

楽曲-景観データに基づく音響特徴量の分析

福本 颯太[†] 奥 健太[†]

[†] 龍谷大学理工学部 〒520-2194 滋賀県大津市瀬田大江町横谷 1-5
E-mail: †t140495@mail.ryukoku.ac.jp, ††okukenta@rins.ryukoku.ac.jp

あらまし 近年、カーオーディオや楽曲再生に関する技術の向上によってドライブ中に好みの音楽を流して楽しむことが可能になった。我々は、ドライブ時の景観に適した楽曲を再生する景観アウェア楽曲推薦システムの研究に取り組んでいる。このようなシステムの実現に向けて、本研究では独自に収集した楽曲と景観の関係データを基にして、景観画像を想起させる楽曲の特徴量を分析する。得られた特徴量を用いて学習モデルを生成し、景観との対応付けが未知の楽曲に対して景観タグの推定を行う。

キーワード 楽曲推薦, 音響特徴, 特徴分析

1. はじめに

カーオーディオや楽曲再生に関する技術の発達によって車の運転中に自信の好みや気分にあった音楽を聴きながらドライブを楽しむことが可能になった。このような状況の中、ドライブ時に適した楽曲推薦システムを構築することでより快適なドライブ環境が実現可能であり、ドライブ時のコンテキストにあった楽曲を推薦するコンテキストアウェアドライブ楽曲推薦システムに関する研究 [1] [2] が行われている。

我々は、ドライブ時の景観に適した楽曲を再生する景観アウェア楽曲推薦システムの研究に取り組んでおり、これまでに景観アウェア楽曲推薦システム [3] を提案してきた。このシステムは、クラウドソーシングにより収集した楽曲-景観関係データに基づき楽曲を景観ベクトルで表現することで、入力された景観にマッチする楽曲を推薦するものである。

本研究では、先行研究 [3] にならって収集した楽曲-景観関係データを音響特徴の観点から分析する。先行研究 [3] では、楽曲を田園系、山林系、水辺系、都市系を要素とした景観ベクトルを定義しているため、それぞれの景観要素の強さと音響特徴量との相関を分析しどのような景観のときにはどのような音響特徴をもつ楽曲が選ばれやすいか検証する。なお、本研究では、音響特徴量として、RMS や ZeroCross, Tempo など 10 種類の特徴量に着目している。景観要素と音響特徴量との関係を明らかにすることで、今後、楽曲-景観関係データが得られていないようなデータに対しても楽曲の景観ベクトル化を行うことができる。

2. 関連研究

2.1 楽曲推薦システム

楽曲推薦システムは、楽曲コンテンツ過多の問題に対し、ユーザの楽曲への嗜好 (ユーザプロファイル) に基づき、ユーザに合った楽曲を見つけることを支援するプログラムである [4]。楽曲推薦システムの例として、Last.fm^(注1) が挙げられる。これは

インターネットラジオの 1 つであり、楽曲を流すほかに楽曲推薦システムを用いて楽曲の推薦を行っている。Last.fm では、ユーザがお気に入りとして保存した好きな楽曲に関係した情報に加え、ユーザの聴取履歴に基づいたユーザの楽曲への嗜好を予測することで楽曲推薦を行っている。

2.1.1 コンテキストアウェア楽曲推薦システム

ユーザが楽曲を聴取する際どのような楽曲を選択するのかは、ユーザのそのときのコンテキスト (時間や場所, 感情, 現在の活動状況など) に影響を受けるという考えがある。これに基づき、ユーザのコンテキストを考慮した楽曲推薦システムが注目されている [5]。

Kaminskas ら [6] は、コンテキストの一つである位置情報に着目した楽曲を推薦する問題に取り組んでおり、タグベースおよび知識ベース手法を組み合わせたハイブリッド型の楽曲推薦システムを提案している。本研究では、ドライブ時に関係するコンテキストに基づく楽曲推薦を目指すためにドライブ時に変化する景観をコンテキストとして扱う。Baltrunas らは、車内で楽曲推薦を行うシステム [2] を提案しており、道路風景や天候などのコンテキスト因子と 10 種類の音楽ジャンルとの関連性について評価値付けの観点から分析している。本研究では、楽曲と景観のマッチングされたデータを基に景観特徴ベクトルと音響特徴ベクトルに基づいた分析を行う。

