

楽曲から想起される色彩イメージを利用した楽曲推薦システム*

菅野 桃花[†], 伊藤 克亘[†],

1 はじめに

音楽と視覚情報の調和は、情報をより効果的に伝達することにつながるため [1], 音楽が作品全体に与える影響は大きい。そこで、手軽に画像に既存の音楽をつけられる検索システムに需要があるのではないかと考えた。本研究では、視覚情報の中でも特に色に着目し、画像にマッチした音楽を既存の楽曲の中から検索できるシステムを作成する。

2 色彩イメージを利用した楽曲推薦システム

本研究で作成するシステムの概要について説明する。まず、使用者は画像ファイルをシステムに入力する。次に画像ファイルの色情報を抽出する。色情報はHSV(色相・彩度・明度)とRGBのそれぞれ3つの要素を想定している。

次に楽曲推薦アルゴリズムへすすむ。ここでは、事前に楽曲の楽曲情報や音響特徴量をまとめたデータベースを作成する。入力された画像の色情報と、調査結果から導き出した色に対応する楽曲情報・音響特徴量を用い、データベースから適当な楽曲を推薦する。推薦された楽曲と入力の画像の印象がふさわしいかどうかについて、性能評価を実施し評価してもらう。

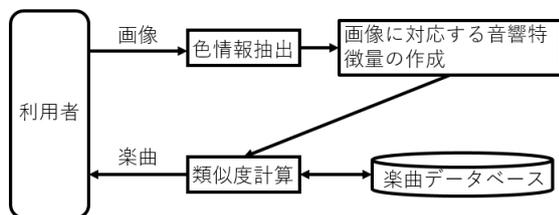


図 1. システムの概要

3 楽曲推薦システム作成のためのアンケート調査

本研究では最初に、色情報に対応する音響特徴量を作成するための調査を行う。具体的には、楽曲から連想する色彩イメージに関する主観評価実験を行い、得られたデータを画像と楽曲の対応付けに用いる。

本調査では、一般の人々に対して色彩イメージについての調査を行うが、共感覚の一つである色聴に関する従来研究 [2, 3, 4, 5] や色調保持者以外も含めて楽曲と色彩(画像・映像も含む)の関係に着目した従来研究 [6, 7] を参考にしてアンケート内容を作成する。共感覚とは、ある物理刺激に対してその感覚器官以外に属するはずの感性反応を引き起こす現象である。色聴とは、音を聴いて色が見える現象であり、このような現象が起こる人はごく一部に限られている。

従来研究では音と色との対応がおおよそ1対1でとれていること、またその対応は個人性を持たないことが明

らかになっている [2]。また楽曲と対応の取れる色の数は20~30色であることも明らかになっている [3, 8]。ここでの楽曲と色の対応については、楽曲を聴取し色を選択する実験に基づくものである。これらのことから、20~30色の対応が確認されることを目指し、後述する楽曲情報や音響特徴量が重複しない楽曲をアンケートに使用する。今回のアンケートではYouTubeのオーディオライブラリの中から30曲を選び、全ての曲を30秒後に切り取ったものを使用する。被験者は研究室の学生6名(男性5名, 女性1名)を対象に行う。回答はWeb上のアンケートフォームで行う。このアンケートフォームは、3年次の予備実験で実施した紙のアンケートや色聴について調査しているアンケートを基に設計し、先述した色彩イメージの回答は画面上のカラーピッカーで行ってもらう。また楽曲に対して色彩イメージがもてない場合は色彩を感じないという選択肢を選んでもらう。

4 楽曲推薦システムの構築

システムの構築は既存の楽曲分類・楽曲検索・楽曲推薦等の研究 [9] を参考に行う。システムは大まかに、画像からの色情報の抽出、抽出した色情報から適切な楽曲を推薦するアルゴリズムの作成の2つの段階に分けられる。後者はさらに、楽曲データベースの作成、色情報に対応する音響特徴量の作成、データベース内の検索システムの作成の3つに分けられる。

4.1 画像からの色情報の抽出

画像からの色情報の抽出は、画像を32色に減色したのちに、使用されたピクセル数が最も多い色を出力する。画像からカラーコードを取得したのちにRGBに変換し、HSVを求める [10]。

4.2 楽曲推薦アルゴリズム

4.2.1 データベースの作成

推薦される楽曲群のデータベースを作成する。楽曲はYouTubeライブラリの中から約500曲を使用する。先行研究 [8, 11] から、ボーカルが加わることで、楽曲の色彩イメージに対する歌詞の影響などを考慮する必要があることが判明している。本研究ではボーカルのない楽曲や、歌詞が意味を持たない楽曲など、出来る限り歌詞の影響を受けない楽曲を中心に使用する。また、データベース化する音響特徴量は先行研究 [9, 12, 13] を参考にし、RMS energy, Low energy, Tempo, Zero crossing, Roll off, Brightness, Inharmonicity, Mode, Key clarity, Spectral irregularity, Keyの11種類を用いる。音響特徴量の分析はMIRtoolboxを用いて行う。

4.2.2 色に対応する音響特徴量群の生成

先述のアンケート調査の結果から、多変量回帰を行うことで、ある色情報にふさわしい音響特徴量を算出するための行列を計算する。式は以下ようになる。

$$y_i = X_i \beta + e_i, i = 1, \dots, n$$

*: A music recommendation system based on query by color contents based music retrieval Momoka Kanno (Hosei Univ.) et al.

