

# 印象の近い音楽と画像の組み合わせを自動選出する手法

小田瑞穂 伊藤貴之

お茶の水女子大学 理学部情報科学科

## 1. 概要

MIDI や MP3 などのデジタル音楽記録形式の普及により、近年では計算機上での音楽鑑賞が非常に普及している。また一方では、聴覚だけでなく視覚に訴える計算機技術も非常に発達している。そこで聴覚だけでなく視覚にも同時に刺激を与えるような音楽提供方法により、さらに直感的に音楽に親しむことが可能になると考えられる。

そこで本研究では、音楽情報を与えると、コンピュータがその音楽の印象を推測しそれに近い印象の画像を選択し、同時に提示する手法を提案する。本研究では画像や音楽の特徴を自動抽出して数値化し、その特徴から画像や音楽の印象を推測する。この推測のために本研究では、人間による印象も数値化し、特徴から印象を求める数式を用いる。この数式の決定には応答曲面法を用いる。

この研究成果は例えば、曲ごとにアイコンを作り、多くの曲を一覧表示する、という目的に応用可能であると考えられる。また、本報告で提案する手法の原理は、音楽から画像を選ぶだけでなく、画像から音楽を選ぶ、という逆の目的にも応用が可能である。

## 2. 関連研究

大山らは画像に印象の合うように音楽を自動アレンジする手法を提案している[1]。また印象に基づいて楽曲を検索するシステムについての研究がある[2]。しかし、この研究では画像と音楽の組み合わせは実現していない。また、類似度の高い曲に類似度の高い色を割り当てるインタフェースの研究[3]も存在している。

## 3. 提案内容

本研究では、「特徴値」「評価値」という2種類の数値をもって、画像および音楽を評価する。ここで特徴値および評価値の定義は以下のとおりである。

**[特徴値:]** 画像および音楽の特徴を表す値で、コンピュータによる自動抽出が可能なもの。

**[評価値:]** 画像または音楽に対して、人間による主観的な印象を表す値。本研究では、画像や音楽の印象を表

す数種類の形容詞を用意し、その形容詞の各々に対して適用度を主観評価させた値を評価値として用いる。例えば「明るい」という形容詞に対し、与えられた画像や音楽の適用度を0から10までの11段階のいずれかの値で主観評価した値を用いる。よって本研究では、用意した形容詞の数だけの評価値が用いられる。なお本研究では、画像と音楽の両者に対して同じ形容詞を用いるものとする。

ここで特徴値と評価値を、以下のように定式化する。

画像の項目の特徴値を  $P(p_1, p_2, \dots, p_l)$  とする。

音楽の項目の特徴値を  $M(m_1, m_2, \dots, m_m)$  とする。

両者共通の評価値を  $A(a_1, a_2, \dots, a_n)$  とする。

(ただし  $l, m, n$  は正整数)

$P$  から  $a_k$  を求める数式を  $a_k = f_k(P)$  とする。

$M$  から  $a_k$  を求める数式を  $a_k = g_k(M)$  とする。

関数  $f$  と  $g$  は応答曲面法を使って求める。応答曲面法とは、 $a_k$  と  $x_i$  を与え、その点群の近傍を通るような曲面を生成できるように係数  $0, i, i_j$  を決定する方法である。本報告では関数  $a_k$  を、応答曲面法による以下の数式により表現する。

$$a_k = 0 + i x_i + i i x_i^2 + i_j x_i x_j$$

$a_k$ : 人間による画像や音楽の印象の、 $k$  番目の評価値

$x_i$ : コンピュータが計算する画像( $P$ )や音楽( $M$ )の特徴

本研究では最初に準備段階として、 $s$  個のサンプル音楽と  $t$  個のサンプル画像を用意する。これらを評価者に提示し、評価者がもつ印象を数値化することで、評価値  $a_1 \sim a_k$  を決定する。つまりここには  $(s+t)$  個の評価値が存在することになる。

続いて  $s$  曲のサンプル音楽について、特徴値  $M(m_1 \sim m_m)$  を求める。この結果を以下の式

$$a_1 = g_1(M)$$

:

$$a_k = g_k(M)$$

に代入し、 $g_1 \sim g_k$  の各々に対して応答曲面法を適用し、係数  $0, i, i_j$  を確定する。

同様に、 $t$  枚のサンプル画像について、特徴値  $P(p_1 \sim p_m)$  を求める。この結果を以下の式

$$a_1 = f_1(P)$$

:

$$a_k = f_k(P)$$

に代入し、 $f_1 \sim f_k$  の各々に対して応答曲面法を適用し、

---

A method for automatically coupling images and music those bring similar impressions.