2.2 音響特徴解析ツール

近年では、音響特徴を解析する際に用いる様々なツールが開発されている。

Marsyas^(注2) は、コマンドラインから音響特徴量を抽出が可能のほか、基本的な音響信号処理に対して有用な独自のスクリプトがある。

Jaudio^(注3) は、GUI に楽曲ファイルを入力し必要な音響特徴量を指定することで簡単に音響特徴量を抽出することが可能である。また、GUI 上で出力形式が XML 形式か ARFF 形式かの選択ができるため、分析手法によってデータの形式を簡単に

(注1) : <http://www.last.fm>

(注2) : <http://marsyas.info/>

(注3) : <http://jaudio.sourceforge.net/>

変更することができる。

MIRtoolbox^(注4) は、MATLAB 上に実装された音響特徴量の分析パッケージであり、MATLAB は行列計算などの関数が豊富であるため、機械学習などを考慮する場合において有用である。

3. 準備

3.1 定義

定義 1 : 景観要素. 奥らの先行研究 [7] では、予備実験により、道路景観を構成する基本要素として、田園系 (r), 山林系 (m), 水辺系 (w), 都市系 (u) の四つの要素を定義した。これらの要素を景観要素とよぶ。

定義 2 : 楽曲-景観関係データ. 楽曲-景観関係データは楽曲集合 T と 4 種類の景観要素の関係を表したデータであり、二通りの表現方法がある。一つは楽曲-景観関係行列による表現であり、もう一つは楽曲-景観関係テーブルによる表現である。本研究では、後者の楽曲-景観関係テーブルの表現を用いる。

楽曲-景観関係テーブルは次の関係スキーマをもつ：

```
tracks-landscape(track_id, rural, mountain,
                 water, urban)
```

ここで、`track_id` は楽曲 ID である。`rural, mountain, water, urban` は各景観要素に対応するカラムであり、それぞれ田園系、山林系、水辺系、都市系の各要素に対応する。1 楽曲は 1 タプルで構成される。ここで、`track_id = $t_i \in T$` における景観要素 $l \in \{\text{rural, mountain, water, urban}\}$ のカラムには、楽曲 t_i が景観要素 l に関連する確率を表す。以降、これらカラムの値のことを景観要素値とよぶ。なお、任意の楽曲 $t_i \in T$ について景観要素値の総和は 1 となる。

3.2 楽曲の音響特徴

本研究では、楽曲の音響特徴量として下記の特徴量に着目する。なお、括弧内の記号は以降の参照の際に用いる：

RMS (RMS)

音響信号の実効値を表す。音響信号の物理的な強度と関連する。算出式は次式であり、 n は解析窓内のサンプル数、 x_i は第 i 番目のサンプル値を示している：

$$y_{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i)^2} \quad (1)$$

ZeroCross (ZC)

音響波形の振幅値が 0 との交差回数を表す。ノイズの量を表す指標として知られている。算出式は次式であり、 x_i は第 i 番目のサンプル値を示しており、 t は解析窓の開始時刻、 t' は解析窓の終了時刻を示している。 θ は交差の有無を表しており、交差した場合に 1 を、そうでない場合に 0 の値を出力する：

$$y_{zc} = \sum_{i=t}^{t'} \theta \begin{cases} 1 & (x_i x_{i-1} < 0) \\ 0 & (x_i x_{i-1} > 0) \end{cases} \quad (2)$$

SpectralCentroid (SC)

パワースペクトルの重心点の周波数を表す。音響信号の明るさに関係する指標である。算出式は次式であり、 x は周波数を示し、 $f(x)$ は x の時のパワーの値を示す：

$$y_{sc} = \int x f(x) dx \quad (3)$$

Spread (SP)

パワースペクトルにおいて、SpectralCentroid からの分散値を表す。算出式は次式である：

$$y_{sp} = \sigma^2 = \int (x - y_{sc})^2 f(x) dx \quad (4)$$

SpectralRolloff (SRO)

パワースペクトルにおいて、低周波数帯域から全体の 85% を占めるエネルギー量を指す周波数の値を表す。音響信号の明るさに関係する指標である。算出式は次式であり、 $M_t[k]$ は t 番目の窓をフーリエ変換して得られるスペクトルにおける、 k 番目の周波数ビンにおける振幅値を示す：

$$y_{sro} = 0.85 * \sum_{k=1}^K M_t[k] \quad (5)$$

Brightness (BN)

