

# 印象に基づく楽曲検索システムにおける程度語の理解

Interpretation of Modifiers Representing a Degree for Impression-based Music-retrieval

熊本忠彦\*1

Tadahiko Kumamoto

太田公子\*2

Kimiko Ohta

\*1 独立行政法人 情報通信研究機構

National Institute of Information and Communications Technology

\*2 フリー

Impression-based music retrieval is the best way of finding pieces of music that suit the preferences, feelings, or mental states of users. This kind of system presents a relatively short set of words that denotes impressions and users select and estimate one or more words from this set with graphical user interfaces to input their target impressions into the system. With this approach, an increased number of words presented leads to increased costs for the user in terms of the time and labor required to enter impressions. Conversely, a set of words that is too limited will often leave users unable to accurately represent their target impressions. We, therefore, are developing a natural language interface that understands spontaneous inputs of the users' impressions and generates query vectors from them for the impression-based music-retrieval system we developed. In this paper, we focus on a method for generating appropriate query vectors from impression words with modifiers representing a degree.

## 1. まえがき

大量にある楽曲データの中から、ある特定のものを探し出すというよりも、ユーザが持つ何らかの判断基準（例えば嗜好や感性、心的状態など）に合致するものを見つけ出すという場合には、印象に基づく楽曲検索システムが有用である [辻 97, 佐藤 01, 池添 01, 熊本 02b]。

このようなシステムへの印象の入力は、従来、システムが提示する (1) 複数個の印象語の中から 1 つ以上を選択する [辻 97] (2) 複数個の印象語の中から 1 つ以上を選択し、それぞれを多段階評価 (例えば 5 段階評価) する [佐藤 01] (3) 複数個の印象尺度 (印象語の対からなる評価尺度) の中から 1 つ以上を選択し、それぞれを多段階評価する [池添 01, 熊本 02b]、といった方法で行われる。そのため、提示される印象語の数が多いと、入力したい印象を表す適切な印象語 (もしくは印象語の組み合わせ) を決定するのに時間と労力を要してしまい、少ないと、決定できなくなってしまう。そこで、我々は、語彙的にも言い回し的にも自分の言葉で入力できる自然言語インタフェースの実現に取り組んでいる [熊本 03]。

さて、自然言語インタフェースならではの特長の一つとして、程度語の使用が挙げられる。すなわち、単に「面白い曲」や「優しい感じの曲」のような印象語による指定だけでなく、「かなり」や「やや」のような程度語を用いて、より微妙な違いを表現できる点である。そこで、本稿では、程度語によって修飾された印象語から印象に基づく楽曲検索システムへの入力となる検索条件 (印象ベクトルと呼ぶ) を生成するための手法を提案するとともに、被験者 40 名による評価実験を行い、提案手法の有効性を検証する。

## 2. 前提となる楽曲検索環境

### 2.1 印象ベクトル (検索条件) の形式

自然言語インタフェースが実装される楽曲検索システム [熊本 02b] において用いられている検索条件 (すなわち印象ベクトル) の形式を示す。

連絡先: 熊本忠彦, 独立行政法人情報通信研究機構けいはんな情報通信融合研究センター, 〒619-0289 京都府相楽郡精華町光台 3-5, kuma@nict.go.jp

表 1: 我々が提案している印象尺度

番号	印象尺度を構成する印象語
1	静かな — 激しい
2	落ち着いた — 忙しい
3	爽やかな — 重苦しい
4	明るい — 暗い
5	荘厳な — 軽々しい
6	ゆったりとした — 窮屈な
7	綺麗な — 綺麗でない
8	楽しい — 悲しい
9	気持ちが落ち着く — 気持ちが高揚する
10	心が癒される — 心が傷つく

印象ベクトルは、10 次元のベクトルであり、各要素は表 1 に示した 10 個の印象尺度 [熊本 02a] に対応している。すなわち、各要素の意味は対応する印象尺度を構成する印象語の意味によって定義付けられ、その値は印象尺度の評定スケール (7 段階評価値) に準拠する。例えば、楽曲の楽しさ/悲しさに関する印象を表すには、印象尺度 8 に点数を与えればよく、その点数が「とても楽しい (7 点) — 楽しい (6 点) — 少し楽しい (5 点) — どちらとも言えない (4 点) — 少し悲しい (3 点) — 悲しい (2 点) — とても悲しい (1 点)」という評定スケールのどのあたりに位置しているかによって、印象の程度 (度合い) が示される。このとき、印象ベクトルは、

$$(nil\ nil\ nil\ nil\ nil\ nil\ nil\ x\ nil\ nil)$$

という形式で表される。但し、記号  $x$  は印象の程度を表す実数値 (0 点 ~ 8 点) であり、記号  $nil$  は *don't care* 項を意味する。

### 2.2 楽曲から生成される印象ベクトル

我々は、楽曲から抽出される物理的特徴量 (音の高さ、長さ、強さ、音色に関する時系列データ) を説明変数、印象尺度  $m$  ( $m = 1, 2, \dots, 10$ ) における評価値 (被験者 100 名が 80 曲を聴取し、付与した 7 段階評価値の平均値) を目的変数とする重回帰分析 (変数増加法) を行い、楽曲からその印象を記述す

