

音楽ジャンル分類精度向上のための楽曲内ジャンル境界の検出

中村 宏己[†] 黄 宏軒^{††} 川越 恭二^{††}

[†] 立命館大学大学院 情報理工学研究科 〒525-8577 滋賀県草津市野路東1丁目1-1

^{††} 立命館大学 情報理工学部 〒525-8577 滋賀県草津市野路東1丁目1-1

E-mail: [†]nakamura@coms.ics.ritsumeai.ac.jp, ^{††}huang@fc.ritsumeai.ac.jp, ^{†††}kawagoe@is.ritsumeai.ac.jp

あらまし 本稿では、音楽ジャンル分類精度向上のための楽曲内ジャンル境界の検出手法を提案する。音楽ジャンル分類の研究ではこれまで楽曲を1つのジャンルに分類する手法が数多く提案されてきた。しかし、多様化する音楽のジャンル分類をより正確に実現するには、複数のジャンルを有する楽曲にも目を向ける必要がある。複数のジャンルを有する楽曲とは、例えばロック音楽のソロ部にジャズの音階を用いたものや、ジャズ音楽の導入部にオーケストラを用いるものなどが挙げられる。このような楽曲に対して正確にジャンル分類を行うためには、ジャンルが変化するタイミングを抽出し、そのタイミングで楽曲を区間分割後、区間ごとに音楽ジャンルの分類を行う必要がある。そこで本研究ではMFCCとクロマベクトル、リズムグラムを用いた楽曲セグメント分割手法を開発し、楽曲内ジャンル境界の検出を行う。キーワード 音楽ジャンル境界、音楽ジャンル分類、実音響信号、MFCC、クロマベクトル、リズムグラム

Detecting Music Genre Boundaries for Accuracy Improvement of Music Genre Classification

Hiroki NAKAMURA[†], Hung-Hsuan HUANG^{††}, and Kyoji KAWAGOE^{††}

[†] Graduate School of Science and Engineering, Ritsumeikan University
Nojihigashi 1-1-1, Kusatsu, Shiga, 525-8577 Japan

^{††} College of School of Science and Engineering, Ritsumeikan University
Nojihigashi 1-1-1, Kusatsu, Shiga, 525-8577 Japan

E-mail: [†]nakamura@coms.ics.ritsumeai.ac.jp, ^{††}huang@fc.ritsumeai.ac.jp, ^{†††}kawagoe@is.ritsumeai.ac.jp

Abstract In this paper, we propose a novel method to detect music genre boundaries for the improvement of music genre classification(MGC) accuracy. In the existing studies of MGC, a music piece is mainly classified into one music genre. However, in order to realize more accurate music genre classification, we need to pay attention to a music piece having some music genres. In this paper, segment dividing method using MFCC, chroma vector, and rhythm vector is described in order to detect music genre boundaries.

Key words Music Genre Boundary, Music Genre Classification, Real Signal, MFCC, Chroma Vector, Rhythmogram

1. はじめに

近年、インターネット配信やポータブル機器に利用される音楽ファイルの増加に伴い、音楽情報検索の研究が盛んに行われるようになった。特に音楽ジャンル分類によるラベル付けは非常に重要な課題として多くの研究がされている[1]。従来の研究には、低音旋律パターンやドラムのリズムパターンを抽出し、クラスタリングや機械学習によりジャンル分類を行うものが多い[2][3]。このような音楽の音色や周期性に着目した研究は、音楽のジャンル分類に対して非常に有効である。

ジャンル分類に関する多くの研究は、単一ジャンルの楽曲を

分類対象としている。このため、同一楽曲内でジャンルが変化する楽曲に対して、従来手法はジャンル分類精度が悪くなる問題点がある。これは、前述した音響特徴の周期性を用いていることが原因に挙げられる。楽曲内でのジャンル変化時は、演奏楽器の変化により音色や音域が変わり、ビートの変化によりリズムパターンが変化する。そのため、ジャンルが変化する楽曲のジャンル分類には、音響特徴の変化を考慮した特徴抽出が必要である。また、音楽ジャンル分類研究で利用されるデータセット[4][5]はBlues, Classical, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae, Rockなどのジャンルを内包しているが、複数のジャンルを有する楽曲は扱っていない。

