

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

土 橋 佑 亮[†] 北 原 鉄 朗^{†,††} 片 寄 晴 弘^{†,††}

音楽ジャンルは楽曲検索において有力な指標となる。音響信号を対象とした音楽情報検索の研究では、音色やリズムといった低次の特徴量を混合音全体から抽出するのが主流であるが、ユーザーの嗜好に合わせた検索を目指すには個々の楽器パートからの特徴量抽出が必要である。本稿では楽器編成の中で重要な役割を果たし、かつ複音からの分離が比較的容易なベースパートの特徴量を取り入れたジャンル推定を扱う。まず、ベースラインの特徴量を PreFEst を用いて得られた基本周波数から抽出し、従来より用いられてきた音色やリズム系の特徴量も用意する。マハラノビス距離を用いた 6 ジャンルでの識別実験において、ベースラインの特徴量を取り入れることで全ジャンル総合の認識率が 54.3 % から 62.7 % に向上した。また Music Island を利用して注目する特徴量に応じての楽曲の島を作成し、ユーザーの嗜好に対しての柔軟な楽曲分類を実現した。

Music Genre Classification of Audio Signals from Bass Part

YUSUKE TSUCHIHASHI,[†] TETSURO KITAHARA^{†,††}
and HARUHIRO KATAYOSE^{†,††}

Music genres play an important role in music information retrieval(MIR). Most of the previous studies on MIR for audio signals have used low-level features, such as timbre and rhythm from a mixture of sound, but acoustic features could be extracted from individual instrument parts to achieve user-adaptive MIR. In this paper, we deal with music genre classification using acoustic features extracted from the bass part, which plays an important role and the fundamental frequency of which can be comparatively easily estimated. First, the paper describes feature extraction about the bass part from pitch information obtained with PreFEst. We also prepare features about timbre and rhythm, which have been used so far. Experimental results of 6-genre classification by using the Mahalanobis distance show success rates of 62.7 % (with bass-part features), against 54.7 % (without bass-part features). Finally, we built Music Islands by browsing different views, and achieved flexible music classification for user's preference.

1. はじめに

人々の音楽の楽しみ方はここ数年で大きく変化し、ネットから配信された楽曲をハードディスクに大量に保存して聴くようになった。それに伴い楽曲を検索する際の新たな検索技術として MIR(Music Information Retrieval) の研究分野が注目されている。現在 MIR の研究技術は楽曲の「内容に基づく手法」と膨大な数のユーザーによる「評価を用いた手法」に大別され、後者においてはアマゾン等の Web サイト上で既に実用化に至っている。しかし日々新たに作成される楽曲や、あまり多くの人が聞かない少数派の楽曲など、ユー

ザーの評価を十分に獲得できていないものに対しては機能しない問題点がある。一方、前者は楽曲データそのものから抽出した情報を用いて検索を行うので、こういった問題が発生しない。

楽曲データから抽出すべき内容として、Rock や Jazz 等の音楽ジャンルが考えられる。音楽ジャンルは楽曲検索における 1 つの指標となるものであり、近年では国際会議 ISMIR¹⁾ において音楽ジャンル解析が主要テーマの 1 つとして扱われている。楽曲データから音楽ジャンルの情報を抽出する研究は、MIDI データを対象にしたものと音響信号を対象としたものに分けられる。MIDI を対象とした特徴抽出においては、Basili ら²⁾ や McKay ら³⁾ によって楽曲のパート編成やピッチ、音長といったような高次の特徴量が提案されている。一方で音響信号の分野においては、Dixon ら⁴⁾ による楽曲のリズムパターンや、Tzanetakis ら⁵⁾ や Pampalk ら⁶⁾ による音色やリズム特性といった低

[†] 関西学院大学大学院理工学研究科
Graduate School of Science and Technology, Kwansei
Gakuin University

^{††} 科学技術振興機構 CREST CrestMuse プロジェクト
CrestMuse Project, CREST, JST

次の特徴量が主流となっている。世の中に存在する楽曲の大多数を占める音響信号を対象とした技術の向上が期待される一方、Pampalk は博士論文の中で「低次の特徴量だけを用了楽曲検索手法には限界がある」と指摘しており、解決策としてより高次の特徴量の設計が望まれている。

我々はこの音響信号の楽曲データを対象とした「内容に基づく手法」の分野において、かねてから音楽ジャンル解析におけるベースパートの重要性に注目してきた⁷⁾。ベースパートは後藤の PreFEst⁸⁾ において混成音楽からの音源分離が報告されているように、単独パートの特徴量として抽出が可能である。またリズムとハーモニーの両面を支える重要なパートであり、後述のようにジャンルごとの演奏に特徴が見受けられる。文献⁷⁾では MIDI データに対してベースパートの有効性は確認したものの、音響信号に対して確認するには至っていなかった。そこで本研究では音響信号を対象とし、ジャンル識別において従来から用いられてきた音色やリズム系の特徴量に加え、高次の特徴量としてベースラインの特徴量を取り入れ、その有効性を示すことを目的とする。

