

# 楽曲データの印象表現に基づいた一覧表示の一手法

草間 かおり<sup>†</sup> 伊藤 貴之<sup>†</sup>

<sup>†</sup>お茶の水女子大学理学部 〒112-8610 東京都文京区大塚 2-1-1

E-mail: <sup>†</sup>{kaori, itot}@itotlab.is.ocha.ac.jp

**あらまし** 近年のマルチメディア技術の発達により、今日では個人が PC やオーディオプレイヤーに楽曲を保有することが主流となった。しかし保有する楽曲数の増加に伴い、ユーザが聴きたい楽曲を見つけることが困難となる場合がある。本報告では、曲名やアーティスト名等のメタデータに依存せず、旋律や調性などの楽曲特徴や印象に基づいて楽曲を検索するツールを提案する。本手法では、楽曲の印象を瞬時に直感的に認識するために、各楽曲データから特徴を検出し、その特徴量に基づいて印象画像を自動生成する。それと同時に、楽曲を階層型にクラスタリングする。そしてそのクラスタリング結果を、印象画像を用いて一覧表示する。

**キーワード** 感性、可視化、印象、楽曲特徴、クラスタリング

## An Impression-based Visualization Technique for Music Collections

Kaori KUSAMA<sup>†</sup> Takayuki ITOH<sup>†</sup>

<sup>†</sup> Faculty of Science, Ochanomizu University 2-1-1 Ohtsuka, Bunkyo-ku, Tokyo, 112-8610 Japan

E-mail: <sup>†</sup>{kaori, itot}@itotlab.is.ocha.ac.jp

**Abstract** Thanks to recent evolution of multimedia technology, we used to store tunes in computers or audio players. The more tunes computers store, the more difficult it is to search the tune they want. This paper proposes a tool that assists to search for the demanded tunes based on impression, not based on metadata (e.g. title and/or artist name). The technique first extracts features of tunes, and then generates impression pictures of the tunes automatically, so that users can recognize the impression of tunes instantly and intuitively. At the same time, it hierarchically clusters the tunes based on the features. It finally visualizes the music data by using the impression pictures.

**Keyword** KANSEI, visualization, impression, musical feature, clustering.

### 1. 概要

音楽の再生機材として PC やポータブル音楽プレイヤーが主流になり、その内蔵記憶装置（ハードディスクや半導体メモリ）の記憶容量の増大に伴い、個人が保有する楽曲数が膨大化している。これにより、ユーザが聴きたい楽曲を見つけるのが困難になることが多くなると考えられる。多くの場合においてユーザは、曲名やアーティスト名などのメタデータをもとに楽曲を検索する。しかし、これらのメタデータに頼らずに、旋律や調性などの楽曲特徴や印象に基づいて楽曲を検索する技術には、まだ検討の余地があると考えられる。例えばカフェやバーの BGM の選曲のように、楽曲のタイトルやアーティストにこだわらず、場所、時間、状況によるムードに基づいて楽曲を選曲したい場面があるとす。あるいはユーザが、アーティストや作曲者に関わらず、ある楽曲について似た曲調の楽曲を見つきたい場合があるとす。このような場合には、楽曲を印象に基づいて分類し、特定の印象を有する楽曲群を簡単に提示できるツールがあると便利である。

以上の背景に基づいて本報告では、膨大な楽曲をメタデータに依存することなく、楽曲そのものの特徴から、印象に基づいて整理、分類し、わかりやすく一覧表示するツールを提案する。

### 2. 関連研究

#### 2.1. 音と感性

感性語による音楽データベース探索システムに関する研究[1]では、楽曲データのサンプリング個数や周波数強度の値に対して、感性語との相関性をそれぞれ求め、楽曲の特徴と感性語の関係を、ニューラルネットワークを用いて検出している。

コレスポンデンス分析による楽曲の特徴認識[2]では、楽曲に対する印象を表す形容詞を評価することによって、音楽の特徴認識には【重い—軽やかさ】、【スピードのある—ゆったりしている】、【power のある—power のない】という 3 軸で認識すると述べている。

#### 2.2. 色と感性

色空間と感性の反映方法[3]では、単色刺激から受ける感性情報の因子として Evaluation(評価性)、

Potency(力量性あるいは潜在性), Activity(活動性あるいは躍動性)があること, また2色配色の刺激から受ける感性情報の因子として Harmony (調和性) が存在することを述べている. このことから, 色を用いて感性を表現することが可能であると考えられる.