パワースペクトルにおいて、全エネルギー量に含まれる 1,500Hz 以上のエネルギー量の割合を表す。音響信号の明るさに関係する指標である。

Flatness (FN)

ある時刻において周波数解析を行った区間のパワースペクトルを周波数軸方向に任意のサブバンドに分け、全サブバンド数のパワーの相乗平均を相加平均で除算することで得られる値を表す [8]。高さをもった音らしさに関連する。算出式は次式であり、 n はサブバンドの値を示し、 x はパワーを示している：

$$y_{bn} = \frac{\sqrt{\prod_{n=0}^{N-1} x(n)}}{\left(\frac{\sum_{n=0}^{N-1} x(n)}{N}\right)} \quad (6)$$

Roughness (RN)

ある時間区間内でのピーク値を計算し、得られたすべてのピーク値の対からすべての不協和音の平均をとることで得られる値を表す [8]。音の粗さを表す指標である。

Tempo (TP)

オンセット時刻から周期性を検出することで得られる値を表す。楽曲の速さに関連する。

PulseClarity (PC)

拍の明瞭さの推定値を表す [9]。音の時間的継続に関連する。

4. データ前処理

本研究では、楽曲-景観関係データに基づき、景観と楽曲の音響特徴量との関係を分析する。本章では、その分析のために必要な楽曲-景観関係データおよび音響特徴量について述べる。

(注4) : <https://jp.mathworks.com/matlabcentral/fileexchange/24583-mirtoolbox?requestedDomain=www.mathworks.com>

4.1 楽曲-景観関係データの収集

3.1 節で定義したとおり、楽曲-景観関係データは楽曲集合 T と 4 種類の景観要素（田園系，山林系，水辺系，都市系）の関係を表したデータである。このような関係データを作成するためには、楽曲 t_i が各景観要素にどの程度の確率で関連するかについて知る必要がある。楽曲 t_i がどの景観に関連しているかという判断は、人間の感性に委ねる部分が多い。実際にはその感性は個人に依存することも考えられるが、本研究では、先行研究 [3] にならってクラウドソーシングを用い、まず一般的な感性の獲得を試みる。クラウドソーシングにより得られた楽曲-景観関係データを楽曲-景観関係テーブルとして登録する。

任意の楽曲には楽曲ファイル（MP3 形式，WAV 形式など）が含まれる。事前に全楽曲について特定の区間を切り出し、評価用楽曲ファイルとしておく。楽曲 $t_i \in T$ について、以下の手順によりクラウドソーシングを用いて楽曲-景観関係データを収集する：

(1) クラウドソーシングの作業（以降、ワーカ）は評価用楽曲ファイルを聴取する。

(2) ワーカは聴取した楽曲と最もマッチしていると感じる景観イメージとして、「田園景観」、「山林景観」、「水辺景観」、「都市景観」、「いずれにも該当しない」の選択肢の中から一つを選択する。何らかの不具合により楽曲が再生されなかった場合は「楽曲が再生されない」を選択する。

1 楽曲あたり複数名のワーカが回答する。1 名のワーカによる 1 楽曲に対する回答を 1 タプルとして

`workers-tracks-landscape` テーブルに登録する。このテーブルは次の関係スキーマで定義される。

```
workers-tracks-landscape(worker_id,
                           track_id, landscape)
```

ここで、`worker_id` はワーカ ID、`track_id` は楽曲 ID、`landscape` は選択された景観要素をそれぞれ示す。

`workers-tracks-landscape` テーブルを基に、楽曲-景観関係テーブル `tracks-landscape` を作成する。ここで、楽曲 t_i に対し景観要素 $l \in \{\text{rural, mountain, water, urban}\}$ のカラムの景観要素値として登録する。ただし、ここでの比率とは各景観要素値の総和を 1 としたときの比率とする。

4.2 楽曲の音響特徴量の抽出

楽曲からの音響特徴量の抽出には MATLAB の `MIRtoolbox`^(注5) を用いる。`MIRtoolbox` による音響特徴量の抽出手順は下記のとおりである：

- (1) 対象の楽曲を MATLAB に入力する。
- (2) 解析窓長と解析窓シフト長を設定する。
- (3) 楽曲から任意の特徴量を抽出する。
- (4) 得られた特徴量を全体平均したのちに楽曲-音響特徴テーブルに登録する。