多様化する音楽のジャンル分類を実現するためには、単一のジャンルを有する楽曲だけではなく、複数のジャンルを有する楽曲にも目を向ける必要がある。複数のジャンルを有する楽曲とは、例えば、導入部にオーケストラを起用したメタル音楽やソロにジャズの和音とリズムを用いたロック音楽などが挙げられる。このような楽曲に対して高精度なジャンル分類を行うためには、ジャンルが変化する時刻で楽曲を分割し、分割区間ごとに音楽ジャンル分類を行う必要があると考える。

そこで本稿では、音楽ジャンル分類精度向上を目的とした楽曲の時間構造におけるジャンル境界検出手法を提案する。ジャンル境界検出にあたって、音色、和音、リズムの3つの音楽要素に着目した。なぜならば、これらの要素は音楽を形成する重要な音響特徴であり、音楽ジャンルの変化と共に大きく変動する可能性が高いと考える。本稿では、上記3つの特徴量に対してセグメント分割手法を適用し、楽曲内でのジャンル境界を検出する。

2. 音響特徴の変化量

曲中におけるジャンル変化時は、一般に音色と和音、リズムが変わりやすい特性を持つため、それらの変わり目を推定することによってジャンル境界の検出が可能であると考えられる。

本研究では、ジャンル境界検出に向けて、ジャンル変化に関わる音響特徴量 MFCC、クロマベクトル、リズムグラムに着目し、それらの変わり目を推定する。なお、各音響特徴の変わり目の推定には Foote らが開発した Novelty 検出法 [6] を用いる。

2.1 音響特徴量 [7] [8] [9]

• MFCC(メル周波数ケプストラム)

スペクトル包絡に対して、人間の周波数知覚特性を考慮した重み付けを行った特徴量である。人間の周波数知覚特性を考慮しているため、音色の数値表現に適している。本研究では、ケプストラムの低次成分 13 次元までを用いる。MFCC の例を図 1 に示す。

• クロマベクトル

フーリエ解析によって得られる周波数領域のエネルギーをオクターブの 12 音に折り返して足し合わせた特徴量である。クロマベクトルはオクターブの 12 音による表現であるため、倍音成分の影響が少なく和声の構成要素の解析が容易である。クロマベクトルの例を図 1 に示す。

• リズムグラム

短時間フーリエ変換の結果を元に算出したスペクトル束列に対して区間をずらしながら自己相関解析を行った特徴量である。スペクトル束列にたいして自己相関を行うことで、リズム周期をより正確に表すことが可能である。ずらす区間に 8 秒間、ずらす間隔に 0.5 秒を用いる。リズムグラムの例を図 1 に示す。

2.2 自己類似度行列 [10]

音響特徴量は、音楽ジャンルが一定である区間は反復し、音楽ジャンルが変化する区間は大きく変動すると考える。そのため、音響特徴の変化量の算出には自己類似性を用いることが有効である。自己類似性は、各フレームを全ての他のフレームとを比較することにより算出する。MFCC、クロマベクトル、リズムグラム

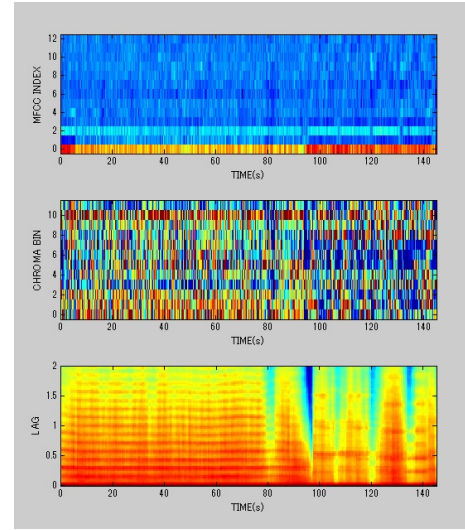


図 1 MFCC(図:上) クロマグラム (図:中) リズムグラム (図:下)

