

# 共感覚の可視化アート

井上泰助† 蔡東生†  
福本麻子‡ 棚沢順††

本論文では音符と絵画の色を感性語を介さない直接の対応(ノンバーバルマッピング)を試みる。このマッピングは音楽と絵画に内在する Zipf の法則によって決定される。具体的には音符と色のヒストグラムをとり、それぞれの頻度順に対応付けることによって、音と色を対応付ける。さらに、音符のヒストグラムを共感覚を用いて色付けし、2つのヒストグラムを Earth Mover's Distance を用いて比較し、最も近い音楽と絵画を探す方法を提案する。

## Visualization based on Synesthesia

Taisuke Inoue† Dongsheng Cai†  
Asako Fukumoto†† Jun Kurumisawa†††

We try to perform a nonverbal mapping between musical note and colours in paintings. The mapping order has been determined by Zipf's Law. We obtain the histogram of the musical notes and paintings color and map the ranking between these two information. In order to map these two histogram we use Earth Mover's Distance to search the most similar pairs of music and painting.

### 1. はじめに \*

テレビのコマーシャル、映画などで見られるように、音楽と映像がマッチしたものであれば、お互いの印象を強めることができる。このような音楽と映像がどのようなパラメータで結びついているか解明することができれば、ゲーム、CM、アートなどのマルチメディア作品の創作支援に応用できると考えられる。現在メディア間の関係を表すには感性語を介して表現することが大半であるが、この方法には限界がある。

そこで本論文では音楽から音の分布を取り、共感覚を用いて音と色を対応付けて、Earth Mover's Distance(EMD)[1]という尺度を用いて画像の色の分布と比較し、ある音楽に対してマッチした画像を選ぶ方法を提案する。

### 2. 音とイメージの融合

音楽とイメージの融合というと、ウォルト・ディズニーの作品「ファンタジア」が音楽映画として色彩と音楽を融合させた最高傑作として知られている。しかし、この作品において音楽と色彩の融合は感性によっておこなわれており、科学的ではない。その他、現在の音楽とイメージの融合はウィンドウズメディアプレイヤー、Martin Wattenberg の「Shape of Song」、永野哲久氏の「Monalisa: 音の影」などのような(1)音楽の視覚化(画像の音化)、または Music ビデオのような(2)

映像との組み合わせ、また申請者が開発をした(3)Ziplayer[2]、などを挙げられる。

#### 2.1 従来手法

##### (1)音楽の視覚化

音楽の視覚化はオシロスコープに色とプリミティブな形を与えたものが主流である(図1)。円、グリッドなどプリミティブな形で描くと現代的なデザインが施されてかっこよく、さらにコンピュータで計算しやすく、プロジェクターで表示しやすいという利点がある。しかしながら、そこには工夫や広がりをもたせる余地がなく、絵画がもつ複雑な表現能力は見えてこないため、音に対応した視覚化であっても、それ自体に心を動かされたり、鑑賞したりといった体験はうまれてこない。



図1. ビートルズの Eleanor Rigby を Shape of Song に読ませたイラスト

\* 「Shape of Song」: あらかじめ用意された音楽もしくは MIDI ファイルをアップロードすると、曲を解釈して、様々な形の半透明の円を描いてくれというソフトウェア

<http://www.turbulence.org/Works/song/method/method.html>

\*† 筑波大学  
Tsukuba University  
‡ 慶応大学  
Keio University  
\*†† 千葉商科大学  
Chiba University of Commerce



図2 Ziplayer

## (2)映像と音楽の組み合わせ

ミュージックビデオやカラオケなどのように具体的な映像に音楽を組み合わせることがある。しかしながら、音楽と絵画を融合するメディアとして、具体的な映像と音楽を使う場合を考えると、以下のような問題点が挙げられる。

- 1.イメージを合わせることはできるが、音と画像(絵画)のパラメータを対応させにくい。
- 2.コンサートで映像を使うと、観客の注意が聴覚より視覚の方に奪われてしまい、音楽を鑑賞する障害になるため、演奏者・作曲家に敬遠される傾向がある。
- 3.抽象性が失われるため、演奏者(描画者)にも鑑賞者にも広がりがなく、視覚的に固定された、予測可能なものにしかならない。

## 2.2 Ziplayer[2]

Ziplayer は統計的手法を用いて絵画と音楽を融合するメディアプレイヤーである(図2)。Ziplayer では zipf の法則に基づいて音と色を対応付け、演奏する。音楽について zipf の法則は、ドビュッシーやモーツァルトなどの調和の取れた音楽はべき乗則に従っており、シェーンベルグなどの無調音楽は従わないことが Zanette らによって報告されている[3]。また、申請者らによって印象派の画家がパレット上で使った色とその使用量が zipf の法則に従っていることが確認されている[4]。

Ziplayer では次のような手順で音と色を対応付け、演奏している。

### (1) 絵画の解析

・1-1. 絵画の画像ファイルからパレットカラーを抽出する。

・1-2. 絵画をパレットカラー毎にレイヤーへ分割する。

### (2) 音楽の解析

・2-1. 音楽ファイル(MIDI形式)から NoteOn/NoteOff イベントを抽出して音程毎の出現回数を取得し、各音程を出現回数の多い方から順に並び替える。