本分析において、楽曲の入力は `miraudio` のコマンドを使用した。また、`mirrms` や `mirzerocross` などのコマンドを利用しオプションの `Frame` と `SP` を用いることで指定した解析窓区間の音響特徴量を抽出した。`mirstat` のコマンドを使用して楽曲の各窓から抽出した特徴量の全体平均を指す `mean` の値を楽曲の特徴量とした。なお、出力は CSV 形式とした。

以下に例として、RMS の音響特徴量を抽出した際のコマンドを記す：

```
function main(in)
    audio=in;
    DefFrame=1024;
    % RMS
    rms=mirrms(audio, 'Frame', DefFrame, 'sp');
    STAT_rms=mirstat(rms);
    M_rms=STAT_rms.Mean;
```

ここで、`audio` は入力した楽曲を示し、`DefFrame` は解析窓長を示す。

出力された CSV ファイルを基に楽曲-音響特徴テーブルを作成する。楽曲-音響特徴テーブルは次の関係スキーマで定義される：

```
tracks-features(track_id, rms, zc, ..., pc)
```

ここで、`track_id` は楽曲 ID であり `tracks-landscape` テーブルの `track_id` を参照する外部キーである。`rms, zc, ..., pc` は 3.2 節で説明した 10 種類の音響特徴量をそれぞれ示す。

5. 楽曲-景観関係データに基づく景観要素と音響特徴量の関係分析

楽曲-景観関係データに基づき、景観要素と音響特徴量の関係性について分析する。具体的には下記の観点から分析、評価を行う：

- (1) 景観要素値と音響特徴量との相関分析 (5.2 節)
- (2) 景観要素間における音響特徴量の差異の分析 (5.3 節)
- (3) 音響特徴量に基づく景観ラベル予測性能の評価 (5.4 節)

5.1 データセット

楽曲データセットとして、『クラシック名曲サウンドライブラリー』^(注6) のクラシック楽曲の 663 曲を利用した。楽曲の研究利用については当サイト作成者の許諾を得ている。

この 663 曲の楽曲について、4.1 節で述べた方法により、楽曲-景観関係テーブル `tracks-landscape` を作成した。クラウドソーシングにおける評価用ファイルとして、各楽曲ファイルから楽曲長の中心から前後 30 秒間の区間を抽出した。ワーカには 1 楽曲あたり抽出した 30 秒間すべてを聴取してもらった。1 楽曲あたり 30 名のワーカに回答してもらった。したがって、 $663 \times 30 = 19,890$ のタプルからなる `tracks-landscape` テーブルを作成した。

また、同データセットについて、4.2 節で述べた方法により、1 楽曲あたり 10 種類の音響特徴量を抽出し、楽曲-音響

(注5) : <https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox>

(注6) : <http://classical-sound.seesaa.net/>

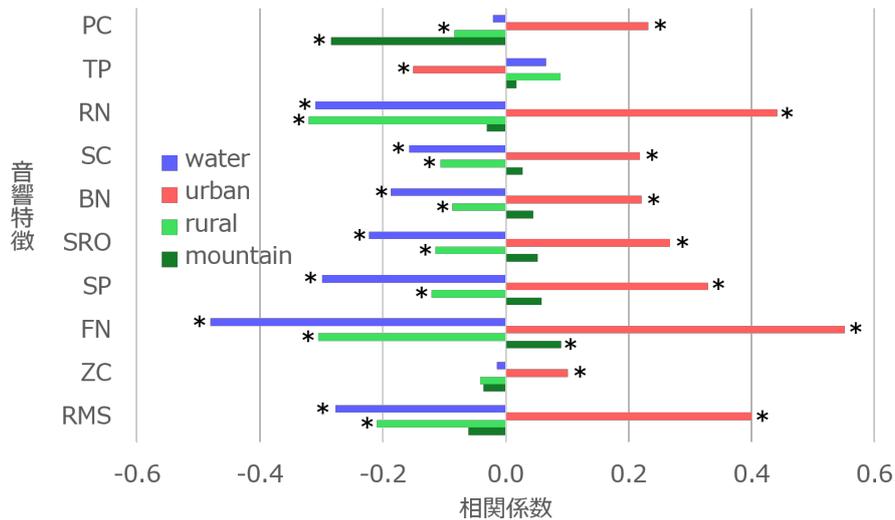


図 1 景観要素値と音響特徴量との相関係数（解析窓長：1,024）．縦軸は音響特徴の属性を表し，横軸は相関係数を表す．ここでは，四つの景観要素それぞれと各音響特徴量との相関係数を示している． t 検定により有意に相関がみられたもの ($p < 0.05$) には * を記している．