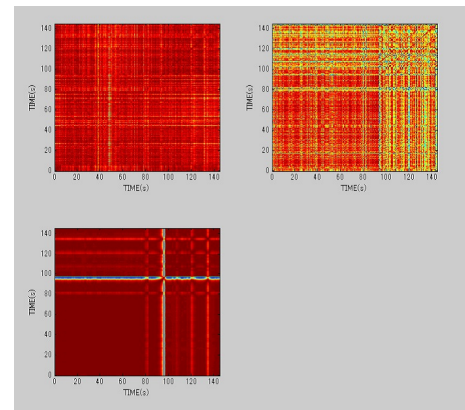


図 2 MFCC-SDM(図:左上) クロマグラム-SDM(図:右上) リズムグラム-SDM(図:下)

ムの自己類似度行列 SDM(Self-Distance Matrix) は音響特徴量 $(x_1, x_2, \dots, x_n)$ を用いて $SDM(i, j) = d(x_i, x_j)$ for $i, j \in \{1, 2, \dots, N\}$ で表す。また、類似度算出手法は様々考えられるが、本手法では一般的であるコサイン類似度を用いる。コサイン類似度 d は以下の式で算出する。

$$d(x_i, x_j) = \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|} \quad (1)$$

図 2 に各特徴量の SDM の例を示す。

2.3 自己類似度行列を用いた音響特徴の変化量算出

自己類似度行列 SDM に対して Novelty 検出法を適用し、各音響特徴の変化量 Novelty を求める。Novelty は $SDM(N \times N)$ サイズ) とガウシアンカーネル行列 $M \times M$ (with $M < N$) を SDM の対角成分に沿って畳み込むことで求める。Novelty のピークは SDM の対角成分上のブロックの角を示し、特に高値であるピークは変化が大きいことを示す。また、ガウシアンカーネル行列はガウス動径関数により重みを付加された正方行列であり、 2×2 のチェッカー図柄に似た構造を持つ。

図 3, 4 にガウシアンカーネル行列のイメージ図と Novelty 検出法により算出した各特徴量の変化量をそれぞれ示す。本稿で

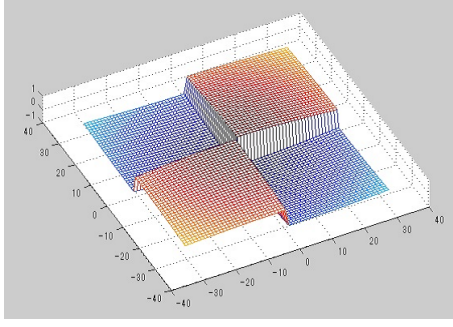


図 3 ガウシアンカーネル

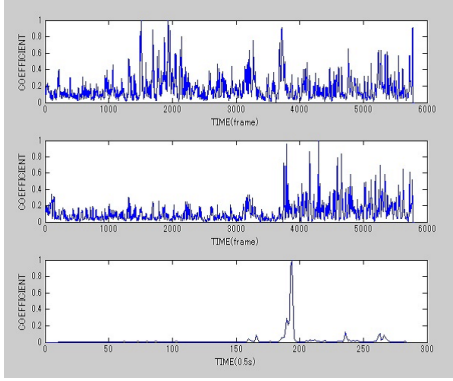


図 4 各特徴量の Novelty

は、ガウシアンカーネル行列のサイズは各特徴量の 1 秒間のフレーム数を用いる。

さらに、Novelty 検出法に加えて、ノイズを低減するために移動平均を用いた平滑化を行う。移動平均の区間は、ガウシアンカーネル行列のサイズ ϕ を用いる。移動平均 SMA(Simple Moving Average) は、Novelty の特徴量 $(n_1, n_2, \dots, n_m)$ 、区間 ϕ を用いて以下の式で算出する。

$$SMA_i = \frac{n_i + n_{i-1} + \dots + n_{i-\phi+1}}{\phi} \quad (2)$$

3. ジャンル境界検出

3.1 セグメント分割手法

本稿で提案するセグメント分割手法を説明する。

本セグメント分割法はまず、音響特徴の変化量である Novelty の時間的変化のデータに対して、与えられた閾値を用いてセグメントの 2 分割を行う。次にセグメント分割後、ジャンル境界時刻を求める。

