

旋律に潜むジャンルの特徴

鈴木 崇也† 長谷川 智史† 穴田一‡

東京都市大学大学院 工学研究科† 東京都市大学 知識工学部 経営システム工学科‡

1. 研究背景及び目的

現在,世界中で様々な楽曲が作曲,演奏され,人々の耳に届いている.また,楽曲にはそれぞれ様々な特徴があると言われ,特に楽曲のメインとなる旋律にその特徴が表れると考えられている.それに関連して,アーティストや作曲家,音楽ジャンル等の様々な特徴を楽曲の旋律から見出そうとする研究[1][2]が多数行われている.その中でも,音楽情報検索(MIR)において有効な指標として用いられている音楽ジャンルに関する研究が盛んに行なわれており,旋律から各ジャンルの特徴を抽出し,ジャンル判別等へ応用する試みがなされている.しかし,従来の研究[2]では Jazz と Classical の 2 種類のみなど,現存する音楽ジャンルに対して判別を行なっている種類数が少なく,ある程度限定したジャンルのみでしか分析を行っていないため,旋律に表れる各ジャンルの明確な特徴を見出すことは出来るとは言えない.そこで本研究では,旋律に潜むジャンルの特徴を捉えた新指標を提案し,その有効性の検証を行う.

2. 提案手法

2.1 分析に用いるデータ

本研究では,分析に用いるデータとして MIDI を用いる.MIDI の旋律パートのみを抽出し,そこから音高と音長に関するデータを抽出する.音高は MIDI のノートナンバーをそのまま用いる.音長は,MIDI に記載されている値を抽出した後,MIDI の 4 分音符の長さを表す TimeBase 値の 4 倍の値でそれらの値を割り,1 小節を 1 の長さに規格化する.また,それぞれの音が発音してから次の音が発音するまでの長さに各音符の音長を調整する.ここで,無音時間の長さが 1 以上のものは休符とする.また,和音を含む楽曲は除外する.

以上の操作を行い得られた音高,音長,休符のデータを「旋律データ」とし,分析を行っていく.

2.2 特徴量

楽曲の旋律パートの楽譜を見ると,音符数や長さは様々だが,音高の変動には,緩やかに上昇と下降を繰り返す山のような形状が多く見られる.またゲシュタルト心理学のよい連続の原理により,同一の方向性を持った緩やかな音高変化を 1 つのグループとして知覚すると考えられている.そこで,本研究ではその山の形にジャンルの特徴が表れると考え,次の図のように音高が上昇し,下降し,再び上昇する 1 つ前の音符までを「音高フレーズ」と定義した.



図 1. 音高フレーズ

音高が下降から上昇に転じる音符をフレーズの開始音とし,再び下降から上昇に転じる音符の手前の音符をフレーズの最後の音とする.A,B はそれぞれフレーズを表し,重複している音符 a は両方のフレーズに含めることとする.

各楽曲を音高フレーズに分割した後,「フレーズに含まれる音符数」,「フレーズの高さ」,「フレーズの幅」,「フレーズの形状」,「フレーズの非対称性」という 5 つの特徴量を計算する.フレーズの高さとは,各音高フレーズに含まれる音符の最高音高と最低音高の差,フレーズの幅とは各音高フレーズの音符の合計音長を表す.フレーズの形状は次の図のように,フレーズの最初の音符と,フレーズの山の頂点(以下:頂点)の音符,そして最後の音符を直線で繋いだものに対して,フレーズに含まれる音符がどの程度ずれているかで定義する.

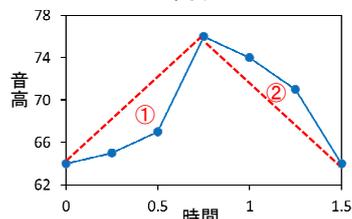


図 2. フレーズの形状の例

横軸は時間,縦軸は音高を表し,各点は音符を表す.フレーズの最初の音符を時間 0 とし,折れ線はそれぞれの音符を繋いだもので,丸い点は音符を表す.破線はフレーズの最初の音符と頂点の音符及び,フレーズの最後の音符と頂点を繋いだ直線を表す.この破線に対して,折れ線がどの程度ずれているかを山の形状として考え,破線とによって形成される図形①,②の面積を計算する.便宜的に破線より音高が下の部分で形成された①の図形の面積は負とし,上の部分で形成された②の図形の面積は正とする.そしてそれぞれの面積の和を求め,それをフレーズの最初の音符から最後の音符の 1 つ手前までの合計音長で割った値をフレーズの形状とする.