まず第 2 章ではベースラインの特性と特徴量について述べる。次に第 3 章では後続の章での実験に用いる各種特徴量の設定を行い、第 4 章ではマハラノビス距離を用いた識別実験を行う。最後に第 5 章では Music Island⁶⁾を用いた特徴量に応じての楽曲マップの作成を行い、音楽ジャンルの定義のあいまいさに対処する。

2. ベースラインの特性と特徴量設定

本章では楽器構成の中でのベースラインの特性とジャンル別のベースライン特性について述べ、音楽ジャンル識別に有効な特徴量候補の考察と設定を行う。

2.1 ベースラインの特性

全ジャンルの楽曲を通してベースパートは

- 低域から和声構造やリズムを支え、楽曲全体の印象を形成する土台となる
- ほぼ全ての楽曲に存在し、楽曲中の大部分で演奏が継続、無音区間の割合が極めて低い

等の性質があり、楽器構成の中でも極めて重要な役割を果たすと共に、ジャンルによつての演奏パターンにも特徴がある。『ベース演奏技法』⁹⁾によると、「ロックは音使いを複雑にしないで、8 分音符を連打するのが定番」「ダンス系の楽曲ではオクターブの跳躍をともなった繰り返しパターンが典型的」「ジャズは経過音を頻繁に用いたウォーキングベースラインが主流」等の記述がある。本研究では「音高のバリエーション



Fv1: 楽曲全体で 1 フレーム以上使用した異なる音高の数: 6 (A2, C#3, D3, E3, D#3, B2)
Fv3: 短時間窓内での異なる音高の数の時間方向の平均値: 3.5 (1 小節目 A2, C#3, D3)
(単位時間を上記の 1 小節目とする場合) (2 小節目 E3, D#3, C#3, B2)
Fv5: 出現回数の多い音高 1 位と 2 位の音高差: 1 (1 位と 2 位は D3 と C#3)

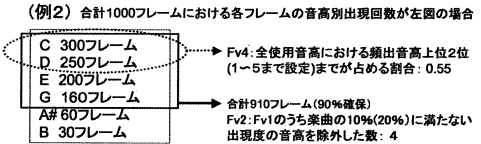


図 1 音高のバリエーションに関する特徴量

に関わる特徴量」と「音高の推移に関わる特徴量」を設定し、後に相関係数やジャンル分離尺度を用いて圧縮することを前提として網羅的な抽出を行う。

2.2 音高のバリエーションに関わる特徴量

音高のバリエーションに関する特徴量 (11 個) を以下のように設定する。なお () 内は特徴量の要素数を表している。また各特徴量の抽出例を図 1 に示す。

- Fv1: 楽曲全体で 1 フレーム以上使用した異なる音高の数 (1)
- Fv2: Fv1 のうち楽曲の 10 % (20 %) に満たない出現度の音高を除外した数 (2)
- Fv3: 短時間窓 (2 秒に設定) 内での異なる音高の数の時間方向の平均値 (1)
- Fv4: 全使用音高における頻出音高上位 i 位までが占める割合 ($i = 1, \dots, 5$) (5)
- Fv5: 出現回数の多い音高 1 位と 2 位の音高差 (1)
- Fv6: 最も出現回数の多い音高の平均周期 (1)

Fv1 は使用する音高の数が限定されがちな Rock 系 (Metal や Punk 含む) の楽曲と、経過音を頻繁に用いる Jazz の特徴抽出を意図したものであり、Fv2 ではさらに雑音の影響を無視できる可能性を考慮した。Fv3 は Electronic 系の楽曲に見られる、短いフレーズでよく動くパターンの抽出を、また Fv4 は頻出する特定の音高のみで楽曲の大部分をカバーするジャンルの検出を意図したものである。Fv5 と Fv6 は Tzanetakis⁵⁾ がピッチ特徴量の抽出の際に設定した特徴量と同じものを用意した。

2.3 音高推移に関わる特徴量

音高の推移に関する特徴量 (10 個) を以下のように設定する。なお () 内は特徴量の要素数を表している。また各特徴量の抽出例を図 2 に示す。

- Ft1: 単位時間あたりの推移回数の平均値 (1)
- Ft2: 隣接音との各推移パターンの割合 (5)
- Ft3: 連続 2 回の推移における各パターン割合 (4)



