の評価スコア r_0 は次の式(1)で与えられる。

$$r_0 = 2(x - 50) / 100 \quad (1)$$

次に聞いた楽曲の近傍の楽曲のスコアを計算する。まずはその楽曲からの距離が 1 以内の楽曲を選び、選ばれた曲

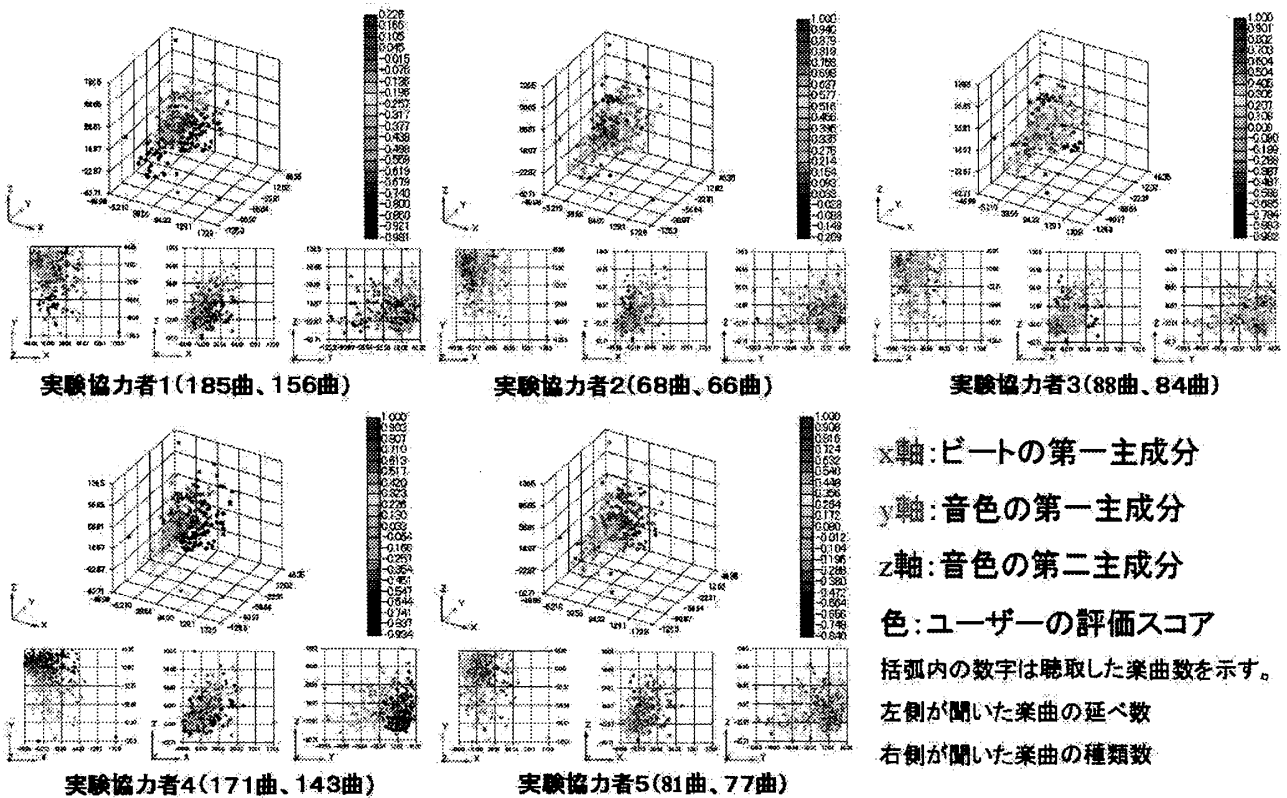


図4: 音楽的嗜好の可視化実験結果

が40曲未満であれば距離を1大きくして2以内の楽曲まで選ぶ。この操作を次は3、次は4というように繰り返し、選ばれる楽曲が40曲を超えた時点でやめる。この結果40~45曲の楽曲が選ばれる。なお40という数字にしっかりした根拠があるわけではなく、今後検討する必要がある。