### (3) 音程と色彩の割り当て

・3-1. 音程とパレットカラーを、それぞれ出現回数の多い方から順に割り当てる。

## (4) 演奏

・4-1. 音楽の進行に合わせて、各音符に対応したパレットカラーを絵画上に表示する。

ziplayer では従来の音楽とイメージの融合と異なり、画家の使った独特の色使いで音楽を聞くことができる

ziplayer では統計的手法のみを用いて音と色を対応付けてきたが、本論文では共感覚を用いた認知的手法も用いて音と色を対応付ける。音楽だけを聴くより、共感覚現象がみられると、記憶力が向上されたり印象が強まる傾向がある。脳科学の分野で、色や音、テキストなどの刺激が脳のどの領域を活性化するかが明らかになっており、1つの刺激を与えるより多くの領域を活性化させるため、理解や記憶の反応が強まるといわれている。これによって、作者が音楽の強調したいパラメータを選び画像と結びつけることによって、作者の意図を聞き手に伝えることができるのではないかと考えられる。

## 3. 認知的手法による音と色の対応

ここで絵画の要素(パラメータ)と音楽の要素を整理する。

### ●絵画の要素

(1)色 (2)マチエール(+ハイライト) (3)構図

### ●音楽の要素

(1)キー(調) (2)ピッチ(音高) (3)音色 (4)強弱

メディアを構成するパラメータとイメージの対応関係を、感性語を解さず表現する対応付けをノンバーバルマッピングという。本プロジェクトでは統計的手法と認知的手法の2方向よりマッピングを行なうことを目的としている。

認知的手法では、共感覚の中で色聴現象に着目し、音楽における調、音、音色のパラメータを変化させた際に色聴保持者が感じる色の対応付けの規則性を、明度、色相、彩度のパラメータを用いて表現することを目的とする。

共感覚とは、ある刺激に対して通常の感覚だけでなく異なる種類の感覚をも生じさせる一部のみにみられ

る特殊な知覚現象をいう。例えば、共感覚を持つ人には7という文字に青い色を感じたり、音階のミの音に緑色を感じたり、ハンバーグの形が苦い味に感じたりする。特に共感覚の中でも音楽や音に対して色を感じるものは色聴といわれる。

共感覚は古くはギリシャ時代からその存在が語られてきた。ニュートンは音と色の関係を報告しており、ファインマンは数式を色で語っている。しかし、人間の感覚の問題であったので、その存在は”思いこみ”であるとか、個人の”錯覚”、”あいまいな”感覚”と思われてきた。しかし、1999年以降、fMRI、PETでそのメカニズムが明らかになるにつれその存在は否定できないものになっている。

色聴保持者の中にはバラツキがあるものの、とくに絶対音感をもつ人々の間では、ある程度同じようなマッピング（色と音の対応づけ）がみられることが報告されている[5]。しかし、色聴によるマッピングは非常に複雑であり、まだ音と色の関係ははっきりしていない。そこで今回は音程と言う概念のみを用いてマッピングする。特に今回は実際の音でなく、ある音に対応する音符に対して色を感じるという共感覚を用いてマッピングを試みた。インターネット上にある音符に対する色の対応の報告を図3に示す。

音符と色の対応は HSV 色空間の H(色相)で表すことができ、これを用いてド～シの音に色を割りあてる。しかし、#やbに対応する色の報告はなされていないので、それぞれ隣の2つの音の色の HSV 色空間での中間色を割り当てることにした(図4)。

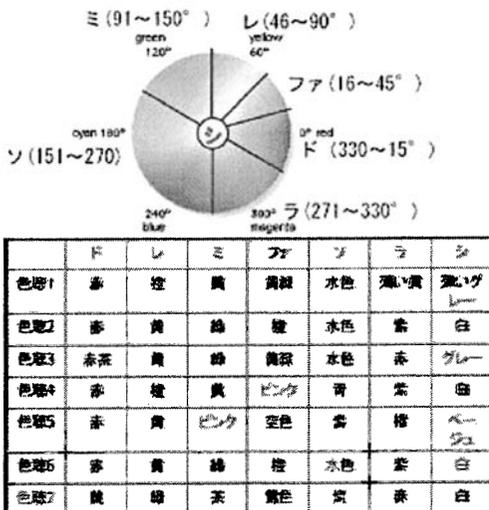


図3 色聴保持者の音符と色のマッピング

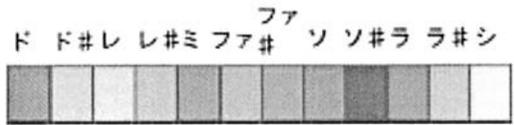


図4 音と色の対応

#### 4. 提案手法

認知的手法を用いてある画像に合った音楽を選ぶ方法を提案する。具体的には音楽から音程の分布をとり、各音程に共感覚に基づいて色を割り当てる。また、画像から同様に色の分布をとり、Earth Mover's Distanceでこれらの分布を比較する。

