

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

頭川 愛<sup>†</sup>, 酒向 慎司<sup>†</sup>, 北村 正<sup>†</sup>

<sup>†</sup> 名古屋工業大学大学院工学研究科

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なされているが、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。従来まで、Hevner が提案した 8 つの印象語群を用いて印象空間を生成していた。しかし、主観評価実験の結果、よい精度が得られなかった [1]。その原因として、第一に Hevner の研究が 70 年以上前に行われており、現代の意味とずれている可能性が考えられる。ソーシャルタグを用いた近年の研究 [2] においても、Hevner の研究とは異なる結果が示されている。第二に、英語を和訳する際に意味が多少変わる言葉が存在することが挙げられる。したがって、最近の日本語による印象空間が望まれるが、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。

本研究では、聴取実験によるデータを用いて多くの人に対応する一般的な印象空間を生成する。楽曲の印象を表すものにふさわしい指標を定義し、楽曲特徴量との関係性を多変量解析によって求める。そして、評価実験により楽曲と言葉の対応が的確であるかを検証した。

2 実データによる印象空間の生成

2.1 聴取実験

実データとして、先行研究 [3] で実施した聴取実験を利用する。この実験では、評価対象として RWC 研究用音楽データベース [4] のクラシック音楽を使用している。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 サンプル用意している。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 (表 1) について SD 法で印象評価をさせた。表 1 の左側の感性語を 1、右側の感性語を 7 とした 7 段階評価である。

2.2 印象空間の生成

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。一般的な印象空間を生成するために平均より大きく外れた評価データを除く。因子分析をして得られた印象空間上で、それぞれの感性語の座標ができるだけ離れていることが望ましいと考え、それぞれの感性語の最も近くにある感性語との距離を調べた。評価平均との差  $d$  を変化させて調べた二乗距離平均の推移を図 1 に示す。寄与率から印象空間は 2 次元と決定した。この結果より、全 166600 データのうち  $d < 2$  の評価である 154702 データを使用した。

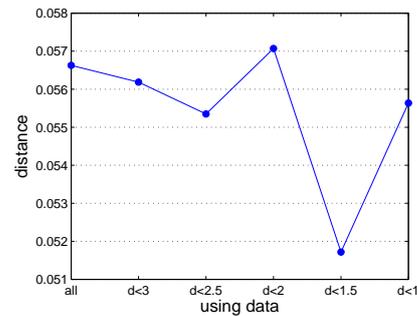


図 1: 最も近くにある感性語との二乗距離平均

表 1 の右側にある言葉の印象空間  $S$  上の座標を図 2 に示す。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 因子は「陽気な」「明るい」などが大きい値を示していることから「明るさ」を意味しており、第 2 因子は「慌しい」「激しい」などが大きい値を示していることから「激しさ」を意味していると考えられる。

表 1: 感性語対

番号	感性語対	
1	軽い	重い
2	明るい	暗い
3	しみりした	うきうきした
4	迫力のある	静かな
5	穏やかな	激しい
6	陰気な	陽気な
7	のびやかな	抑えたような
8	速い	遅い
9	優雅な	荒々しい
10	寂しい	賑やかな
11	慌しい	のんびりとした
12	重厚な	軽快な
13	安らぐ	緊張した
14	華やかな	素朴な

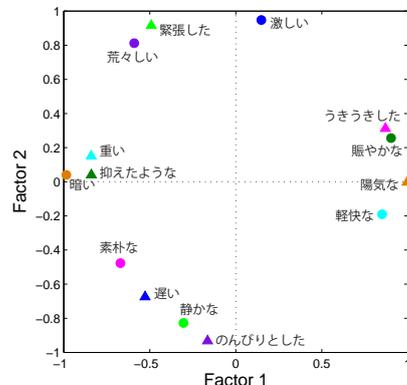


図 2: 印象空間  $S$

### 3 楽曲の写像

楽曲を印象空間  $S$  に写像するため、聴取実験で用いた楽曲から抽出される特徴量と第1因子、第2因子の関係を明らかにする。

#### 3.1 楽曲特徴量

楽曲特徴量として本研究ではテンポ、音量、音高、長調と短調の音量差、それぞれの平均と標準偏差を用いる。抽出には、MIRToolbox[5]を使用した。

#### 3.2 印象空間と楽曲特徴量の関係

聴取実験で得られた結果を重回帰分析することによって、楽曲特徴量から印象空間  $S$  上の座標を決定する係数を導く。それぞれの軸を表す指標として、因子負荷量が最も高い感性語対から第1因子を「陰気な-陽気な」、第2因子を「慌しい-のんびりとした」とする。これらの評価平均と楽曲特徴量の関係を調べるにあたり、印象空間  $S$  の座標の範囲を  $(-1, 1)$  とするため、評価平均  $e$  を式(1)を用いて  $e'$  に変換する。

$$e' = \frac{e - 4}{3} \quad (1)$$

評価平均  $e'$  と前述の楽曲特徴量を用いて重回帰分析を行う。

### 4 任意の言葉の空間写像

印象空間  $S$  上に配置した感性語を使って、より多数の言葉の座標を決めたい。ここでは、言葉の類似性に着目し、感性語と類似した言葉は原点とその感性語の間に配置できると考える。