カラーシステム[4]では, 色を warm-cool 軸, soft-hard 軸からなる2次元の感性空間に配置することによって, 印象を表現している. その際感性空間における単色配色の分布には偏りがみられ, 感性空間に色が存在しない領域が存在する. 一方多色配色では, 感性領域全体を捕うことができる. また多色配色では, 色の多様な組み合わせ, 色相やトーンの違い, 清色や濁色の違いにより, 様々な効果が生まれる. このことから単色配色よりも多色配色の方が, 色の選択に十分な検討が必要だが的確な印象を表現できる.

### 2.3. 楽曲の印象と画像

印象の合う画像で楽曲を表現する手法に MIST[5]がある. MISTではサンプルとなる楽曲や画像に対して, その感性語適合度をユーザに回答させ, 楽曲や画像の特徴量との相関性を学習させる. それを基に任意の楽曲や画像の感性語適合度を自動算出し, 楽曲の印象に適した画像をアイコンとして表示する.

Music icons[6]では, ニューラルネットワークを用いて適した幾何学模様の画像を選択している. Music icons は予め用意した画像を選択しており, 本研究では印象画像を自動生成することを目的としている.

### 2.4. 楽曲選択の支援

D. Liu[7]は, 楽曲を音量, リズム, 音質に基づいて階層的にクラスタリング手法を提案している. この手法ではまず, 音量によって高階層クラスタを生成した後, リズム, 音質で低階層クラスタを生成する. しかし本手法は楽曲分類に専念するものであり, 表示方法は提案していない.

ユーザによる楽曲選択を支援するインタフェース技術として, Musicream[8]が提案されている. この手法では, マウス操作によって特定の楽曲と類似した楽曲をくっつける機能を有する.

## 3. 処理手順

本手法では楽曲データとして CD に採用されている音響データ書式(wav ファイル)を用いて, (1)楽曲の特徴量の検出, (2)印象画像の生成, (3)クラスタリング, (4)一覧表示の4段階に従って処理を施す.

### 3.1. 特徴量検出

楽曲データから特徴量を検出する. 我々は表1に示す特徴量を, MIRtoolbox[9]を用いて検出している.

楽曲には, 次第に曲の印象が変わるものや, 一つの楽曲に複数の印象を合わせ持った曲があり, 楽曲全体の特徴量の一つに定めるのは困難である. そこで, 一

つの画像を生成するには一定区間ごとに特徴量を検出する. 次第に楽曲が変化する場合各区間の特徴量の平均値をその楽曲全体の特徴量とし, 特定の時刻から明確に印象が変わる楽曲に対しては, 印象が変化する前と後に分けて複数の特徴量を1曲の特徴量とする. 現段階では, 楽曲の中間部のある5秒間から得た特徴量を楽曲全体の特徴量として暫定的に定めている.

表 1. 楽曲特徴

特徴量	説明
RMS energy	音量
Low energy	弱音の割合
Tempo	テンポ
Zero crossing	波形が0値をとる回数
Roll off	85%を占める低音域の割合
Brightness	1500Hz以上の音域の割合
Roughness	不調和な音の多さを示す値
Spectral irregularity	音質の変化の大きさ
Inharmonicity	根音に従っていない音の量
Key	主に使われている音
Mode	major と minor の音量の差

### 3.2. 印象画像の生成

本手法では, ユーザが特定の印象を有する楽曲を直感的に発見するために, 抽象的な印象画像を提示する. 旧来から芸術分野では, 目で見た風景や様子を楽曲で表現する, 曲を聴いた印象を絵画で表現する[10]という試みがなされてきた. このことから, 楽曲の印象を画像で表現することは有効であると考えられる. また人間の感覚には, 音楽や音を聞いて色を感じる「色聴」という知覚があること[11]から, 音と色の印象は密接な関係にある. また, 音と色の印象はしばしば同一の形容詞で表現される. よって, 視覚的に表現され得ない楽曲の印象を可視化する際には, 色印象に重きを置いた印象画像を媒体として表示することにより, 楽曲の背景知識や歌詞の意味よりも楽曲特徴に基づいた印象を重視して, 視覚的表現ができると考えられる.