図2 音高推移に関する特徴量

隣接音との推移パターンは半音進行、順次進行 (3度未満)、跳躍進行 (3度以上)、6度以上の跳躍進行、オクターブ進行の5種類を意味する。また連続2回の推移におけるパターンは1回目と2回目それぞれ順次進行か跳躍進行かの組み合わせ4種類を意味する。Ft1はベースラインの音高推移が激しいと予想されるJazz等のジャンルとルートキーピングが基本となるRock系のジャンルの分離を意図している。Ft2とFt3はジャズ等のウォーキングベースやエレクトリック系の音が跳ねるフレーズなど、各ジャンルで特徴のある推移パターンを検出する特徴量として網羅的に設けた。

またこれらの特徴量の抽出対象となる基本周波数の解析においては一般的にオクターブ違いの推定が起こることが多い。この場合を想定し、「音高」とは別にオクターブを無視した「音名」に対しても同様の特徴量を設定し (ただし Ft2 のオクターブ推移に関しては抽出不可能)、合計41個をベースラインの特徴量とする。

3. 処理の流れ

本研究の全体像を図3に示す。前章で紹介したベースラインの特徴量の抽出には、後藤の PreFEst⁸⁾ を利用して得られる低域に存在する基本周波数 (F0) の候補を用いる。さらに従来からジャンル識別に利用されてきた「音色の特徴量」と「リズムの特徴量」の抽出も行い、次元圧縮の過程を経て以降の章での実験に用いる特徴量とする。

3.1 ベースラインの基本周波数推定

入力音響信号を対象に10msの窓幅で短時間フーリエ変換を行い、ピークの存在する周波数とパワーを抽出する。次にベースパートのみを対象とするためローパスフィルタをかける。文献8)と同じく周波数域を260Hz以下 (A0~B3) とする。その後 PreFEst⁸⁾ を用いて、どの周波数にどの程度の強さの基本周波数成分があるかを時刻ごとに推定し、程度の大きいものから順に第1候補、第2候補とする。

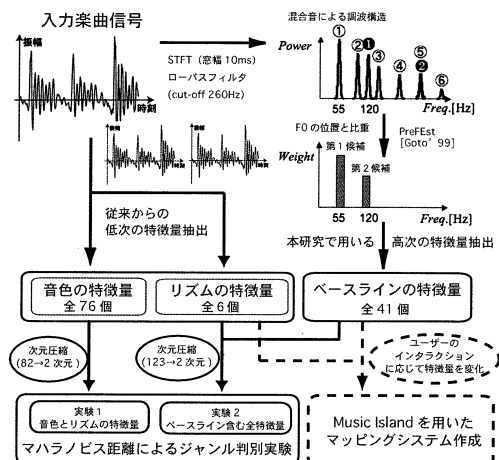


図3 実験の全体像

PreFEstを用いて得られた低域に存在するF0は必ずしもベースパートのF0と一致するとは限らない。我々は雑音除去として前後間の連続値として見ることで、よりベースラインとして可能性の高い基本周波数へと補正を行う。具体的には以下の2つの処理を行う。

- 第2候補のF0が前フレームで第1候補となっている場合の第2候補採用
- 前後5フレームの合計11フレームにおける最頻値による中央値の修正

3.2 ベースラインの特徴量

3.2により得られたベースラインのF0を対象に、2章で述べた特徴量 (全41個) を抽出し、「ベースラインの特徴量」とする。

3.3 音色の特徴量

音色系の特徴量としてはSTFT (Short-Time Fourier Transform) に関するもの (全66個) とMFCC低次5次元までの平均と分散 (全10個) の合計76個を用いる。STFTに関する特徴量は西山らの特徴量¹⁰⁾を用いた。これはTzanetakisら⁵⁾がジャンル識別で、またLuら¹¹⁾が音楽のムード解析で用いたものを基に設定したものである。STFTに関する特徴量は4096サンプルの窓幅ごとに図4に示す全33種類の特徴量を抽出し、それぞれ対象とする楽曲の全演奏時間における平均と分散をとったものである。またサブバンドとは周波数帯域で分かれた帯のことであり、1(低周波数域)~7(高周波数域)を意味する。