3.1.1 セグメント分割条件

セグメント分割のための 3 つの条件を説明する。

音楽のジャンル変化は楽曲の構成に沿った演出であることから、使用頻度が増えると楽曲自体を破綻させる可能性が高い。そこで、楽曲内でのジャンル変化の最大回数 $maxgenre$ を設定し、ジャンル境界の可能性が低い時刻を検出対象外とする。これにより、ジャンル境界検出の精度向上を期待する。本セグメント分割法では、ジャンル変化最大回数よりセグメント最大個数 $maxseg$ を設定することで、ジャンル境界検出の精度向上を期待する。セグメント最大個数 $maxseg$ はジャンル変化最大回

Algorithm 1 Segments Dividing

INITIALIZATION: Segments S_0 ; Threshold ϕ , $maxseg$;

OUTPUT: candidate boundaries C

```

1:  $i \leftarrow 0$ ;
2:  $j \leftarrow 0$ ;
3: while (1) do
4:   if  $i > j$  then
5:     break;
6:   else if  $j \geq maxseg \ \&\& \ i \geq maxseg$  then
7:     break;
8:   end if
9:
10:  segment values  $s \in S_i$ ;
11:  if  $check(s) = true$  then
12:     $t = max(s.arrayindex)$ ;
13:     $C.add(t)$ 
14:
15:    if  $check(s[1...t - \phi]) = true$  then
16:       $S_{j+1} = s[1...t - \phi]$ ;
17:       $j++$ 
18:    end if
19:
20:    if  $check(s[t + \phi + 1 : end]) = true$  then
21:       $S_{j+1} = s[t + \phi + 1 : end]$ ;
22:       $j++$ 
23:    end if
24:
25:  end if
26:   $i++$ ;
27: end while
28: return  $C$ ;

```

図 5 Algorithm1

数 $maxgenre$ を用いて $maxseg = maxgenre + 1$ で算出する。

また、セグメント内で変化量の大きい時刻を多数検出した場合、そのセグメントはジャンル境界を含まないと考えることが妥当である。これは、ジャンル境界の変化量は他の時刻に比べて明らかに大きいため、セグメント内の変化量に大きな差がない場合は、ジャンル境界がないと推測できるからである。このため、閾値 τ を導入する。変化量が閾値 τ 以上である時刻の数が、セグメントの最大個数 $maxseg$ を超える場合は、そのセグメントにジャンル境界はないと判断する。

さらに、セグメントの長さが極端に短い場合についても、そのセグメントにジャンル境界はないと考えることが妥当である。一般に音楽は、繰り返し構造を持つことが知られており、繰り返しの基本単位として主に 4 小節が用いられる。これはジャンルの変化についても同様で、ジャンルが変化する区間は最小でも 4 小節分の長さが必要である。そこで、楽曲のテンポ $tempo$ を用いてセグメントの最小サイズ $segsz$ を以下の式で算出する。なお $tempo$ の算出には、Lartillot [11] らを参考にした。

$$segsz = (60/tempo) * 16 \quad (3)$$

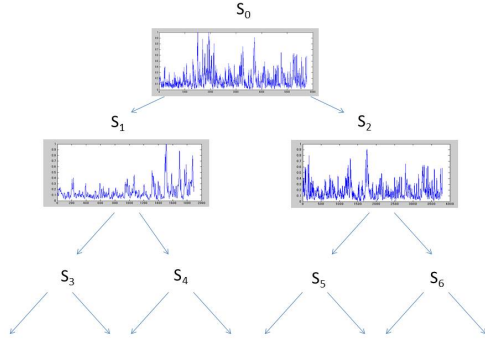


図 6 セグメント分割手法のイメージ図

Algorithm 2 check(s)

INITIALIZATION: Segment Values s ; Threshold τ , $segsz$, $maxseg$;

OUTPUT: boolean;

```

1:  $s \leftarrow s / \max(s)$ 
2: if  $length(s) > segsz \ \&\& \ length(s < \tau) \leq maxseg$  then
3:   return true
4: else
5:   return false
6: end if