特徴テーブル `tracks-features` を作成した．音響特徴量の抽出対象は，先の評価用ファイルと同一の区間とした．切り出したファイル形式は wav 形式であり，サンプリングレートは 16,000Hz，ビットレートは 256kbps である．なお，解析窓長は 1,024，2,048，4,096，8,192，解析窓シフト長は各解析窓長の 1/2 とした．

以上の二つのテーブル，`tracks-landscape` と `tracks-features` を自然結合したテーブル `tracks-landscape` × `tracks-features` を分析に用いる．

5.2 景観要素値と音響特徴量との相関分析

どの音響特徴量がどの景観要素値に関連するか分析する．本節では，`tracks-landscape` × `tracks-features` を基に，楽曲の各景観要素値と各音響特徴量との相関関係を分析する．

図 1 は解析窓長を 1,024 としたときの各景観要素値と各音響特徴量との相関係数を示したものである．図の縦軸は音響特徴の属性を表し，横軸は相関係数を表す．ここでは，四つの景観要素それぞれと各音響特徴量との相関係数を示している． t 検定により有意に相関がみられたもの ($p < 0.05$) には * を記している．

以下，図 1 を基に景観要素ごとに相関関係を考察する．

a) 田園系要素

田園系要素と RMS, FN, RN に弱い負の相関がみられた．このために，音が弱くやや暗い印象を与えるものが水辺系を想起させる曲であると考えられる．

b) 山林系要素

山林系要素と PC に弱い負の相関がみられた．このために，拍があまり明確でなく曲中のテンポの変化があるものが山林系を想起させる曲であると考えられる．

c) 水辺系要素

水辺系要素と FN に負の相関が，RMS, SP, SRO, RN に弱い負の相関がみられた．このために，特定の楽器の音色が強く感じら

れ，音が弱くやや暗い印象を与えるものが水辺系を想起させる曲であると考えられる．

d) 都市系要素

都市系要素と RMS, FN, RN に正の相関が，SP, SRO, BN, SC, PC に弱い正の相関がみられた．このために，音が強いと感じ様々な楽器の音色が混ざり合い聴こえるほか，明るい印象を与え拍がとらえやすいものが都市系を想起させる曲であると考えられる．

全体的な傾向として，田園系要素と水辺系要素は相関関係のある音響特徴が類似することがわかる．これは田園景観と水辺景観は開放的なのでどかなイメージをもつという点で共通的な特徴を有するためであるといえる．実際に田園系要素が強い楽曲および水辺系要素が強い楽曲を聴取してみると，穏やかでゆったりとした楽曲が多いことが確認できた．

田園系要素，山林系要素，水辺系要素が多くの音響特徴に対し負の相関関係がみられたのに対し，都市系要素は多くの音響特徴に対し正の相関関係がみられた．これは前者の 3 景観が自然景観であり，都市景観は非自然景観であるという点に起因していると考えられる．実際に都市系要素が強い楽曲については，他の楽曲に比べ特に活動的で賑やかな曲調が目立つことが確認できた．

5.3 景観要素間における音響特徴量の差異の分析

5.2 節では音響特徴量との相関を景観要素ごとに分析した．その結果，例えば田園系要素と水辺系要素は類似する特徴がみられるなど，景観要素間においても何らかの関連性があることを確認した．そこで，本節では，景観要素間における音響特徴量の差異について分析する．

5.3.1 楽曲データに対する景観ラベルの付与

景観要素間の分析を容易にするために，楽曲データセット内の各楽曲に対し，いずれか 1 個または 0 個の景観ラベルを付与する．いま，楽曲 t_i のタプルが $t_i = (t_i^1, t_i^2, t_i^3, t_i^4)$ であるとすると， t_i^k は楽曲 t_i の景観要素 k に対応する景観要素値を表す．

表 1 景観ラベルごとの各音響特徴量の平均値（解析窓長：1,024）．行は景観ラベルを表し，列は音響特徴の属性を表す．景観ラベル対ごとに U 検定を行ない，すべての景観ラベル対で Bonferroni 法により有意差が得られた数値 ($p < 0.01/3$) はボールド体で示している．