#### 3.2.1. 色の割当

まず, 印象画像を生成する色を決定する. 前述の通り, 多色配色の方が単色配色よりも豊かな印象を表現することが可能だが, 配色の組み合わせパターンが膨大となり検討が必要なので, 現時点では単色を扱う.

本手法ではカラーイメージスケール[4]をもとに印象画像の色を選択する. カラーイメージスケールは, warm-cool 軸と soft-hard 軸の2次元の感性空間に色を分布させたものである. 楽曲もこの感性空間に分布させ, 空間上で最も距離の近い色をその楽曲の色とする.

そこで, warm-cool 軸, soft-hard 軸に適応する楽曲

特徴量を考える。一般にメジャーコードに明るい・暖かいといったポジティブな印象を表わすコードであり、マイナーコードは冷たい・悲しいといったネガティブな印象を表わすコードであることから、マイナーコードとメジャーコードの音量差を示す特徴量 **mode** を **warm-cool** 軸に割り当てることにした。また低音が少なければ軽い・やわらかい音、低音が多ければ固い・重い音と表わすことに基づいて、低音域の音量を示す特徴量 **roll off** を色の感性空間の **soft-hard** 軸と同等に扱い、楽曲を色の感性空間に分布させることにした。

その後、楽曲の座標値  $(m_{wc}, m_{sh})$  と最も近い座標の色をその楽曲の色とする。

$$color = \min \sqrt{(m_{wc} - c_{wc})^2 + (m_{sh} - c_{sh})^2}$$

### 3.2.2. デザインの生成

続いて、デザインの文法[12]に基づいて、また楽曲の特徴量に応じてデザインを生成する手法を提案する。本報告では一例として、**RMS energy** を背景のグラデーション、**tempo** を円の個数、**Spectral irregularity** を円の配置バランス、**Roughness** を円の半径のばらつき、**Brightness** を星の個数に割り当てるデザイン生成を示す。このデザイン生成の概要を図1に示す。

下からのグラデーションは重みと広がりを感じさせることから、楽曲の重厚感を表すことができる。**RMS energy** は音量を表す楽曲特徴量であるので、**RMS energy** を背景となるグラデーションに割り当てる。

**tempo** は、速い-遅いという形容詞のみならず細かい-粗いと表すことができる。デザインで細かさをオブジェクトの数で表現している。

**Spectral irregularity** は音質の変化の大きさを表す値なので、変化が大きい程不安定である。それをデザインで表現するには、オブジェクトをランダムに配置するインフォーマル構造を採用する。

構図全体が適切な均衡を保つとき、バランスが取れた構図と言える。**Roughness** は不調和な音の多さを示す値なので、値が大きい程バランスが取れていないと考えられる。そこで均衡を崩したデザインをオブジェクトのサイズを不均一にすることによって表す。

**Brightness** は高音域の割合を示している。高音域が多いことを示すには、きらめき・輝きを示す星形のオブジェクトを扱い、高音域の割合が多い程星形のオブジェクト数を増やして表現する。

以上によって生成された3枚の画像を重ね合わせることによって、一枚の画像を生成する(図2参照)。

以上のデザイン生成手法はあくまでも、文献を参考にして我々の主観的に基づいてデザインであり、まだ検討の余地が十分にある。デザイン生成結果の妥当性について今後検証し、被験者からの意見をもとに、楽

曲の印象に合うデザインを考案する予定である。

絶対評価ではなく、一覧表示する楽曲に対して相対的な印象画像を生成するために、楽曲特徴量  $f$  において楽曲特徴量のしきい値  $f_i$  を以下のように定める。