3.4 リズムの特徴量

リズムの特徴量として楽曲のビートの周期性や、その強弱を抽出する。前処理を図5に示す。3秒間の窓幅ごとに自己相関を用い、各時間のビートのピークを

STFTに関する特徴量(全33種類) ()内は特徴量要素数

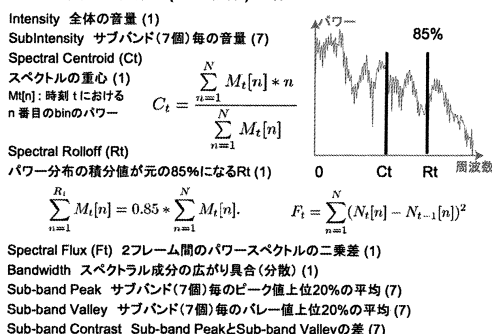


図4 音色に関する特徴量 (STFT) の内訳

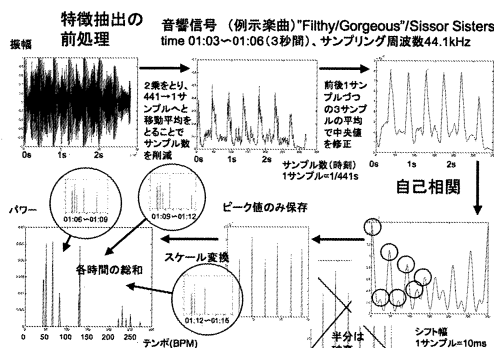


図5 リズムに関する特徴量抽出の前処理

成すパワーを抽出し、時間軸方向での総和を求める。変換されたスケールを対象に、以下に示す合計6個をリズムの特徴量とする。()に要素数を示す。

- 最も大きなピークのパワーがパワーの総和に対して占める割合 (t_1)(1)
- 2番目に大きなピークのパワーがパワーの総和に対して占める割合 (t_2)(1)
- 上記 t_1 と t_2 の割合 (t_1/t_2)(1)
- 最も大きなピークが示すテンポ (1)
- 2番目に大きなピークが示すテンポ (1)
- 単位時間 (3 秒間) あたりのパワーの総和 (1)

3.5 特徴量の次元圧縮

特徴量によってクラスを分離する能力は異なり、各特徴量のクラス分離尺度を (クラス内分散 / クラス間分散) によって算出する。また抽出された特徴量はお互い相関の高いものが含まれている可能性が高く、ジャンル識別の性能を低下させる一因となる。そこで我々は相関係数が一定の値 r より高い組み合わせにおいて、クラス分離尺度の低い特徴量を削除する方針で次元圧縮を行う。最後に残った特徴量の分離尺度の大

きい上位 s_1 を対象に主成分分析を行い、さらに分離尺度の大きい主成分上位 s_2 を実験に用いる特徴量として選択する。圧縮の過程で設定した r, s_1, s_2 は実験に応じて設定する。

3.6 設定した特徴量による実験への応用

以上の過程によって抽出した「ベースラインの特徴量 (全 41 個)」「音色の特徴量 (全 76 個)」「リズムの特徴量 (全 6 個)」を実験に用いる。4 章ではジャンル認識率と特徴量の有効性を示すことを目的としたマハラノビス距離を用いた識別実験を、5 章では Music Island を用いた楽曲のマッピングについて議論する。

4. マハラノビス距離を用いたジャンル識別実験

前章で述べた特徴量を用いてジャンル認識実験を行い、ベースラインの特徴量の有効性を示す。

4.1 データセット

識別対象となる音楽ジャンルは ISMIR2004 のジャンル分けコンテスト¹²⁾ と同様、[Pop/Rock], [Metal/Punk], [Electronic], [Jazz/Blues], [Classical], [World] の 6 ジャンルに設定、学習用、実験用の音楽ファイルは ISMIR から web 上¹³⁾ で配布されている音源の一部 (各ジャンル 50 曲 × 6 ジャンルの 300 曲) を用いる。また識別には計算量を考慮し演奏時間のうちの 1 分間のみに焦点をあてる。ベースパートが演奏を行わない可能性がある前奏を避け、抽出対象を各楽曲演奏開始 1 分から 2 分の 60 秒間に固定した。

4.2 識別手法

ジャンル識別の際に用いる距離尺度としてマハラノビス距離を採用する。マハラノビス距離とは分布の広がりや考慮に入れた確率的距離であり、クラス c (各ジャンル) に対するマハラノビス距離 D の 2 乗は次式で定義される。

$$D_c^2 = (\mathbf{x} - \boldsymbol{\mu}_c) \boldsymbol{\Sigma}_c^{-1} (\mathbf{x} - \boldsymbol{\mu}_c) \quad (1)$$

ここで

$\mathbf{x} = (x_1, x_2, \dots, x_n)^t$: 実験楽曲の特徴量縦ベクトル
 $\boldsymbol{\mu}_c = (\mu_{c1}, \mu_{c2}, \dots, \mu_{cn})^t$: クラス c の平均値縦ベクトル (t は転置行列)
 $\boldsymbol{\Sigma}_c^{-1}$: クラス c の母集団の分散・共分散行列

各ジャンルのクラスに対するマハラノビス距離を比較し、実験楽曲から最も距離の近いクラスにあたるジャンルを第 1 推定として決定する。また実験の際には用意した音楽ファイルのうち各ジャンルの楽曲を訓練データ 49 曲、実験データ 1 曲に分けて 50 回繰り返す Leave-one-out cross validation を用いる。

表 1 実験結果の比較 (認識率)