```

図 7 Algorithm2

3.1.2 セグメント分割手法の処理内容

図 5 に示す Algorithm1 は、音響特徴の変化量 *Novelty* を入力として、ジャンル境界候補の時刻を出力するアルゴリズムである。Algorithm1 では音響特徴の変化量 *Novelty* の時間的変化データをセグメントとして扱う。処理の流れを以下で説明する。

まず、全ての *Novelty* 時間的変化データである初期セグメント S_0 がセグメント分割条件を満たしているかを check() を用いて判定する (11 行目)。もし S_0 がセグメント分割条件を満たす場合、 S_0 内の最大値を取る時刻 t をジャンル境界候補とする (12 行目)。ここで条件を満たさない場合は、 S_0 にジャンル境界候補は存在しない。次に、 S_0 を時刻 t でセグメント S_1, S_2 に分割する。ここで、セグメント S_1, S_2 が、セグメント分割条件を満たしているかを判定し、条件を満たしていれば、そのセグメントを生成する (15-23 行目)。ただし、 S_1, S_2 は $t \pm \phi$ の時刻範囲を含まない。この処理は閾値 ϕ を用いてノイズを取り除くために行う。以下、 S_0 を S_1 に置き換えて処理を繰り返す。ループの終了条件は全てのセグメントが分割条件を満たさない、あるいはセグメントの個数が最大個数 $maxseg$ を超える場合である (4-7 行目)。

図 6 にセグメント分割手法のイメージ図を示す。

次に図 7 に、図 5 内で用いた check() のアルゴリズムを説明する。check() は、セグメント s がジャンル境界を含むか否かを判定する関数である。まず、セグメントのベクトル s を最大値 $\max(s)$ で正規化する (1 行目)。次に、セグメント分割条件を満たしているかを判定する。check() が判定するセグメント分割

条件は以下の 2 つである。

- セグメント s の長さが $segsz$ を超えている
- セグメント内の閾値 τ 以上の時刻の数が $maxseg$ 個以内

3.2 ジャンル境界時刻の検出

楽曲内でジャンルが変化する時、全ての音響特徴が同時に変化するのではなく、ある程度のタイムラグが発生する場合がある。そのため本研究ではジャンル境界の検出にあたって、音響特徴の変化のタイムラグを考慮した検出を行う。すなわち、時間区間 lag を各特徴量の変化のタイムラグとすると、時間区間 lag 内に各特徴量のジャンル境界候補が全て存在する時、その区間内の先頭にあるジャンル境界候補をジャンル境界時刻として検出する。

本研究では lag に楽曲の 1 小節分の長さとして定義し、楽曲のテンポ $tempo$ を用いて $(60/tempo) * 4$ で表す。また、MFCC、クロマベクトル、リズムグラフのジャンル境界候補時刻の集合をそれぞれ mft , cht , rht とする。

3.3 ジャンル境界検出の具体例

ジャンル境界を含む楽曲を対象にセグメント分割手法を適用し、ジャンル境界検出を行った。テストデータは、演奏者 EugenCieroJerry による楽曲 JerryVanRooyen を用いる。事前に音楽有識者 3 人にアンケートを取り、それより求めたジャンル境界の正解時刻は 98 秒付近である。この楽曲に対して閾値 $\tau = 0.7$, $maxseg = 8$ を用いてセグメント分割手法を適用した。その結果を以下に示す。なお、各要素の時間単位は秒である。

$$mft = \{54.8, 43.5, \underline{99.8}, 39.0, 52.7, 59.5, 137.9\}$$

$$cht = \{121.7, 101.6, 124.5, \underline{100.6}, 114.3, 126.5\}$$

$$rht = \{97.0, 95.0, 118.5, 83.0, 96.5, \underline{98.5}, 132.5\}$$

次に、この mft , cht , rht に対して、区間 lag 内に同時に存在するジャンル境界候補を求める。この楽曲の $tempo$ の値は 80 であるため、3.2 節で定義した式を用いて、区間 lag は $(60/80) * 4 = 3.0$ 秒と算出できる。従ってこの区間 lag 内に同時に存在するジャンル境界候補は $mft = 99.8$, $cht = 100.6$, $rht = 98.5$ である。最後に、これらの区間内の先頭の特徴である $rht = 98.5$ をジャンル境界として検出する。ジャンル境界検出の例を図 8 に示す。