$$f_i = \frac{i(\max(f) - \min(f))}{n} + \min(f) \quad (1 \leq i \leq n)$$

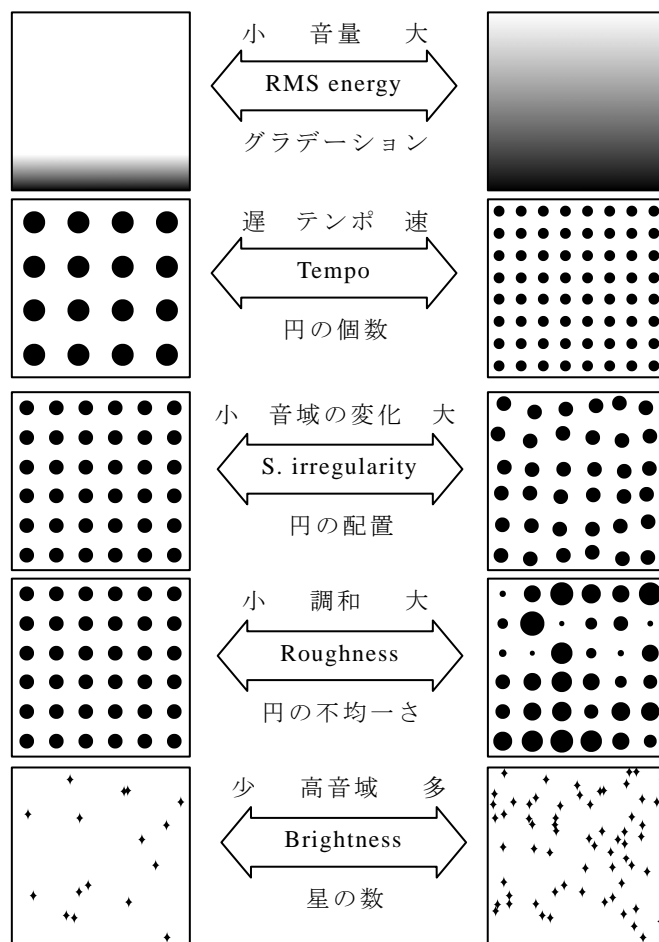


図1. デザインの方法

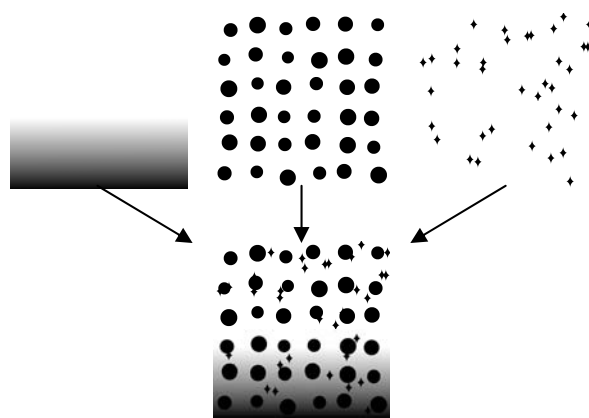


図2. 画像の生成

### 3.3. クラスタリング

続いて、楽曲を特徴量に基づいて階層的にクラスタ

リングする。本研究では、多次元変数クラスタ分析におけるウォード法を用いている。

### 3.4. 一覧表示

生成された画像およびクラスタリング結果を、階層構造を有する画像群の可視化手法 CAT[13]を用いて一覧表示する。図 3(左)に示すように CAT では、各楽曲を表す印象画像をサムネイル表示し、サムネイルを長方形の枠で囲うことでクラスタを表現する。さらに CAT は、ズーム率に合わせた詳細度制御機能として、ズームイン時は各々の画像サムネイルを表示し、ズームアウト時にはクラスタを示す長方形領域を代表画像に置き換えて表示する (図 3(右)参照)。このように、CAT は階層化された画像群に対するズーム操作によって、直感的に情報を絞り込みながら閲覧できる。

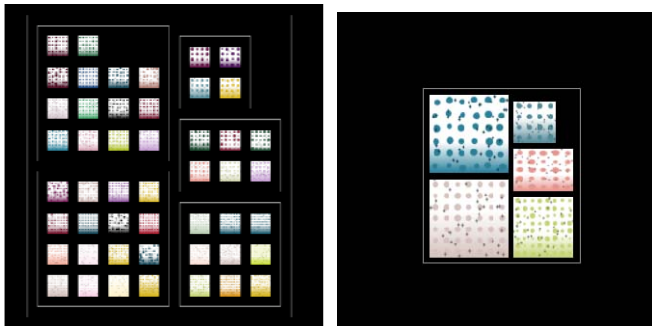


図 3. CAT での表示結果

## 4. まとめ

楽曲を印象に基づいて直感的に一覧表示する手法として、楽曲の特徴量から印象画像を自動生成して、楽曲のクラスタリング結果と合わせて表示する手法を提案した。以下に今後の課題を示す。