(実験1) 音色とリズムの特徴量		全ジャンル Total正解率 54.3%					
	Pop &Rock	Metal &Punk	Elect ronic	Jazz &Blues	Classi cal	World	
Pop&Rock	21	6	11	0	0	3	
Metal&Punk	13	42	8	0	0	4	
Electronic	3	1	5	0	0	0	
Jazz&Blues	7	1	12	48	4	7	
Classical	2	0	9	0	31	20	
World	4	0	5	2	15	16	
正解率	42%	84%	10%	98%	62%	32%	

(実験2) ベースラインを加えた全特徴量		全ジャンル Total正解率 62.7%					
	Pop &Rock	Metal &Punk	Elect ronic	Jazz &Blues	Classi cal	World	
Pop&Rock	23	5	10	5	0	7	
Metal&Punk	10	42	1	1	0	1	
Electronic	8	0	23	1	0	6	
Jazz&Blues	5	2	6	37	2	10	
Classical	0	0	3	1	45	8	
World	4	1	7	5	3	18	
正解率	64%	84%	46%	74%	90%	36%	

(縦軸) 識別結果ジャンル / (横軸) 実験データのジャンル

我々はベースラインの特徴量の有効性を示すため、以下の2パターンで実験を行う。

- (実験1) 音色とリズムの特徴量 (全 82(76 + 6) 個) による実験
- (実験2) ベースラインを加えた全特徴量 (全 123 個) による実験

それぞれ識別に有効な特徴量を前章で述べた次元圧縮の過程に従って圧縮する。実験1においては相関係数のしきい値 ($r = 0.65$)、分離尺度による特徴量選択数 ($s_1 = 8$) とし、実験2においては ($r = 0.5$)、($s_1 = 7$) とする。また最終的にジャンル識別に用いる主成分は、両実験とも分離尺度において上位の2次元 ($s_2 = 2$) とする。

4.3 実験結果

両実験の結果を表1と図6に示す。全ジャンル総合の認識率を比べると音色とリズムの特徴量のみを用いた実験1の54.3%に対し、ベースラインを含む特徴量での認識率は62.7%に上昇した。ジャンル別に見るとElectronicにおける認識率向上が著しい。楽曲を主成分の軸でプロットした図6を見比べると、実験2においての第2主成分(因子負荷量による意味解析の結果、ベースパートにおける使用音高の多様性)がElectronicの分離に貢献していることがわかる。

また、各実験において識別に有効であった上位の特

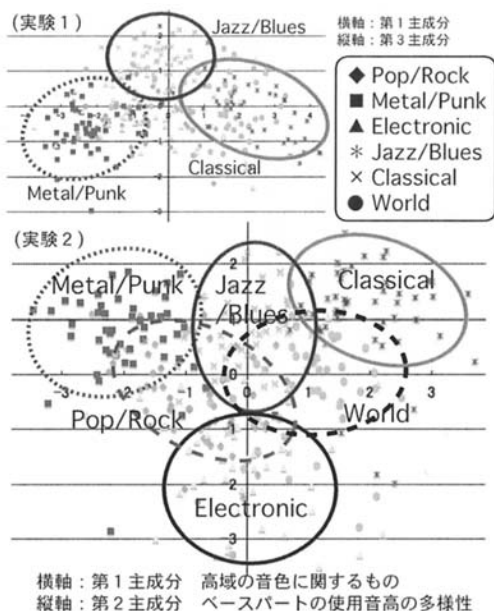


図 6 実験結果の比較 (散布図)

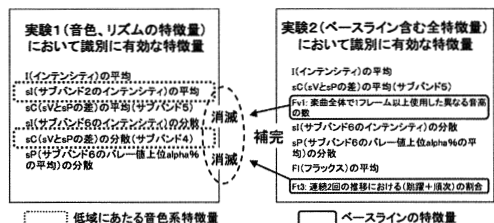


図 7 各実験の識別に有効な特徴量の比較

徴量を図7に示す。ほぼ同等の特徴量が上位に並び、中、実験2では実験1で上位を占めた低域の音色に関する特徴量が消え、代わりにベースラインに関する特徴量が名を連ねている。これは低域における特徴量は、音色系よりもベースラインの特徴量の方が識別に有効であることを示唆している。以上の実験結果より、ベースラインの特徴量はジャンル識別に有効であり、他の特徴量と併用することにより全体の識別精度が向上することが示された。

特定の楽曲において識別が上手くいかなかった原因の1つとしては、ベースラインのF0推定において雑音を取りきれていない問題が考えられる。図8に例としてGreenDayの『Minority』における雑音除去後のベースラインのF0を示す。図より雑音はベースが音高推移するタイミング、すなわち拍節においての発生が顕著である。原因としてバスドラム等の低音にあ

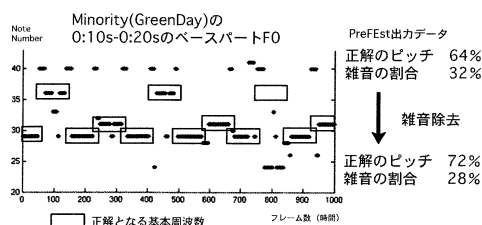


図 8 雑音除去後のベースライン基本周波数 (F0)