そして選ばれた近傍の楽曲Aのスコア r_A は式(2)によって得られる。なお式(2)において d_A は聞いた楽曲と選ばれた楽曲Aの組織化された空間における距離を示す。また w_A は距離が大きくなるに従ってスコアの絶対値を小さくするための重みにあたる。

$$r_A = w_A \times r_0 \quad w_A = (100/(100 + d_A)) \quad (2)$$

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。このように1曲聞くごとに40曲ほどに対するスコアが得られる。

次に得られた r を用いて最終的なスコア R を求める。

4時間のうちに実験協力者は多数の楽曲を聞くが、4時間経った時点で楽曲Aのスコア r_A がN個得られていたとき、楽曲Aに対する最終的なスコア R_A は式(3)で表される。

$$R_A = \sum_N r_A / \sum_N w_A \quad (3)$$

これはN個得られた r_A を、重み w_A によって加重平均を取ることにあたる。このようにして得られた R_A が楽曲Aに対する最終的なユーザーの評価スコアとなる。

4.3 可視化実験結果

図4に実験協力者5人のそれぞれの楽曲に対する評価スコアの分布を示す。図4において上段は3次元の分布、下段左、中、右はそれぞれ3次元分布をx-y平面、x-z平面、y-z平面

面へと射影した図を示している。特徴空間内の点一つが一曲の楽曲を表しており、点の色が赤いほどユーザーのその楽曲に対する評価スコアが高く、青いほど楽曲に対する評価スコアが低いことを示している。

ユーザーはこのような結果を見て色の赤青から自分がどのあたりの楽曲が好きそうで、どのあたりの楽曲が好きではなさそうか知ることができる。

各実験協力者間のスコアを比較する。

実験協力者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などの現代の商業音楽である。実験協力者は20代前半の男性であり、やはり流行の現代の商業音楽に耳が慣れているため、このあたりの楽曲を好んで聞くというのは納得できる話である。

しかし実験協力者4は実験協力者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クラシック音楽やブルース音楽が多く含まれる領域である。実験協力者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、それが嗜好スコアに反映されていると言える。

また実験協力者5はxの値が高く、yの値が高い部分のスコアが高い。

さらに例えば実験協力者2と実験協力者4に注目すると、実験協力者4はy値の高低によってスコアの高低が分かれる傾向にあるが、実験協力者2はどちらかと言うとx値の高低によってスコアの高低が分かれる傾向がある。

4.4 推測精度評価

本研究における音楽的嗜好の推測精度を評価するために、先ほどの5人のうち評価実験に協力してもらうことが

表1: 推測精度評価実験の結果

楽曲	1	2	3	4	5	6	7	8	9	10	11	12
協1	3	2	2	4	6	3	4	6	2	2	3	3
協2	6	4	4	6	5	4	1	5	3	6	6	3
協4	3	5	4	3	5	2	2	3	3	3	4	2

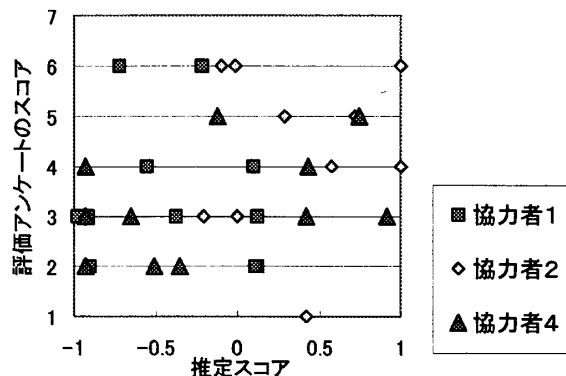


図5: 推定スコアと評価アンケートスコアの関係

できた3人の実験協力者に対してそれぞれ以下のことを行った。まず未聴取の楽曲のみをスコアの低い順に並べる。そしてテストデータとしてスコアの最も高い4曲とスコアの最も低い4曲、さらにその間を5等分するように4曲を選び、実験協力者にその12曲を聞いてもらい数値評価をさせる。

表1にその結果を示す。表1において曲番号が小さいほどシステムが推定したスコアが高い。また「協1」は実験協力者1を示し、番号は図4の実験協力者番号と対応する。評価の値は1~7の7段階であり、大きいほど楽曲に対する高い評価を表す。