#### 4.1 代表語の配置

感性語に類似した言葉を代表語と呼ぶ。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[6]の単語間の共起確率を類似関係を表す尺度として用いる。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、 $F(w_1, w_2)$  を  $w_1, w_2$  の共起頻度としたとき、式(2)で定義される。

$$P(w_1, w_2) = \frac{F(w_1, w_2)}{F(w_1) + F(w_2)} \quad (2)$$

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。そのため、「軽い」を「軽さ」など全ての感性語を名詞に変換して代表語を決定する。感性語(名詞)と共起確率が高い上位300語を、代表語として原点と感性語を結んだ直線上に配置する。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

$$(x_{1R}, x_{2R}) = (k * P(r, E) * x_{1E}, k * P(r, E) * x_{2E}) \quad (3)$$

となる。このとき、 $k$  を  $k * P(R, E)$  が0から1の値を取るよう設定する必要がある。経験的に  $k = 50$  とした。

#### 4.2 入力語の写像及び検索方法

入力語  $W$  に対して、代表語  $R_1, R_2, \dots, R_n$  の  $n$  個と共起するとき、 $W$  の座標  $(x_{1W}, x_{2W})$  は以下のように算出する。

$$\left( \frac{\sum_{j=1}^n P(W, R_j) * x_{1R_j}}{\sum_{j=1}^n P(W, R_j)}, \frac{\sum_{j=1}^n P(W, R_j) * x_{2R_j}}{\sum_{j=1}^n P(W, R_j)} \right) \quad (4)$$

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。

### 5 評価実験

印象空間上で言葉と楽曲が的確に対応しているかを検証するため、評価実験を行う。

### 5.1 実験方法

聴取実験で用いた感性語28語を入力語としたときの検索結果を調べる。評価平均が表1の左側の感性語では2以下、右側の感性語では5以上の楽曲を正解とし、検索結果として上位3曲のうち正解がいくつあるかを評価する。

### 5.2 実験結果

感性語を印象空間上に配置した様子を図3に示す。対になる感性語は同じマークとなっており、図中の番号は表1の番号に対応している。これより、感性語対は原点を中心とした対称に近い値を示しているものが多いことがわかる。全28語の感性語のうち、上位3曲中全てが正解であったものが5語、2曲が正解であったものが7語、1曲が正解であったものが11語、正解がないものが5語であった。平均は1.43曲であるため、出力された楽曲のうち半数が入力語の印象を表現している楽曲であると言える。しかし、正解がない感性語も存在するため、更なる改善が必要だと考えられる。

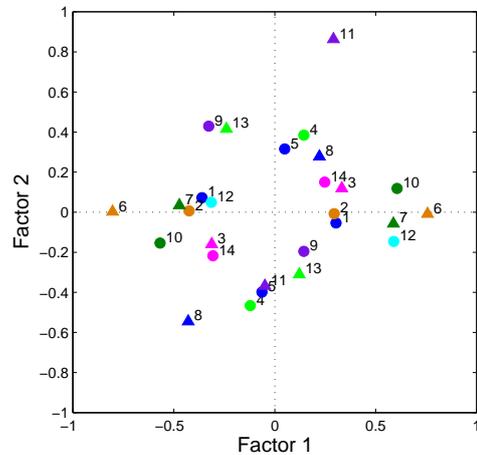


図3: 感性語の写像

### 6 むすび

本研究では、任意の言葉を用いる楽曲検索の実現に向け、楽曲と言葉の印象を表す空間を生成した。聴取実験の結果より、楽曲の印象と関係が深いと考えられる2軸(「明るさ」、「激しさ」)を決定し、それぞれの軸と楽曲特徴量の関係を明らかにした。感性語を入力語とした評価実験を行った結果、検索結果として出力された上位3曲のうち平均で1.43曲が入力語の印象を表した楽曲であることがわかった。

今後の展望として、他の楽曲特徴量の使用、楽曲の写像方法の改善の他に、生成した印象空間を用いての主観評価実験が挙げられる。将来的には、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, 単語共起頻度データベースを使用した任意の言葉の印象に合った楽曲検索, FIT2012, pp.513-514.
- [2] Cyril L. et al., Music Mood Representations From Social Tags, ISMIR 2009, pp.381-386.
- [3] 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.
- [4] RWC 研究用音楽データベース, <http://staff.aist.go.jp/m.goto/RWC-MDB/index-j.html>
- [5] Lartillot O. et al., MIR in Matlab(II): A Toolbox for Musical Feature Extraction from Audio, In Proc. ISMIR, pp.237-244, 2009.
- [6] ALAGIN 言語資源・音声資源サイト, <http://alaginrc.nict.go.jp/>