表 2 各実験の識別に有効な特徴量の比較

横軸: 実験楽曲 縦軸: 識別ジャンル	Pop/Rock		Metal/Punk		Electronic		Jazz/Blues		Classical		World	
	T	F	T	F	T	F	T	F	T	F	T	F
Pop/Rock	7	4	3	6	4	5			2			1
Metal/Punk	3		7	3								
Electronic		6		1	4	5						
Jazz/Blues					1		10	8				2
Classical									10	1		
World					1					7	10	7
合計正解人数	11/20		10/20		9/20		18/20		11/20		17/20	

各ジャンル2曲ずつ実験 (Tは計算機が正解した楽曲、Fは不正解した楽曲)

数字は10名の被験者による第1推定ジャンルの合計人数

Total正解率 63%

マハラノビス距離による実験において計算機が識別したジャンル

計算機の間違いと人間の間違いが一致した箇所

る F0 がより優勢な F0 として選ばれた可能性が高い。この問題の解決方法としては打楽器の分離を行う吉井ら¹⁴⁾ や宮本ら¹⁵⁾ の手法を取り入れたドラムパートによる雑音の除去が有効となる可能性がある。

4.4 被験者によるジャンル識別との比較

識別が上手くいかなかった他の原因として、音楽ジャンルという概念の定義があいまいであることが考えられる。楽曲の中には多ジャンルの要素を含んでいるものや、評価者によって識別が異なるものが存在する。そこで被験者を募り、人間によるジャンル識別結果と実験 2 において計算機が識別した結果を比較する。

被験者は日常生活で音楽をよく聞く人 10 名、識別を行う楽曲は全 6 ジャンルの楽曲のうち実験 2 において計算機が正解した楽曲と不正解した楽曲各 1 曲ずつの合計 12 曲とする。被験者には最も尤もらしいと感じるジャンル 1 つを第 1 推定として回答してもらった。表 2 に識別結果を示す。実験結果によると、識別機が誤認識した楽曲は被験者も同じ間違いをしている箇所が多いことが示された。また全楽曲総合での正解率は 63 % となり、人間による識別は我々が実装した識別機の性能とさほど変わらない結果となった。

マハラノビス距離を用いた実験を通して、ベースラインの特徴量がジャンル識別において有効であることを示した。また実験により人間によるジャンル識別と比べた結果、計算機におけるジャンル識別の精度向上は限界に近いことも示された。

5. Music Island を利用したジャンル分析

音楽ジャンルの定義はあいまいであり、楽曲とジャンルを完全に 1:1 に対応させることは困難である。人はそれぞれ音楽の楽しみ方が異なり、音楽を聴く際に音色を重視する人や特定の楽器パートを重視する人など、類似尺度はユーザーに依存する性質を持つ。そこで本章では、Elias Pampalk の提案した Music Island⁶⁾ を用いることでユーザーのインタラクションに応じた類似楽曲のマッピングを行う。

5.1 SOM アルゴリズムと Music Island

Music Island⁶⁾ には自己組織化マップ (SOM) のアルゴリズムが用いられている。SOM とは Kohonen¹⁶⁾ により発表された教師なし学習ニューラルネットワークであり、多次元のデータを圧縮して低次元のマップを描くことで似た構造を持つデータは近くに、そうでないものは離れた位置に配置する。Music Island は可視化システムにあたり、データが密集する場所で島を形成し、更に密集度の高い場所では山、低いところでは海を形成する。Music Island は概念空間を視覚化する方法として非常に優れている。

5.2 Music Island 適用の向けての設定

学習に用いる特徴量として「音色とリズムの特徴量 (全 82 個)」と「ベースラインの特徴量 (全 41 個)」の 2 種類を用意する。「音色とリズムに関する特徴量」とは 3 章で扱った「音色の特徴量 (全 76 個)」と「リズムの特徴量 (全 6 個)」を組み合わせたものにあたる。それぞれ 3.5 節で示した相関係数とジャンル分離尺度を用いた次元圧縮を行う。各設定値は「音色とリズムの特徴量」において ($r = 0.65, s_1 = 8$)、リズムの特徴量においては ($r = 0.5, s_1 = 5$) とし、最終的に主成分分析により得られたジャンル分離能力の高い上位 3 次元ずつ ($s_2 = 3$) の合計 6 次元を利用する。上位 3 次元の累積寄与率 (元の特徴量の情報をどの程度反映しているかを示す指数) は「音色とリズムの特徴量」において 79 %、「ベースラインの特徴量」において 84 % を保持している。