この図より,各フレーズの「フレーズの形状」が求まる.フレーズの非対称性は次の図のように,フレーズの最初の音符から頂点までの形状(実線)と頂点から最後の音符までの形状(破線)がどの程度異なっているかで定義する.

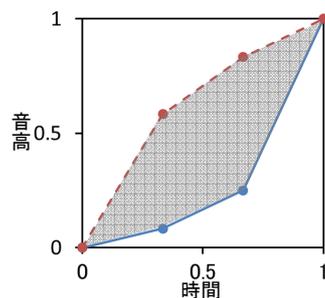


図 3. フレーズの非対称性の例

横軸は時間,縦軸は音高を表す.図 2 のフレーズの音高と時間を,最初の音符が 0,頂点が 1 となるようにそれぞれ調整する.同様に,頂点から最後の音符までの音高と時間を,頂点が音高 1,時間 1,最後の音符が音高 0,時間 2 となるように調整する.そして,時間 1 を基準とし折り返し,重ね合わせる.実線はフレーズの最初の音符から頂点まで,破線は頂点から最後の音符までの折れ線を表し,2 つの折れ線のずれである色づけされている部分の面積をフレーズの非対称性と定義する.

The Feature of Music Genres in the Melody

†Graduateschool of Engineering, Tokyo City University

‡Faculty of Knowledge Engineering, Tokyo City University

この図より、「フレーズの非対称性」が求まる。以上が音高フレーズの特徴量となる。

音高フレーズは音高の変動から形成される山であったが、その山の変動について着目してみると、音高の変動と同様に山のような形状が見られることがわかった。そこで、1つの音高フレーズを1つの音符と見立て、その変動から形成される山を1つのフレーズと考えることとする。音高フレーズを1つの音符と見立てるために、各音高フレーズの平均音高と幅を求める。平均音高 p_{ave} は以下の式で計算される。

$$p_{ave} = \frac{\sum_{k=1}^n p_k l_k}{\sum_{k=1}^n l_k} \quad \dots(1)$$

ここで、 n は音高フレーズに含まれる音符数、 p_k とは l_k それぞれ音高フレーズに含まれる k 番目の音符の音高と音長を表す。これにより、どの程度の音高がどの程度鳴っていたのかを考慮した平均音高が求まる。これを音高フレーズの音高と見立て、音長はその音高フレーズのフレーズ幅とする。フレーズへの分割方法は音高フレーズと同様にし、これを「平均音高フレーズ」とする。平均音高フレーズからも音高フレーズと同様の5つの特徴量を抽出する。

以上、音高フレーズ、平均音高フレーズから計算された10種類の特徴量ごとの平均、分散の計20種類の値を用いてジャンル判別実験を行い、有効性の検証を行なう。

3. 分析対象ジャンル及び楽曲

本研究で分析対象としたジャンルは、土橋らの研究[3]で用いられている midi データが配布されているサイトで分類されているジャンルから選定し、Rock, Pop, Blues, Country の4ジャンルとした。実験に用いる楽曲は、上記のサイトを含むフリーサイトから収集した各ジャンル46曲の計184曲である。

4. 実験

本研究で提案した特徴量の有効性を検証するため、ユークリッド距離を用いた k -近傍法によるジャンル判別実験を行なった。各ジャンルから10曲、計40曲をテストデータとし、残りの144曲を学習データとした。

ジャンル判別実験に用いる特徴量として、本研究では20種類の特徴量を用意した。ここで k -近傍法を行なう前に、各特徴量の値を $[0, 1]$ に規格化する。その方法として、各特徴量の184曲の平均と標準偏差(σ)を計算し、平均から 2σ 引いた値が0、足した値が1となるように調整する。また、はずれ値が存在すると判別精度の低下に繋がる可能性があるため、0を下回る値は0.1を上回る値は1とする。また 3σ についても同様な操作を行い、 2σ で規格化した特徴量、 3σ で規格化した特徴量それぞれについて実験を行う。