Mizuho Oda, Takayuki Itoh

Ochanomizu University

{miz-oda, itot}@itolab.is.ocha.ac.jp

係数  $a_i$ ,  $i_j$  を確定する。

以上により  $a_1 \sim a_k$  を求める数式が完成する。これ以降、新しい音楽データや画像データを追加しても、 $a_1 \sim a_k$  を求める数式を用いることで、計算機は画像や音楽の印象を推測できる。

表 1. 特徴値と評価値の一覧

画像 P( $p_1, \dots$ )	評価値 A( $a_1, \dots$ )	音楽 M( $m_1, \dots$ )
色相	明暗	休符含有率
彩度	静動	平均音程
明度	重軽	テンポ
面積	攻防	単位時間当音数
要素数	華素 幸悲 硬柔 複単	

以下に具体的な特徴値と評価値を挙げる。特徴値は音楽や画像から数値として取り出しやすくするため、音楽は MIDI を、画像は SVG を使い、そこから数値として取り出しやすいものを特徴値として抜粋した。評価値は SD 法 (Semantic differential method) [4] が提示する形容詞群の中からいくつかを抜粋し、これについて評価者に採点させた結果を用いている。本研究で用いた特徴値と評価値の一覧を、表 1 に示す。

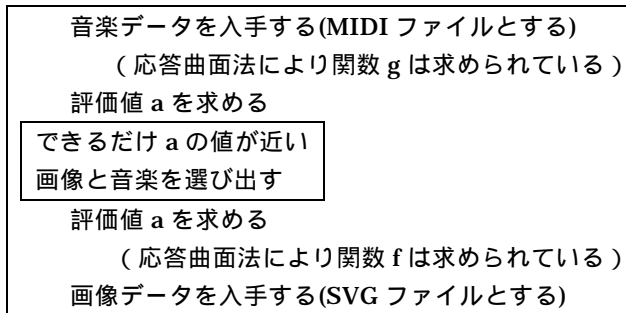


図 1. 印象の近い画像と音楽の選出

本研究では、評価値をベクトルであるとみなし、その距離が小さいものを「印象が近いもの」と判断する。例えば A の項目が 3 個のとき、画像の評価値を  $A_1(a_{11}, a_{12}, a_{13})$  とし、音楽の評価値を  $A_2(a_{21}, a_{22}, a_{23})$  と仮定する。3 次元座標系で  $A_1$  と  $A_2$  を座標にプロットするし、印象の近いものを選ぶには 2 点の距離が最小である点を探し出せばよいことになる。2 点間距離は  $(a_{11}-a_{21})^2 + (a_{12}-a_{22})^2 + (a_{13}-a_{23})^2$  によって求められるので、距離が最小の画像を選び出せばよい。A の項目を 3 項目から  $n$  項目に拡張したときには、 $\sum_{k=1}^n (a_{1k}-a_{2k})^2$  の値が最小になる組み合わせを与える画像を選べばよい。以上の処理をまとめると図 1 のようになる。

#### 4. 実行結果

まず準備段階として、サンプル画像とサンプル音楽に対して特徴値と評価値を求め、この値から散布図を生成した。この散布図から、評価値に対応する各々の印象語について、相関性の強い特徴値を画像と音楽に対して視覚的に選択した。この結果を表 2 に示す。

この結果を用いて、各々の評価値ごとに応答曲面法を適用して数式をつくり、評価値の距離の近い画像と音楽を選択する。現時点では、画像、音楽の両方の特徴値を 2 つ以上適用可能な、上から 4 項目だけについて応答曲面法を適用した。選択結果とその評価については講演にて論じる。

表 2. 相関性の強い特徴値の視覚的な選出結果

画像の特徴値	評価値	音楽の特徴値
彩度, 明度, 面積	静的な 動的な	音程, 音数
彩度, 面積, 数	重い 軽い	テンポ, 音程
彩度, 面積	攻撃的 保守的	音程, 音数
彩度, 面積, 数	華やか 素朴な	テンポ, 音程, 音数
彩度	硬い 柔らか	休符含有率
彩度, 明度	明るい 暗い	テンポ

#### 5. まとめ

本研究では、印象の近い音楽と画像を組み合わせる手法を提案した。今後の課題として、表 2 の下 2 項目についても応答曲面法を適用できるようにしたい。また、音楽については、数値以外の特徴 (例えば音色など) や、1 曲の中での曲想の変化などを考慮できるようにしたい。画像については、静止画だけでなく CG アニメーションなども対象にしたい。

#### 謝辞

応答曲面法は京都大学小山田耕二教授からいただきました。この場をお借りして御礼申し上げます。

#### 参考文献

- [1] 大山、伊藤、DIVA:画像の印象に合わせた音楽自動アレンジの一手法の提案、第 68 回情報処理学会全国大会講演予定、2006。
- [2] 熊本、太田 印象に基づく楽曲検索システムにおける程度語の理解 The 18<sup>th</sup> Annual Conference of the Japanese Society for Artificial Intelligence, 2004
- [3] 後藤、後藤、Musicream:楽曲を流してくっつけて並べることでできる新たな音楽再生インタフェース、WISS2004, pp.53-58, 2004.
- [4] Osdoog, C.E., Suci, G.J. Tannenbaum, P.; The measurement of meaning, Univ. Illinois Press, (1957)