

# 任意の言葉による楽曲検索のための楽曲特徴量の自動選択

Automatic selection of music features for music retrieval system by arbitrary words

頭川 愛†  
Ai Zukawa

酒向 慎司†  
Shinji Sako

北村 正†  
Tadashi Kitamura

## 1. はじめに

近年の楽曲検索において、キーワードを用いた検索以外にも多様な手法が提案されている。その一例として感性検索が挙げられ、ユーザが求める楽曲の印象を入力とすることから、具体的な曲名などがわからないときにも有効であり、未知の音楽を発見できる効能も期待できる。

このような感性に基づいた楽曲検索の研究はこれまでもいくつか提案されているが、一定の感性語から選択する方法など、検索システムの入力方法が限定的なものが多かった。しかし、音楽を言葉で言い表す際には様々な語が用いられ、複数の語や文章のような形で表現されることも多いことから、音楽を形容するための語は種類や数に制約のない形が望ましいと考えた。

楽曲と言葉の印象を表す指標として、各軸を代表的な感性語とした印象空間を考える。印象モデルはこれまでもいくつか提案されてきているが、少人数の評価データを用いているなど確立されたものは存在しない。そのため、聴取実験によるデータを使用して本研究に適した印象空間を生成した。得られた空間に、経験的に決めた8つの楽曲特徴量を用いて楽曲を、言葉の類似度を用いてあらゆる言葉を写像した [1]。しかしながらその際用いていた楽曲特徴量は、すべて印象と関係しているとは言えず、楽曲が適切に写像できていない可能性があった。

そのため、本研究では有用な楽曲特徴量だけを選択して用いる手法を提案する。印象空間の各軸によって関係している特徴量は異なると考えられるため、それぞれについて変数選択を行い、楽曲が正しく写像されるかを評価実験によって検証した。

## 2. 任意の言葉の写像

任意の言葉の印象を個々に扱うのは困難であるため、印象が既知である言葉との類似度から全ての言葉の印象を表現することを考える。そこで印象を表す指標として印象空間を設定し、その空間に写像することで印象を表す。印象が既知である、すなわち空間上の座標が既知である言葉を決めることができれば、それらとの類似度から入力語の座標を決定することができる。空間上の座標が既知である言葉が多いほど様々な言葉が写像することができるが、多数の言葉の関係を表すのは難しいため、印象空間を生成する際には代表的な印象を表す言葉のみを用いることにする。

### 2.1 印象空間の生成

先行研究 [2] で実施した聴取実験の結果を利用する。RWC 研究用音楽データベース [3] のクラシック音楽を評価対象とし、歌声が含まれていない44曲から与える印象が一定と考えられる15秒間を1サンプルとして100サンプル使用している。被験者119名の男女に全てのサンプルを聴かせ、それぞれに対して14対の感性語についてSD法の7段階で印象評価をさせた。

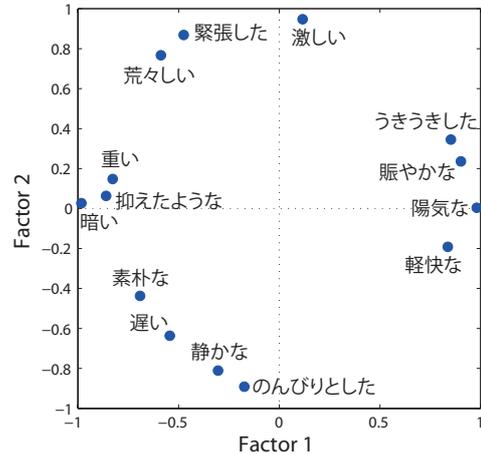


図 1: 印象空間  $S$

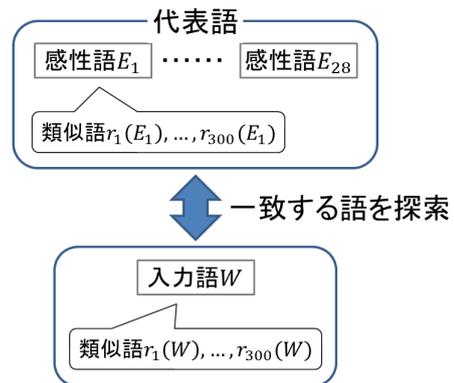


図 2: 任意の言葉の写像方法

聴取実験で得られた評価平均を因子分析し、印象空間  $S$  を生成する。正当な評価がされていないデータを除き、更に一般的な印象空間にするため平均との差が標準偏差以上のものを省いた。因子分析の結果得られた印象空間  $S$  を図 1 に示す。感性語対の対になる言葉は、原点を中心とした点対称の位置に配置する。第 1 因子軸は「明るい」、「陽気な」などが大きい値を示していることから「明るさ」、第 2 因子軸は「激しい」、「慌しい」などが大きい値を示していることから「激しさ」を表していると言える。

### 2.2 代表語の配置

高々28語の感性語との類似度だけでは任意の言葉の写像は困難であると考えられる。そこで、感性語と類似した言葉も印象空間  $S$  上に配置する。類似度として、意味が類似している言葉同士は同じ文書内で使われることが多いと仮定し、ALAGIN [4] の単語間の共起確率を用いる。2つの言葉  $w_1, w_2$  の共起確率  $P(w_1, w_2)$  は、 $F(w_1)$  を  $w_1$  の出現頻度、 $F(w_2)$  を  $w_2$  の出現頻度、 $F(w_1, w_2)$