[†]法政大学 情報科学部

ここで、 X はアンケート 30 問に対する被験者 6 人の回答した色情報を HSV に変換した 180×3 の行列である。 y はアンケートに使用した楽曲の音響特徴量である。この計算では正規化を行う。アンケートに使用した楽曲 30 曲の音響特徴量に対し平均と分散をあらかじめ求め正規化した値で β を求める。求めた β の要素それぞれに対し、分散をかけ平均値を足す操作を行う。色情報は HSV と RGB の 2 種類の値を用いて行う。

4.2.3 データベース内の検索システムの作成

上で求めた、ある色情報に対応する楽曲の音響特徴量とデータベース内の楽曲の音響特徴量の類似度を計算し、類似度の高い楽曲上位 10 曲のタイトルを検索結果として表示する。今回、類似度の計算はコサイン類似度を用いて行う。

5 システムの性能評価

5.1 評価実験

システムの性能評価実験を行う。被験者は学内の生徒を対象として行う。図のような色数の少ない画像を 10 枚用意し、システムにそれらの画像を入力して推薦された楽曲上位 10 曲、すなわち計 100 曲を聴いてもらい、画像のイメージと推薦された楽曲がマッチしているか評価してもらおう。評価は、ふさわしい、ある程度ふさわしい、ふさわしくないの 3 段階で行う。評価結果から、「ふさわしい」と「ある程度ふさわしい」と評価された楽曲を True とし、Precision を求める。色情報に HSV を用いた時と RGB を用いた場合でどのように差が出るか観察し、有用な方を採用する。



図 2. 評価実験に使用する画像の一例

5.2 実験結果

Precision は HSV を用いた場合が 0.52, RGB を用いた場合が 0.57 となった。

5.3 考察

結果から、HSV よりも RGB を用いたほうがより画像にふさわしい楽曲を推薦することができた。色相は環状になっているため、回帰行列を求める際に工夫が必要であると考えられる。現状では色相が 0 を取る画像と 359 を取る画像で異なる楽曲が推薦されてしまうが、本来はごく近い色であるため似た楽曲が推薦される必要がある。RMS energy, Roll off, Brightness が極端に低いことが多い、ピアノやシンセサイザーを主体としパーカッションがなく静かな楽曲は推薦されにくい。また

アンケートで回答されにくい、赤や黄色、または明度が 100 以下の色情報を持つ画像を入力したときに、推薦される楽曲が偏るという問題点がある。これらの問題は、多変量回帰に用いるサンプル数を増やすことで改善することができる可能性がある [8]。さらに、そもそも画像のイメージが、最も多く使用されている色以外の情報に支配されているという可能性がある。人が画像を見るときに、何に注目しているのかという知見を利用することで、推薦結果の精度を上げることができると考えられる。

6 おわりに

本研究では、共感覚の知見を参考に、楽曲から連想する色彩イメージに関する主観評価実験を行い、実験結果を基に多変量回帰で求めた行列を利用した楽曲推薦システムの実装を行った。評価実験の結果から、色情報に RGB を採用した場合の Precision は 0.57 となり、このシステムを利用することで画像に適する楽曲を探す手間を減らせると考えられる。しかし、システムの精度にはまだ改良の余地があると考えられる。改良には、回帰の処理を工夫することや、人間が画像処理を行う上での心理学的な知見などを生かした処理が必要である。また、実用化を考えると、楽曲データベースの充実や、システムの UI を工夫することが必要である。

参考文献

- [1] 岩宮眞一郎, “音楽と映像のマルチモーダルコミュニケーション”, 九州大学出版会, 2000.
- [2] 野村順一, “色の秘密: 色彩学入門”, 文藝春秋, 2015.
- [3] 山脇一宏 他, “音感と色聴感覚”, エンタテインメント感性特集, 2005, Vol.5, No.3, pp.31-37.
- [4] 山脇一宏 他, “カラーイメージスケールを利用した音楽の特徴抽出”, 知能と情報 (日本知能情報ファジイ学会誌), 2005, Vol.17, No.5, pp.615-621.
- [5] 岩井大輔 他, “音と色のノンバーバルマッピング”, 音楽情報科学, 2002, Vol.47, pp.97-104.
- [6] 伊藤貴之, “文書や画像の印象にもとづく楽曲生成”, The 31st Annual Conference of the Japanese Society for Artificial Intelligence, 2017, No.JSAI2017, pp.2C3OS20a2.
- [7] 川野邊誠 他, “楽曲から受ける印象の時系列変化を考慮した楽曲から配色へのメディア変換”, 芸術科学会論文誌, 2006, Vol.5, No.4, pp.95-105.
- [8] 仲村哲明 他, “色彩想起と歌詞の関係に基づく楽曲検索”, 人工知能誌, 2012, Vol.27, No.3, pp.163-175
- [9] 上原美咲 他, “楽曲群のコード進行・メタ情報・楽曲特徴量の統合可視化の一手法”, 情報処理学会全国大会講演論文集, 2015, Vol.77, No.2, pp.395-396.
- [10] 松橋聡 他, “顔領域抽出に有効な HSV 表色系の提案”, テレビ誌, 1995, Vol.49, No.6, pp.787-797.
- [11] 西川直毅 他, “歌詞と音響特徴量を用いた楽曲印象推定法的设计と評価”, 情処学研報, 2011, Vol.2011-MUS-91, No.7, pp1-8.
- [12] 斉藤優理 他, “MusiCube: 特徴量空間における対話型進化計算を用いた楽曲提示インターフェース”, 可視化情報学会論文集, 2014, Vol.34, No.9, pp.17-27.
- [13] 草間かおり 他, “楽曲データの印象表現に基づいた一覧手法の一手法”, 情処学研報, 2009 Vol.2009-MUS-81, No.19, pp85-86.