1. 音楽を解析し、音程の分布をとる。具体的には midi 形式の音楽ファイルから NoteOn/Off イベントを抽出し、音程ごとの出現回数を取得し、図3の色を割り当てる。音高は考慮しない。
2. 画像を解析し、色の分布をとる。今回は k-means 法でクラスタリングを行い、音程の数に近い16色に分色した。
3. 分布を正規化する。これは音楽の分布と画像の分布のスケールの差が大きすぎるためである。

4. 正規化された2つの分布を Earth Mover's Distance(EMD) [1]で比較する。EMDは分布間の距離を表す尺度であり、類似画像検索の分野でよく用いられている。EMDは、分布間の距離を輸送問題として捉え、最適な輸送コストを用いて定義される。EMDを求める際、2つの分布は signature として表現されると、

$P = \{(p_1, w_{p_1}), \dots, (p_m, w_{p_m})\}$  となる。今、分布  $P$  は  $m$  個の特徴量で表現されており、 $p_i$  は特徴ベクトル、 $w_{p_i}$  はその特徴量に対する重みである。同様にもう一方の分布は  $Q = \{(q_1, w_{q_1}), \dots, (q_n, w_{q_n})\}$  となる。ここで、 $p_i$  と  $q_j$  の距離を  $d_{ij}$  とし、 $f_{ij}$  を  $p_i$  から  $q_j$  へのフローとすると、評価関数

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij} \quad (1)$$

が次の制約に従って最小になるような  $F = \{f_{ij}\}$  を求める。

$$f_{ij} \geq 0 \quad (1 \leq i \leq m, 1 \leq j \leq n) \quad (2)$$

$$\sum_{j=1}^n f_{ij} \leq w_{p_i} \quad (3)$$

$$\sum_{i=1}^m f_{ij} \leq w_{q_j} \quad (4)$$

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min \left( \sum_{i=1}^m w_{pi}, \sum_{j=1}^n w_{qj} \right) \quad (5)$$

このとき、P、Q間のEMDは以下の式で与えられる。

$$\text{EMD}(P, Q) = \frac{\sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij}}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}} \quad (6)$$

従来のヒストグラムによる色分布表現では、ヒストグラムのビンが固定であるため表現力が不十分で、またヒストグラム間の比較尺度が人間の感覚に合わないことがあるといった欠点がある。Signatureの場合は対象に応じて要素数、ビンの代表値を変えることができ、柔軟な表現が可能である。また、EMDは他の尺度より人間の感覚に近いことが報告されている。

## 5. 実験

提案手法を用いてゴッホの「The Starry Night」に合った音楽を探す実験をおこなった。Signatureで用いる特徴ベクトルはCIE-L\*a\*b色空間での色ベクトル、分布間の距離はCIE-L\*a\*b色空間でのユークリッド距離とした。比較に用いた音楽と「The Starry Night」とのEMDを表1に示す。音楽は母集団として印象主義音楽の中から選んだ。

表1. 比較に用いた音楽とEMD値

作曲家	曲名	EMD
ドビュッシー	亜麻色の髪の乙女	67.35279
	夢	56.57677
	子供の領分より「ゴリウオーグのケイクウォーク」	51.55793
	アラバスク1番	53.99955
	月の光	52.82109
スコット	アニーローリー	57.5159
ホルスト	木星	52.35121
ラベル	亡き王女のためのパヴァーヌ	48.7227
ファリャ	スペイン舞曲 第1番	58.21979

本手法ではゴッホの「The Starry Nigh」に一番合った音楽としてラベルの「亡き王女のためのパヴァーヌ」が選ばれた。これらを組み合わせてZiplayerで演奏すると、より印象の強い作品が出来上がるのではないかと考えられる。

## 6. まとめ

本論文では共感覚を用いた認知的手法により音と色の対応付けを行い、あるイメージに対してもっともマッチした音楽を選ぶ手法を提案した。Ziplayerにおいて従来の統計的手法によるマッピングだけでなく、この認知的手法を組み合わせることを提案した。今後の課題として、今回の結果の評価、音程以外に調や音色な

どによるマッピングの提案、共感覚の詳細な調査などが挙げられる。

## 参考文献

- 1) Y. Rubner, C. Tomasi and L. Guibas., "The earth mover's distance as a metric for image retrieval," Int. J. Comput. Vision, vol.40, pp.99-121, 2000.
- 2) 福本麻子, 塚田浩二, 蔡東生, 安村通晃 "Ziplayer: 絵画, 文章, 音楽を組み合わせたメディアプレイヤーの提案" インタラクシオン2006論文集, pp.79-80, 2006
- 3) Damián H. Zanette. "Zipf's law and the creation of musical context" Musicae Scientiae, 10(1), pp.3-18, 2006
- 4) 福本麻子, 安村通晃, 蔡東生 "印象派絵画の統計的解析" ヒューマンインターフェース学会 論文誌 vol17 No.2, pp.89-97, 2005
- 5) 長田典子, 岩井大輔, 津田学, 和氣早苗, 井口征士. 音と色のノンバーバルマッピング—色聴保持者のマッピングルール抽出とその応用—, 電子情報通信学会論文誌 A, J86-A(11), pp.1219-1230 2003