	RMS	ZC	FN	SP	SRO	BN	SC	RN	TP	PC
山林景観	0.050	787.225	0.109	1133.654	2131.437	0.251	1142.108	23.440	151.308	0.151
田園景観	0.051	819.977	0.098	1110.064	2074.926	0.249	1138.022	14.792	151.258	0.188
都市景観	0.048	808.190	0.087	1017.089	1889.681	0.215	1073.794	19.136	151.553	0.219
水辺景観	0.071	895.776	0.136	1249.572	2417.769	0.294	1282.384	44.665	149.721	0.241

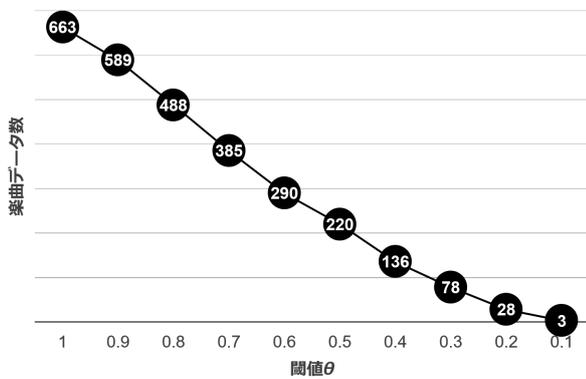


図 2 閾値 θ と景観ラベルが付与された楽曲データ数の関係．横軸は閾値 θ を降順にしたものであり，縦軸は各閾値において景観ラベルが付与された楽曲データ数を表す．白抜き数字は各閾値 θ における楽曲データ数を表す．

このとき，次の手順により楽曲 t_i に対して景観ラベル $l_i \in \{\text{rural, mountain, water, urban}\}$ を付与する：

- (1) 楽曲 t_i の景観要素値の中で最大値をもつ景観要素値 t_i^k を見つける．つづいて，2 番目に大きい値をもつ景観要素値 $t_i^{k'}$ を見つける．
- (2) t_i^k と $t_i^{k'}$ との比率 $t_i^{k'}/t_i^k$ を算出する．
- (3) $t_i^{k'}/t_i^k \leq \theta$ となるとき，楽曲 t_i に景観ラベル $l_i = k$ を付与する．ここで， θ は閾値である．

これは， t_i^k と $t_i^{k'}$ との差が十分に大きいとき，楽曲 t_i の景観ラベル l_i として k が付与されるとみなすという考えからである．いいかえると， t_i^k と $t_i^{k'}$ との差があまりないとき，楽曲 t_i に付与すべき景観ラベルとして一意に決定できないと考えている．今回は，このような一意に景観ラベルを決定できない楽曲は景観ラベル付けの対象から除外する．

つづいて，閾値 θ について議論する．上記の方法では，先述したとおり楽曲データによっては景観ラベルが付与されないこともありうる．閾値 θ を小さく（厳しく）することで，よりその景観要素の特徴が反映された質の高い景観ラベルが付与されるが，一方で分析に使用できる楽曲データ数が減少してしまう．閾値 θ を大きく（緩く）することで，楽曲データ数を多く確保できるが，一方で曖昧な景観要素に対しても景観ラベルを割り当ててしまうため，景観ラベルの質が低くなってしまふ．したがって，十分な楽曲データ数を確保しつつ，なるべく景観ラベルの質を高くする必要がある．

図 2 に閾値 θ と景観ラベルが付与された楽曲データ数との

関係を示す．横軸は閾値 θ を降順にしたものであり，縦軸は各閾値において景観ラベルが付与された楽曲データ数を表す．図 2 に示すとおり，閾値 θ が小さくなるにつれて，楽曲データ数が少なくなっている．図 2 より， $\theta \leq 0.7$ のとき元の楽曲データ数 663 の半数を確保できることが確認できる．したがって，本分析では $\theta = 0.7$ を採用する．ここで景観ラベルが付与された楽曲データ数の内訳は，田園景観ラベルが 63 件，山林景観ラベルが 35 件，水辺景観ラベルが 118 件，都市景観ラベルが 169 件であった．

5.3.2 景観ラベル間の音響特徴量の平均値の差異の分析

景観ラベルごとに楽曲データの音響特徴量の平均値を算出し，景観ラベル間でのその平均値の差異を分析する．表 1 は，景観ラベルごとの各音響特徴量の平均値を示したものである．解析窓長は 1,024 としている．表 1 の行は景観ラベルを表し，列は音響特徴の属性を表す．景観ラベル対ごとに U 検定を行ない，すべての景観ラベル対で Bonferroni 法により有意差が得られた数値 ($p < 0.01/3$) はボールド体で示している．