実験の対象楽曲はマハラノビス距離による実験で用いたものと同じ、50 曲 × 6 ジャンルの合計 300 曲とする。マップは 14 × 14 のノードに設定し、各特徴量によるマップの変化を考察するため、基準となる初期配置を固定する。初期配置の設定には、「ジャンルの情報を与えたメタデータ (6 個) - 特定のジャンルにのみ 1 を、それ以外のジャンルに 0 を設定したもの」を用意し、計 12 次元の特徴量によって作成する。マップに表示されている「PR3」等は「ジャンルの頭文字

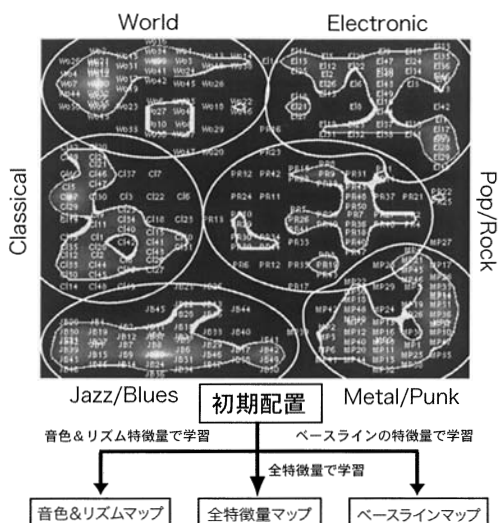


図 9 Music Island と特徴量の比重



図 10 全特微量を用いた Music Island

(Pop/Rock) + 楽曲ナンバー (1~50)」を示している。実験の全体像を図 9 に示す。初期配置マップより、重視したい特徴量を指定することで「全特微量を用いたマップ」「音色とリズムのマップ」「ベースラインのマップ」の切り替えを行える、ユーザー適応型の検索機を目標とする。実際の計算には Matlab の SOM Toolbox(version2)¹⁷⁾ と SDH Toolbox¹⁸⁾ を用いた。

5.3 実験結果

図 10 は基準となるマップから、本研究で設定した全特微量を用いた学習によって作成されたマップである。変化の特徴としては各ジャンルの楽曲集合が範囲を広げており、他ジャンルの範囲と混ざり合うことで

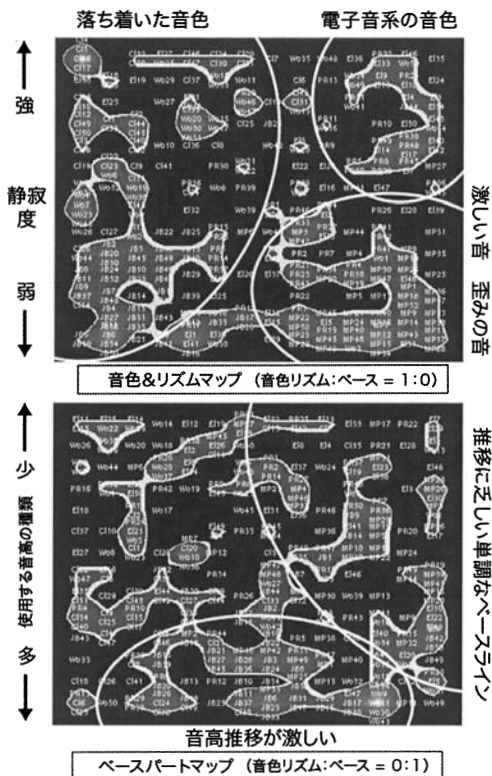


図 11 特定の特徴量だけに注目した Music Island

複数のジャンルのにおける包含関係を表している。特に Pop/Rock や World は集合がほぼ消滅しており、楽曲単位においてそれぞれ類似性の高いジャンルの集合の中に配置されている。具体的には Electronic と Metal/Punk 系の両要素をもつ楽曲の集合や、Classical 風要素をもつ World の楽曲が視覚的に理解できるなど、一つのジャンルへの対応では表現できなかった柔軟な楽曲配置を可能とした。

また図 11 は「音色とリズムのマップ」と「ベースラインのマップ」を示したものである。それぞれジャンルという概念がほとんど消え、各特徴量の特性に応じた類似楽曲の近隣配置が可能となった。「音色とリズムのマップ」においては歪みを伴う激しい音色の楽曲集合や電子音系の音色をもつ楽曲集合、また落ち着いた音色をもつ楽曲集合と位置関係によるその静寂度の違いが確認された。「ベースラインのマップ」においては音高推移の頻度や、楽曲を通して使用する音高のバリエーションに依存した配置が確認された。この配置はベース単独のパートを好んで聴くユーザーにとって有用性が高いと考えられる。

6. 議 論

本研究で用いたような特定パートから抽出した特徴量は、これまでの音楽情報検索ではほとんど用いられてこなかった。この理由の1つに音源分離の難しさがあげられる。一方、ベースやドラムなどの聴感上目立つ楽器については、有効な抽出手法が提案されている^{8),14),15)}。しかし、音楽情報検索への応用には至っていなかった。特定パートから抽出した特徴量を実際に音楽情報検索へ応用し、有効性を確認したのは、本研究がはじめてである。