†名古屋工業大学大学院工学研究科, Graduate School of Engineering, Nagoya Institute of Technology

表 1: 楽曲特徴量

従来	音量
	テンポ
	major と minor の音量差
	スペクトルの変化度合
追加	弱音の割合
	主音の明瞭度
	調性の重心の変化度合
	中心周波数
	85%を占める低音域の割合
	不協和音の多さ
	MFCC13 次元
ゼロ交差数	

を  $w_1, w_2$  の共起頻度としたとき、式 (1) で定義される。

$$P(w_1, w_2) = \frac{F(w_1, w_2)}{F(w_1) + F(w_2)} \quad (1)$$

言語の特性から、形容詞や形容動詞よりも名詞のほうが類似語が共起しやすいと考えられるため、「軽い」を「軽さ」など全ての感性語を名詞に変換する。感性語(名詞)と共起確率が高い上位 300 語を原点と感性語を結んだ直線上に配置する。これらを代表語と呼ぶ。代表語  $R$  の座標  $(x_{1R}, x_{2R})$  は、感性語  $E$  の座標を  $(x_{1E}, x_{2E})$  とすると、

$$(x_{1R}, x_{2R}) = (P(E, r(E)) \cdot x_{1E}, P(E, r(E)) \cdot x_{2E}) \quad (2)$$

となる。代表語を用いて任意の言葉の印象空間  $S$  上の座標を決定する。

### 2.3 入力語の写像

入力語  $W$  及び入力語の類似語 300 語のうち、代表語と  $R_1, R_2, \dots, R_n$  の  $n$  個が代表語と一致したとき、 $W$  の座標  $(x_{1W}, x_{2W})$  は以下のように算出する。

$$\left( \frac{\sum_{j=1}^n P(r(W), R_j) \cdot x_{1R_j}}{\sum_{j=1}^n P(r(W), R_j)}, \frac{\sum_{j=1}^n P(r(W), R_j) \cdot x_{2R_j}}{\sum_{j=1}^n P(r(W), R_j)} \right)$$

## 3. 楽曲の写像

聴取実験の結果を利用して、楽曲特徴量から印象空間  $S$  上の座標を決める写像式を作る。そのために目的変数を評価値、説明変数を特徴量とした重回帰分析を用いる。

### 3.1 楽曲特徴量

従来では、経験的に表 1 の 4 つの特徴量を用いていた。しかし、これら全てが印象と関係していないという可能性があった。また、他にも印象に関係している特徴量があると考えられるため、新たに 8 種類の特徴量を追加しこれらの中から有用なもののみを選択して使用する。

### 3.2 楽曲特徴量の選択

表 1 の特徴量の中から有用なものだけを選択するため、変数増加法を用いる。説明変数は短時間に区切ったフレームごとに求めたそれぞれの特徴量の平均と分散とし、目的変数は各軸を表す指標として第 1 軸では「明るい-暗い」、第 2 軸では「激しい-穏やかな」の評価平均を用いる。変数増加法で有用な変数から追加していき、分散比が 2 未満になるまでの変数を使用する。第 1 因子では 16 個、第 2 軸では 17 個の特徴量が選択された。

表 2: 順位ごとの評価平均

	1 位	2 位	3 位	平均
従来手法	3.60	3.43	3.82	3.62
提案手法	4.03	3.74	3.79	3.85

### 3.3 検索方法

印象空間  $S$  において、入力語と全ての楽曲のユークリッド距離を求め、距離の近い楽曲ほど印象が似ていると判断し、検索結果として出力する。

## 4. 実験

楽曲特徴量を経験的に決めた従来手法と自動選択した提案手法の性能を比較する。

### 4.1 客観評価実験

感性語を入力語としたときにどのような楽曲が検索されるかを調査する。聴取実験で得られた評価平均と最高値との差が 2 未満である楽曲を正解楽曲とすると、検索結果の上位 3 曲が全て正解となった入力語は、28 語中 17 語であった。従来手法で検索を行ったときは 11 語であったため、より正確に楽曲が写像されたことがわかる。

### 4.2 主観評価実験

新しい楽曲の写像式の有効性を検証するため、主観評価実験を行った。感性語ではない言葉 20 語を入力語とし、従来手法で検索された上位 3 曲と、提案手法で検索された上位 3 曲について印象の合致度を 5 段階 (1:合っていない, 5:合っている) で評価させた。被験者は 8 名である。

全体の評価平均を表 2 に示す。提案手法の 1 位の評価平均が唯一 4 を超え、最も高い評価が得られた。従来手法においては 3 位の評価平均が高くなったが、これは従来手法の 3 位と提案手法の 1 位が同じ楽曲になることがあるからだと考えられる。この結果から、従来の特徴量を用いるより選択された特徴量を用いたほうがよいということがわかった。

## 5. むすび

本研究では、印象空間を用いた任意の言葉を用いる楽曲検索の実現に向け、楽曲の写像方法の改善を行った。写像に用いる特徴量を従来までは経験的に決めていたが、多くの特徴量から有用なものだけを自動選択する方法を導入した結果、評価実験で性能が向上したことがわかった。今後の展望として、入力を複数の言葉や文章に対応させることを考えている。

## 参考文献

- [1] 頭川愛 他, 任意の言葉による楽曲感性検索のための感性語による印象空間生成, 情報処理学会第 75 回全国大会, pp.255-256, 2013.
- [2] 岩月靖典 他, 利用者のプロフィールを用いた個人性を考慮した楽曲の印象推定, HCG シンポジウム 2012.
- [3] RWC 研究用音楽データベース, <http://staff.aist.go.jp/m.goto/RWC-MDB/index-j.html>
- [4] ALAGIN 言語資源・音声資源サイト, <http://alaginrc.nict.go.jp/>