この楽曲における正解と提案手法との誤差は最大でも 0.5 秒である。これは、この楽曲の 4 分音符の長さ以下の誤差であるため、提案手法によりジャンル境界の検出が可能となった。

4. 評価実験

4.1 実験条件

提案した楽曲内ジャンル境界の検出手法の有効性を確認するため評価実験を行った。

まず、提案手法の有効性を検証するため、楽曲内ジャンル検出精度の算出を行う。ジャンル境界検出は筆者が調査した限り研究報告がこれまでにない。このため、比較手法として楽曲構造解析手法 [12] を用いる。また、提案手法にはセグメント分割手法において閾値の設定が必要である。閾値設定によりジャンル境界検出精度がどのように変化するかを検証することができる。このため、異なる 2 種類の閾値を設定した。表 1

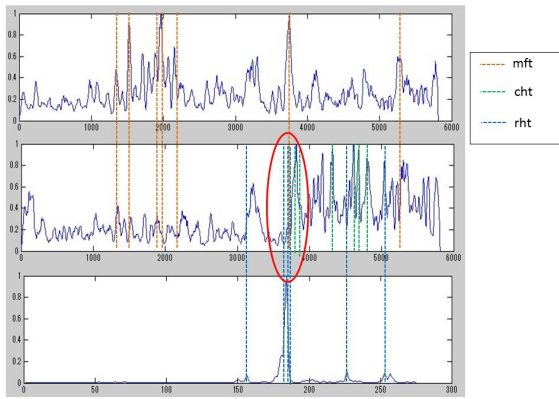


図 8 ジャンル境界検出

に示すような閾値を設定した。2つの提案手法を提案手法 a 、提案手法 b と呼ぶこととする。提案手法 a 、提案手法 b の各閾値を表 1 に示す。

	$\tau(\text{Algorithm2})$	$maxseg$
提案手法 a	0.5	12
提案手法 b	0.7	8

表 1 提案手法 a, b の閾値

実験では、検出されたジャンル境界が正解時刻の前後 α 秒以内にある場合を正解とする。本実験では、閾値 α を各楽曲の 1 小節分の長さとして定義し、以下の式で算出する。

$$\alpha = (60 / tempo) * 4 \quad (4)$$

各実験における評価尺度として適合率、適合率、F 値を用いる。提案手法 a, b と比較手法により検出した各楽曲のジャンル境界の数を F_E 、正解集合の数を F_A とすると、適合率 R_P は $|F_A \cap F_E| / F_E$ 、再現率 R_R は $|F_A \cap F_E| / F_A$ で定義し、F 値は以下の式で算出する。

$$F = \frac{2R_P R_R}{R_P + R_R} \quad (5)$$

N 個の楽曲集合における評価尺度は、それぞれ、平均ジャンル境界検出数 A_E 、平均適合率 A_P 、平均再現率 A_R 、平均 F 値 A_F と定義し、各楽曲の評価尺度の値の合計を楽曲数 N で割ることで求める。本実験において最も有力な評価尺度は A_F とする。この理由は、 A_F が高い手法は最小限のジャンル境界時刻で最大限の正解を検出したことを示唆し、ジャンル境界時刻検出の性能が最も高いからである。

実験には以下に示す 2 種類のテストデータを用いる。

1 つ目は、様々な音楽ジャンル分類システムの評価に利用されてきた GTZAN データセットを元に作成した疑似データである。GTZAN データセット [4] は、16 ビットレート 22050Hz でサンプリングされた 30 秒の楽曲を 1000 曲含み、各楽曲は 10 個の異なるジャンルに属する (blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, rock)。本実験では、このデータセットの中から異なるジャンルに属する 2 つの楽曲を可能な

限りなめらかに連結したものを疑似データとして用いる。疑似データの曲数は 225 曲で、楽曲の長さは全て 60 秒である。また、変化するジャンルの順番は考慮せず全てのジャンルが組み合わせるように作成した。