個々のパートから特徴量を抽出する1つのメリットは、特徴量に重みづけをすることで、リズム重視やベースラインの動き重視など、ユーザの聴きかたに合わせた検索が実現できることである。本稿では、5章でMusic Islandを作る際に特徴量の重みを変えることで、その重みに合ったMusic Islandを得られることが確認できた。これは、ユーザ適応型の音楽検索システムを実現する上での重要な1歩だと考える。

7. ま と め

我々は音響信号を対象とした音楽ジャンル解析において、高次の特徴量であるベースパートの特徴量を取り入れて実験を行った。マハラノビス距離を用いた6ジャンルでの識別においては、従来からの「音色とリズムの特徴量」を用いた実験での54.3%の認識率に対し、「ベースラインを含む全特徴量」を用いた実験においては62.7%の高認識率を実現した。またMusic Islandを用いた実験を行い、重み付けした特徴量に応じてマップを自在に切り替えることでユーザーの聴き方に合わせた楽曲分類を実現した。今後はベース以外のパートにも焦点をあて、ユーザーの様々な好みにより柔軟に対応できるシステムの実現を目指したい。

謝辞 この研究は、西山正紘氏の作成した音色系特徴量の抽出プログラムと、Elias Pampalk 博士によって提供されたMusic Islandのシステムを利用した。ここに記して謝意を表す。

参 考 文 献

- 1) ISMIR 2007, 8th International Conference on Music Information Retrieval, Vienna, Austria, September 23-27, 2007, Proceedings (2007).
- 2) Basili, R., Serafini, A. and Stellato, A.: Classification of musical genre: a machine learning approach, *ISMIR*, pp. 505-508 (2004).
- 3) McKay, C. and Fujinaga, I.: Automatic Genre Classification Using Large High-Level Musical Feature Sets, *ISMIR*, pp. 525-530 (2004).
- 4) Dion, S., Couyon, F. and Widmer, G.: Towards Characterisation of Music via Rhythmic Patterns, *ISMIR*, pp. 509-516 (2004).
- 5) Tzanetakis, G. and Cook, P.: Musical genre classification of audio signals, *IEEE Trans. Speech Audio Process*, Vol. 10, No. 5, pp. 293-302 (2002).
- 6) Pampalk, E., Dixon, S. and Widmer, G.: Exploring music collections by browsing different views, *ISMIR*, pp. 201-208 (2003).
- 7) 土橋佑亮, 片寄晴弘: SOM を用い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, Vol. 2006, No. 90, pp. 31-336 (2006).
- 8) 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 (ISCA Journal)*, Vol. 43, No. 4, pp. 311-329 (2004).
- 9) 藤岡敏則著: ベース演奏技法, 株式会社ヤマハミュージックメディア (2005).
- 10) 西山正紘, 北原鉄朗, 駒谷和範, 尾形哲也, 奥乃博: マルチメディアコンテンツにおける音楽と映像の調和度計算モデル, 情報処理学会研究報告, Vol. 2007, No. 15, pp. 31-36 (2007).
- 11) Lu, L., Liu, D. and Zhang, H. J.: Automatic Detection and Tracking of Music Audio Signals, *IEEE Trans. Audio, Speech and Language Processing*, Vol. 14, No. 1, pp. 5-18 (2006).
- 12) ISMIR 2004, 5th International Conference on Music Information Retrieval, Barcelona, Spain, October 10-14, 2004, Proceedings (2004).
- 13) ISMIR2004 ジャンル分けコンテスト データベース, <http://ismir2004.ismir.net/index.html>.
- 14) Yoshii, K., Goto, M. and Okuno, H. G.: Drum Sound Recognition for Polyphonic Audio Signals by Adaptation and Matching of Spectrogram Templates With Harmonic Structure Suppression, *IEEE Trans. Audio, Speech and Language Processing*, Vol. 15, No. 1, pp. 333-345 (2007).
- 15) 宮本賢一, 立藺真理, ルルージョナトン, 亀岡弘和, 小野順貴, 嵯峨山茂樹: スペクトログラム2次元フィルタによる調波音・打楽器音の分離, 日本音響学会秋季研究発表会講演論文集, Vol. 2007, No. 1-1-7, pp. 825-826 (2007).
- 16) *Self-Organizing Maps*, Springer, 3rd edition (2004).
- 17) SOM Toolbox: Laboratory of Information and Computer Science in the Helsinki University of Technology, <http://www.cis.hut.fi/projects/somtoolbox/>.
- 18) Pampalk, E., Rauber, A. and Merkl, D.: *SDH Toolbox*, <http://www.ofai.at/~elias.pampalk/sdh/>.