表 1 を基に，景観ラベル間で有意差関係をまとめたものを表 2 に示す．表 2 は音響特徴量ごとに示している．表中の M, R, U, W は山林景観，田園景観，都市景観，水辺景観を表している．景観ラベル間において有意差がみられた関係には $**$ ($p < 0.01/3$) を記している．

景観ラベル別にみると，表 1 より，山林景観および田園景観ではすべての特徴量で有意差がみられなかった．また，都市景観では RMS, FN, SP, SRO, BN, SC, RN が，水辺景観では FN, SP, SRO, BN で有意差がみられた．この結果から，今回用いた音響特徴量では山林景観と田園景観の識別は難しいといえる．一方で，都市景観では他の景観要素に比べ，音響特徴量による差別化が容易であるといえる．また，水辺景観についても一部の音響特徴量により差別化が可能であるといえる．

つづいて，表 2 より音響特徴量別にみる．FN, SP, SRO, BN については，山林景観と田園景観の対を除くすべての景観ラベル対で有意差がみられた．これらの音響特徴量を用いることで，景観ラベルの識別が容易になるといえる．一方で，TP については，田園景観と水辺景観の対を除くすべての景観ラベル対で有意差がみられなかった．TP は楽曲のテンポを表す特徴量であるが，今回はデータセットとしてクラシック楽曲を対象としたため，テンポが類似する楽曲が多かった．そのため，今回の分析では TP による差異が明確に現れなかったことが考えられる．

5.4 音響特徴量に基づく景観ラベル予測性能の評価

3.2 節に示した音響特徴量を用いたとき，未知の楽曲に対す

表 2 音響特徴量の景観ラベル間の有意差関係. 表中の M, R, U, W は山林景観, 田園景観, 都市景観, 水辺景観を表している. 景観ラベル間において有意差がみられた関係には ** ($p < 0.01/3$) を記している.

	M	R	U	W		M	R	U	W		M	R	U	W		M	R	U	W		M	R	U	W
M	-		**		M	-		**	**	M	-		**	**	M	-		**	**	M	-		**	**
R		-	**		R		-	**	**	R		-	**	**	R		-	**	**	R		-	**	**
U	**	**	-	**	U	**	**	-	**	U	**	**	-	**	U	**	**	-	**	U	**	**	-	**
W			**	-	W		**	**	-	W	**	**	**	-	W	**	**	**	-	W	**	**	**	-

(a) RMS (b) ZC (c) FN (d) SP (e) SRO

	M	R	U	W		M	R	U	W		M	R	U	W		M	R	U	W		M	R	U	W
M	-		**	**	M	-		**	**	M	-		**	**	M	-		**	**	M	-		**	**
R		-	**	**	R		-	**	**	R		-	**	**	R		-	**	**	R		-	**	**
U	**	**	-	**	U	**	**	-	**	U	**	**	-	**	U	**	**	-	**	U	**	**	-	**
W	**	**	**	-	W	**	**	**	-	W	**	**	**	-	W	**	**	**	-	W	**	**	**	-

(f) BN (g) SC (h) RN (i) TP (j) PC

表 3 検証により得られた混同行列. 行が正解ラベルを表し, 列が学習モデルにより予測されたラベルを表す. rural, mountain, water, urban は, それぞれ田園景観ラベル, 山林景観ラベル, 水辺景観ラベル, 都市景観ラベルを表す. 表中の数値は該当ラベルに予測された楽曲データ数を表す. 例えば, 正解ラベルが rural である楽曲データに対し, 学習モデルにより rural と予測された楽曲データ数は 6 である. 正解ラベルに対し正しく予測されたデータ数はボールド体で示している.

		予測ラベル				計
		rural	mountain	water	urban	
正解ラベル	rural	6	17	17	23	63
	mountain	1	18	9	7	35
	water	10	13	73	22	118
	urban	7	12	17	133	169
計		24	60	116	185	385

る景観ラベルの予測性能を評価する. データセットは, 5.3.1 項で景観ラベルを付与した 385 件の楽曲データ (田園景観ラベル 63 件, 山林景観ラベル 35 件, 水辺景観ラベル 118 件, 都市景観ラベル 169) を用いる. 学習モデルとして単純ベイズ分類器を用いる. 評価は 5 分割交差検証法を用いる. すなわち, データセットを 5 分割したとき 4 組を学習データ, 残り 1 組をテストデータとした対で検証を行い, これを 5 回繰り返す.