2 つ目は、生データとしてジャンル変化を含む市販の楽曲 (CD) を用いた。これらは事前に音楽の有識者 3 人によりジャンル変化を含む楽曲を選定してもらい、それぞれにジャンル変化の時刻を設定した。使用した楽曲数は 30 曲、楽曲の長さは最小で 3 分、最大で 7 分 40 秒、ジャンル変化回数は最小で 1 回、最大で 6 回である。このテストデータは、提案手法の市販の楽曲に対する有効性を検証するために利用する。

4.2 実験結果・考察

4.2.1 ジャンル境界検出

	A_F	A_P	A_R	A_E
比較手法	0.31	0.20	0.71	3.24
提案手法 a	0.29	0.17	0.89	2.72
提案手法 b	0.48	0.42	0.56	1.43

表 2 ジャンル境界検出の評価結果 (疑似データ)

	A_F	A_P	A_R	A_E
比較手法	0.32	0.23	0.52	3.67
提案手法 a	0.30	0.20	0.57	3.25
提案手法 b	0.37	0.29	0.52	3.10

表 3 ジャンル境界検出の評価結果 (CD 音源)

	A_F	A_P	A_R	A_E
Blues + others	0.26	0.15	0.92	2.75
Classical + others	0.29	0.17	0.91	2.52
Country+others	0.30	0.18	0.85	2.40
Disco+others	0.23	0.13	0.86	2.96
Hiphop + others	0.34	0.21	0.88	2.24
Jazz + others	0.42	0.28	0.86	2.04
Metal+ others	0.25	0.15	0.75	2.76
Pop + others	0.24	0.14	0.85	2.81
Reggae + others	0.13	0.07	0.86	3.93
Rock + others	0.27	0.16	0.87	2.83

表 4 各ジャンルの組み合わせと評価結果 (提案手法 a : 疑似データ)

表 2,3 に示すように、提案手法 b の平均 F 値 A_F が高い。この理由として、表 2,3 の A_E に示すように検出したジャンル境界の数に大きな差が見られたことが挙げられる。比較手法は楽曲構造のための境界検出手法であるため、各楽曲に対し多数の境界を検出した。そのため、平均適合率 A_P 、平均 F 値 A_F が大幅に低下した。これに対して、提案手法 b は閾値 τ 、 $maxseg$ を用いて変化量が大きく、かつ最低限のジャンル境界候補を検出したため、ジャンル境界の誤検出が少なくなった。このことから、提案手法 b の平均 F 値 A_F が最も高くなったと考えられる。

また、他の評価尺度 A_R について考察する。

表 2,3 に示すように提案手法 a は平均再現率 A_R が最も高い。この理由は、境界検出手法の相違である。比較手法は各特徴量

の変化量を正規化し、変化量の合計値から境界を検出した。そのため、変化量の小さい特徴量が埋もれてしまうことがよく見られた。例えば、MFCC の時間に伴う変化量が、[4,1,1]、クロマベクトルの変化量が [6,3,20] であるとき、比較手法はこれらの合計値 [10,4,21] から 3 番目の要素をジャンル境界とする。これに対し、提案手法 a では各特徴量からジャンル境界候補を選出し、それらが同時に変化する時刻を検出した。前述の例では、1 番目の要素をジャンル境界として検出する。そのため、提案手法 a がより正確にジャンル境界を検出できたと考えられる。

また、表 3 に示すように市販の楽曲では全ての手法で検出精度が全体的に低下した。市販の楽曲は生データであるため疑似データに比べてノイズが多い。それに伴い音響特徴の誤抽出が増加したことが原因に挙げられる。ノイズの原因として、エレクトーンやシンセサイザーなどの周波数帯が広域になる電子楽器が挙げられる。比較手法は、各特徴量の変化量を合計してジャンル境界の検出を行っている。これに対して、2 つの提案手法は、各特徴量の変化量をそれぞれ用いるので、1 つでも上手く抽出出来ない特徴があると、誤検出の可能性が高まる。また、評価結果より、提案手法 a は平均再現率 A_R 、提案手法 b は平均適合率 A_R と平均 F 値 A_F がそれぞれ比較手法を上回った。したがって疑似データと同様の理由から、市販の楽曲に対しても提案手法は有効であるといえる。