### 楽曲特徴量の決定

現時点では、ランダムに採取した 5 秒間の特徴をその楽曲の特徴量としているが、今後一つの楽曲を 5 秒ごとに分割し、分割したそれぞれから特徴量を検出し、最も適した特徴量をその楽曲の特徴量として扱いたい。

### 多色配色への拡張

単色配色に比べ多色配色の方がよりの確に印象を表現することが可能なので、今後さらに文献調査を進め、適切な多色配色に基づくデザイン生成を試みたい。

### デザインの検討

現時点では、楽曲特徴量を印象画像の色やデザインに直接反映している。これは楽曲特徴量を標識のように表していることに近い。今後はさらに、人間の感性や印象に基づくデザイン生成を目指したい。

### 音楽プレイヤーとしてのユーザインタフェースの開発

楽曲の画像をクリックすると楽曲が再生する、クラスタの代表画像をクリックするとクラスタに属する楽

曲を順に再生する、などの操作を付与させたい。

### 正当性の検証

被験者実験を導入することで、(1)楽曲の印象に適した画像が生成されているか、(2)楽曲が印象に基づいて適切に分類されているか、についての検証を行いたい。

(1)は、楽曲特徴量を検出している 5 秒間の前後 5 秒を合わせた 15 秒間を被験者に聞いてもらい、画像に妥当性があるかを、「適している」「やや適している」「どちらともいえない」「あまり適していない」「適していない」の 5 段階で評価してもらうことを考えている。

(2)は、予め見つけたい楽曲を聴いてもらった後、CAT でズームアウト表示した状態を被験者に見せ、ズームイン操作によって目的の楽曲を見つけられるかを観察することを考えている。

## 文 献

- [1] 大塚, 梶川, 野村, “PCM データに対応した感性語による音楽データベース探索システムに関する研究,” 第 14 回電子情報通信学会データ工学ワークショップ(DEWS2003), 8-p-5, 2003.
- [2] 山脇, 椎塚, “コレスポンデンス分析による楽曲の特徴認識,” 感性工学研究論文集, Vol. 7, No. 4, pp. 659-663, 2008.
- [3] 富岡, 三木, 廣安, “色空間と感性の反映方法,” ISDL Report, 20040621002, 2004.
- [4] 小林, カラーシステム, 日本カラーデザイン研究所 (編), pp. 94-123, 講談社, 東京, 2001.
- [5] 小田, 伊藤, “MIST: 音楽アイコンの自動選択の一手法,” 第 15 回インタラクティブシステムとソフトウェアに関するワークショップ(WISS2007), pp. 115-116, 2007.
- [6] P. Kolhoff, J. Preub, J. Loviscach, “Music Icons: Procedural Glyphs for Audio Files,” IEEE SIBGRAPI, pp. 289-296, 2006.
- [7] D. Liu, L. Lu, and H. Zhang, “Automatic Mood Detection from Acoustic Music Data,” ISMIR-03, 2003.
- [8] 後藤, 後藤, “Musicream: 楽曲を流してくっつけて並べることのできる新たな音楽再生インタフェース,” 第 12 回インタラクティブシステムとソフトウェアに関するワークショップ (WISS2004), pp. 53-58, 2004.
- [9] O. Lartillot, “MIRtoolbox,” <http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox>
- [10] 本江, すぐわかる画家別抽象絵画の見かた, 24-33, 東京美術, 東京, 2005.
- [11] J. Harrison, 共感覚—もっとも奇妙な知覚世界, 新曜社, 2006.
- [12] C. Leborg, Visual Grammar: デザインの文法, ビー・エヌ・エヌ新社, 東京, 2007.
- [13] A. Gomi, R. Miyazaki, T. Itoh, and J. Li, “CAT: A Hierarchical Image Browser Using a Rectangle Packing Technique,” 12th International Conference on Information Visualization (IV08), pp.82-87, 2008.