次に、各特徴量がどの程度判別に有効かわからないため、特徴量ごとに最適な重みを計算する必要があると考えられる。そこで本研究では、遺伝的アルゴリズムを用いて、各特徴量に対する最適な重みを計算することにした。

遺伝的アルゴリズムの設定として、各個体は特徴量数の遺伝子を持ち、それらの遺伝子は各特徴量の重みを表す。重みの値は0以上1未満とし、個体数は200個体とした。適応度は、その遺伝子の持つ重みをそれぞれの特徴量にかけた値で k -近傍法を行なった結果の正答率とした。ここで k -近傍法

は各ジャンルからランダムで8曲、計32曲を選定し、これをテストデータ、残りの112曲を学習データとし実験を行う。正答率は、この操作を4回行い得られた結果の平均正答率とする。テストデータは k -近傍法を4回行なう際、1度選択された楽曲は避け重複しないようにする。近傍数は McKay の研究[4]と同様サンプル数の平方根とし、11とした。

学習の流れとして、まず200個体の各遺伝子に対して初期値を $[0, 1]$ の一様乱数で与える、各個体の適応度を k -近傍法を行い計算する。次に交叉する個体を各適応度の値を用いたルーレット選択で2個体選出し、一点交叉を行なう。また、交叉した個体の各ビットに対して1%の確率で突然変異させる。新しく交叉して生まれた個体の適応度を計算し、最後に適応度が最も低い2個体を淘汰する。これが1ステップの流れとなる。収束条件は最大の適応度を持つ個体が10000ステップ変化しなかった場合とした。

以上の方法で $2\sigma, 3\sigma$ で正規化した特徴量それぞれに対して遺伝的アルゴリズムを行ない、最も高い適応度の個体の重みを用いて、テストデータに対して判別実験を行なった。

5. 結果

184曲から各ジャンル10曲ずつ、ランダムにテストデータを選出する方法で、 $2\sigma, 3\sigma$ で正規化した特徴量それぞれに対して k -近傍法で実験を行なった。近傍数は12とした。 2σ で正規化した特徴量を用いて実験を行なった結果を以下に示す。

表1. 各ジャンルにおける判別率

	Rock	Pop	Blues	Country
Rock	0.40	0.40	0.00	0.20
Pop	0.40	0.20	0.10	0.30
Blues	0.20	0.10	0.50	0.20
Country	0.50	0.00	0.20	0.30

各行名はテストデータのジャンル、各列名は k -近傍法により出力されたジャンルを表す。各数値は、その行名のジャンルのデータが列名のジャンルに判定された割合を表す。例えば Rock の行に着目すると、Rock のデータが Rock と判別された割合が0.40、Rock のデータが Pop と判定された割合が0.40となる。

詳しくは発表で述べるが、表1より Rock の楽曲に対しては、40%、Blues の楽曲に対しては50%とある程度高い判別を得ることができた。このことから、フレーズにジャンルの特徴が表れることが示唆された。詳しい考察及び今後の課題は発表にて述べる。

参考文献

- [1] 三家本祥平, 井手綾香, 出口幸子「楽譜データベースを用いた日本ポピュラー音楽の旋律分析」, 情報処理学会研究報告, 2006-MUS-45, pp.19-24 (2006)
- [2] Ponce de Leon, P.J. and Jose, M.I. 「Musical style identification using selforganising maps」, Proc. International Conference on Web Delivery of Music, pp.82-89 (2002)
- [3] 土橋佑亮, 北原鉄朗, 片寄晴弘「音響信号を対象としたベースラインからの音楽ジャンル解析」, 情報処理学会研究報告, 2008-SLP-70, No.12, pp.217-224 (2008)
- [4] McKay, C. 「Automatic genre classification of MIDI recordings」, McGill University (2004)

※概要の枚数制限上、その他の参考文献は記載しない