最後に、表 4 に提案手法 a による疑似データで連結した各ジャンル別の評価を示す。表 4 に示すように、Classical+others と Blues+others の平均再現率 A_R が高い。これは、Classical と Blues は比較的に静かな楽曲であり、それらが他ジャンルへ変化する際に各特徴量が大きく変化したため、正確にジャンル境界を検出できたと考えられる。一方、表 4 に示すように Metal+others の平均再現率 A_R が低い結果となった。この理由として、Metal の楽曲は音量の変化やノイズが多く、それらを誤認識したことが挙げられる。このような楽曲に対しては、他の特徴量の導入を検討する必要がある。

評価実験の結果、提案手法 b が比較手法を平均 F 値で上回ったことから、セグメント分割法を用いたジャンル境界検出は有効であるといえる。今後はより最適な閾値設定と他の特徴量の導入の検討が必要である。

5. おわりに

本研究では、音楽ジャンル分類精度向上のための楽曲内ジャンル境界検出手法を提案した。提案手法により、ジャンル境界検出の有効性を示し、また、従来では考慮されていなかった複数のジャンルを持つ楽曲に対して、音楽ジャンル分類が適用可能となった。また、提案手法を音楽の有識者が選定した市販の楽曲に適用することで、実際の楽曲で出現するジャンル変化の検出が可能であることを示した。

今後の課題として、提案手法で用いた閾値を自動的に設定することにより、ジャンル境界をより柔軟に検出する方法が挙げられる。そのため、機械学習の導入を予定している。また、ジャンルに関連する他の特徴量を追加することで、ジャンル境界をより正確に検出するよう改良する。また、それにより複数ジャ

ンルからなる音楽の音楽ジャンル自動分類システムの実現と評価を行う予定である。

文 献

- [1] George Tzanetakis, P. C. Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, vol 10, No. 5, JULY 2002
- [2] 角尾 衣未留, George Tzanetakis, 小野 順貴, 嵯峨山 茂樹. 音楽音響信号の低音旋律パターンのクラスタリングと自動ジャンル認識への応用. 情報処理学会研究報告. [音楽情報科学] 2009-MUS-81(1), 1-6, 2009-07-22
- [3] T.Lidy, A. Rauber. Audio music classification using a combination of spectral, timbral, rhythmic, temporal and symbolic features. In Proc. ISMIR, Sep, 2008
- [4] GTZAN: http://marsyas.info/download/data_sets
- [5] RWCDatabase: <http://staff.aist.go.jp/m.goto/RWC-MDB/rwc-mdb-g-j.html>
- [6] J. Foote. Automatic audio segmentation using a measure of audio novelty. In Proc. of IEEE International Conference on Multimedia and Expo, pp 452-455, New York, N.Y., USA, Aug. 2000.
- [7] Daniel P. W. Ellis. Classifying music audio with Timbral and Chroma features. Proc. ISMIR-07, pp. 339-340, Vienna, Austria, October 2007.
- [8] Beth Logan. Mel Frequency Cepstral Coefficients for Music Modeling. In Proc. ISMIR, USA, October, 2000.
- [9] K. Jensen. Multiple scale music segmentation using rhythm, timbre, and harmony. EURASIP Journal on Advances in Signal Processing. Article ID 73205, 11 pages, 2007
- [10] J. Foote, Visualizing music and audio using self-similarity, in Proceedings of the 7th ACM International Multimedia Conference & Exhibition, pp. 77-80, Orlando, Fla, USA, November 1999
- [11] O. Lartillot, Mirtempo: tempo estimation through advanced frame-by-frame peaks tracking, MIREX, 2010
- [12] Paulus, J., Klapuri, A., Music Structure Analysis Using a Probabilistic Fitness Measure And an Integrated Musico-logical Model, in Proc. ISMIR, pp. 369-374, Philadelphia, Pennsylvania, USA, September 14-18, 2008,