また図5にその12曲におけるシステムが推定したスコアと評価アンケートスコアの間を関係を示す。図5を見ると推定スコアと評価アンケートスコアの間で相関が薄く、推定精度が良くないことがわかる。

5. おわりに

本研究では大量の楽曲群を音色、ビートという音響的特徴に基づいて3次元空間に組織化した。

また楽曲が組織化された空間において5人の実験協力者の音楽的嗜好の推測を行い、その結果を楽曲探索の手掛かりとして図4のように可視化した。

その結果実験協力者1~3, 実験協力者4, 実験協力者5の間で違った傾向の結果が得られ、赤青の色によってその違いが表現された。しかし4.4節で述べたように推測精度においては良い結果が得られていない。

このような結果になった理由として人は音色やビートだけでは音楽を聞いていないということが考えられる。音楽は他にもメロディやコード、歌詞といったさまざまな要素を持っている。図4のように各実験協力者間で違った傾向の結果が得られたことから、各実験協力者はそれぞれに音色やビートの好みを持っていて、その好みの抽出ができたと言える。しかし本研究ではそれら以外の特徴量は考慮しておらず、音色やビートは好みであってもメロディや

コード進行が好みでない曲のスコアも高くなってしまふ。このような楽曲がテストデータに入ったことによって推測精度が低くなったと考えられる。

そこで推測精度を上げるためにはメロディなど、音色とビート以外の特徴を抽出する必要がある。

またシステムが4つ以上の特徴量を保持するようになれば、各ユーザーに対する各特徴量の有効度を定量的に評価し、そのユーザーにとって有効度が高い特徴量3つを選んでは可視化する必要がある。

この定量的評価の実現のためには、図4の結果において各実験協力者ごとにどの軸によってスコアの高低が分かれているかに注目する必要がある。また定量的評価はスコアの多次元的配置も考慮されたものであることが望ましい。

またスコアの計算方法が原因で推定精度が下がっていないかどうか確かめておく必要がある。現在ユーザーの評価スコアの計算において、聞いた楽曲に対する評価値を得るために式(1)を、聞いていない楽曲に対する評価値を得るために式(2)を用いているが、これらの式の妥当性については確かめていない。そこで聞いた楽曲に対する評価値に関しては、「ユーザーが何%曲を聞いて飛ばしたか」と「ユーザーのその曲に対する評価」を結びつける関数を検討する必要がある。また聞いていない楽曲に対する評価値に関しては、適切な W_A の計算式を検討するのも手はあるが、計算方法をk-Nearest Neighbor法を用いたものに変え、クロスバリデーションにより適切なkを選ぶようにするという方向性も考えられる。

また推定精度の問題以外にも、可視化結果をユーザーが理解しやすいようにするという課題がある。本研究では特徴抽出の過程で主成分分析を行い、その主成分得点を楽曲の特徴量として採用したが、その結果3次元空間の軸がそれぞれ楽曲のどのような特徴を表しているのかがユーザーにとって理解しにくくなってしまった。

そこで例えばビートの特徴ならば「テンポ」や「裏拍の強さ」といったように、ユーザーが直観的に理解できる特徴を抽出して提示する必要がある。

謝辞: この研究は「RWC 研究用音楽データベース: 音楽ジャンル」[4]を用いました。

参考文献

- [1] K. Hoashi, K. Matsumoto, N. Inoue: Personalization of User Profiles for Content-based Music Retrieval Based on Relevance Feedback, Proceedings of ACM Multimedia 2003, pp. 110-119, November 2003.
- [2] E. Pampalk, A. Rauber, and D. Merkl: Content-based Organization and Visualization of Music Archives, In the Proceedings of the 10th ACM International Conference on Multimedia(MM02), pp 570-579, Juan-les-Pins, France, December, ACM, 2002.
- [3] T. Foote: Content-Based Retrieval of Music and Audio, Multimedia Storage and Archiving Systems II, Proceedings of SPIE3229, pp. 138-147, 1997.
- [4] 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.
- [5] 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