表 3 は検証により得られた混同行列である. 行が正解ラベルを表し, 列が学習モデルにより予測されたラベルを表す. rural, mountain, water, urban は, それぞれ田園景観ラベル, 山林景観ラベル, 水辺景観ラベル, 都市景観ラベルを表す. 表中の数値は該当ラベルに予測された楽曲データ数を表す. 例えば, 正解ラベルが rural である楽曲データに対し, 学習モデルにより rural と予測された楽曲データ数は 6 である. 正解ラベルに対し正しく予測されたデータ数はボールド体で示している.

表 3 より, water, urban については概ね正確に推定されていることがわかる. 一方で, rural, mountain については誤識別が多かった. このことは, 5.3 節の結果からも説明できる. 表 1 からわかるとおり, rural, mountain においては, すべての音響特徴量において有意差がみられなかった. そのため, 今回用いた音響特徴量では rural, mountain を適切に識別することができなかったといえる.

6. おわりに

本研究では景観アウェア楽曲推薦システムの実現を目指し, 事前に収集した楽曲-景観関係データに基づき, 景観要素と音響特徴量の関係性について分析した. 具体的には下記の観点から分析, 評価を行った:

- (1) 景観要素値と音響特徴量との相関分析.
- (2) 景観要素間における音響特徴量の差異の分析.
- (3) 音響特徴量に基づく景観ラベル予測性能の評価.

分析の結果を踏まえ, 景観ラベルごとに識別に有効な音響特徴量について考察した. 今後は, 本研究の成果を基に, 音響特徴量に基づく楽曲の景観ラベルを自動的に付与する手法の開発に取り組む. さらに, 道路景観の推定技術 [7] と組み合わせることで, 景観アウェア楽曲推薦システムの研究に取り組む.

謝 辞

本研究は JSPS 科研費 JP15K12151 の助成を受けたものです. また, 本研究で利用した楽曲データセットは, 『クラシック名曲サウンドライブラリー』 (<http://classical-sound.seesaa.net/>) のサイト作成者のご厚意により使用許諾を頂きました. ここに記して謝意を表します.

文 献

- [1] Aleksandar Stupar and Sebastian Michel. PICASSO To Sing you must Close Your Eyes and Draw. In *[SIGIR2011]Proceedings of the 34th international ACM SIGIR conference on Research and development in Informa-*

tion, pp. 715–724, 2011.

- [2] Linas Baltrunas, Marius Kaminskas, Bernd Ludwig, Omar Moling, Francesco Ricci, and Aykan Aydin. InCarMusic : Context-Aware Music Recommendations in a Car. In *[EC-Web2011]Proceedings of the 12th International Conference on Electronic Commerce and Web Technologies*, pp. 89–100, 2011.
- [3] 糸井勇貴, 奥健太, 山西良典. 楽曲の風景特徴化に基づく風景ウェア楽曲推薦システム. In *DEIM Forum 2017*, pp. A8–3, 2017.
- [4] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor. *Recommender Systems Handbook*. Springer, 2011.
- [5] Marius Kaminskas and Francesco Ricci. Contextual music information retrieval and recommendation: State of the art and challenges. *Computer Science Review*, Vol. 6, No. 2-3, pp. 89–119, may 2012.
- [6] Marius Kaminskas, Francesco Ricci, and Markus Schedl. Location-aware music recommendation using auto-tagging and hybrid matching. In *[RecSys2013]Proceedings of the 7th ACM conference on Recommender systems*, pp. 17–24, New York, New York, USA, 2013. ACM Press.
- [7] 奥健太, 山西良典. 土地被覆図からの景観要素抽出に基づく道路リンクの景観ベクトル化. 情報処理学会研究報告, 第 2017-DBS-1 巻, pp. 1–6, 2017.
- [8] 山田真司, 三浦雅展. 音楽情報処理で用いられる音響パラメータによる音楽理解の可能性. 日本音響学会誌, Vol. 70, No. 8, pp. 440–445, 2014.
- [9] Olivier Lartillot, Tuomas Eerola, Petri Toivainen, and Jose Fornari. Multi-feature Modeling Of Pulse Clarity: Design, Validation And Optimization. In *[ISMIR2008]Proceedings of the 9th International Society for Music Information Retrieval Conference*, pp. 521–526, 2008.