表 2: 楽曲から生成された印象ベクトルの例

印象尺度	1	2	3	4	5
annie_laurie_pi	4.94	5.16	4.26	4.28	4.30
appassionata3_pi	1.70	1.90	3.02	3.16	5.40
arabesque1_pi	3.86	4.07	4.09	3.70	4.40
ave_maria	6.10	5.93	3.91	3.29	4.59
eine_kleine1_str	2.49	3.80	4.45	6.05	4.89
fuge_g_or	3.67	4.10	2.59	2.41	6.00
marche_tsc_pi	2.80	2.73	4.48	5.72	4.19
印象尺度	6	7	8	9	10
annie_laurie_pi	4.96	5.57	3.63	5.01	5.36
appassionata3_pi	3.04	4.69	3.26	2.50	3.92
arabesque1_pi	4.24	5.14	3.89	4.19	4.50
ave_maria	5.15	5.43	3.44	5.05	5.24
eine_kleine1_str	4.20	5.20	5.83	3.76	4.53
fuge_g_or	4.03	5.44	2.88	3.90	4.23
marche_tsc_pi	3.72	4.87	5.55	2.91	4.46

表 3: 印象語解釈ルール (一部)

印象尺度	1	2	3	4	5
哀れな	nil	nil	2.48	2.13	nil
懐かしい	5.01	5.40	nil	4.47	nil
古典的な	5.42	5.56	nil	3.47	5.57
心地よい	5.30	5.61	5.52	5.28	nil
優しい	5.49	5.79	5.62	5.27	nil
力強い	2.13	nil	nil	nil	nil
印象尺度	6	7	8	9	10
哀れな	nil	nil	1.75	nil	nil
懐かしい	5.37	nil	nil	nil	5.72
古典的な	nil	5.51	nil	5.06	5.09
心地よい	5.66	6.01	5.21	6.03	6.13
優しい	5.62	6.01	5.10	5.85	6.16
力強い	nil	nil	nil	2.38	nil

る印象ベクトルを生成する手法を提案した (具体的には, 印象ベクトルの第  $m$  要素の値<sup>\*1</sup> を求める重回帰式 (重回帰式を構成する説明変数とその係数, ならびに定数項) を決定した) [熊本 02b] .

本手法を用いて楽曲 (標準 MIDI ファイル形式) から生成された印象ベクトルの例を表 2 に示す .

### 2.3 印象語から生成される印象ベクトル

ユーザが入力した文から抽出された印象語は, 表 3 に示されたような印象語解釈ルール [熊本 02c] を用いて, 印象ベクトルに変換される .

この印象語解釈ルールは以下のようにして作成された . まず, 被験者 100 人 (男性 50 人, 女性 50 人) に印象語 (全部で 144 語<sup>\*2</sup>) と印象尺度 (10 個) を紙ベースで提示し, その印象語で表現される曲 (例えば「優しい曲」) はどのようなイメージの曲か, ということを 1 つ以上の印象尺度に対する 7 段

\*1 計算値が 0 点より小さい場合は 0 点に, 8 点より大きい場合は 8 点に丸められる .

\*2 我々の印象尺度で使われている 20 語を除いた .

表 4: 印象語を修飾しうる程度語 (一部)

桁違いに, 断然, 超, 最高に, すぎる, めちゃくちゃ, かなり, 非常に, 適度に, なるべく, 割と, 比較的, まあまあ, そこそこ, やや, 心持ち, あまり, さほど, 全然, ちよびり
--

階評価という形で回答してもらった . その結果, 印象語と各印象尺度における 7 段階評価値との対応関係が得られたが, このデータの中には被験者による変動が大きいものや, 不適切な印象尺度と考えた被験者が多いものもあった . そこで, それぞれの印象語毎に, データの平均値  $\bar{x}$  と標準偏差  $\sigma$  を求め, 区間  $(\bar{x} - 1, \bar{x} + 1)$  の間に含まれている被験者の数 (被験者による得点分布を正規分布と仮定した場合の理論上の数) を計算した . そして, この数が被験者 100 人の過半数に相当する 51 人以上であるデータだけを採用することにし, この条件を満たさなかったデータは除去した (実際には記号 *nil* を付与し, *don't care* 項として扱う) . その結果得られた対応表が表 3 のような印象語解釈ルールである .

## 3. 提案手法

### 3.1 程度語に関する検討

自由かつ自然な印象入力を可能にするというコンセプトに基づいて, シソーラス [大野 85] を参考にシステムが受理すべき程度語 119 語を決定した . その一部を表 4 に示す .

さて, 程度語の機能的役割は, その係り先である印象語が持つ「意味」を強めたり, 弱めたりすることと言える . 従って, 程度語を受理できるようにするためには, 印象語から生成された印象ベクトルの大きさを拡大または縮小するための仕組みが必要とされる . 印象尺度の評定スケール上では, 7 段階の評価が可能であり, 「少し」や「とても」によって修飾された印象語も評価することができる . このとき, 印象語そのものに対しては 6 点, 「とても」によって修飾された印象語には 7 点, 「少し」の場合は 5 点という点数が与えられていることから, 程度語によるベクトルの大きさの拡大・縮小はこの評定スケールに準じたものである必要がある . すなわち, ある印象語のある要素が 6 点であるとき, その印象語が「とても (少し)」によって修飾されているなら, 最終的に生成される印象ベクトルの当該要素は 7 点 (5 点) であることが求められる .

### 3.2 程度語による印象ベクトル変換手法

程度語によって修飾された印象語に対しては, 表 5 に示したような程度語定数表を用いて, その印象語から生成される印象ベクトルの大きさを拡大/縮小する . すなわち, ある印象語から生成された印象ベクトルを  $\vec{v} = [v_i]$  ( $i = 1, 2, \dots, 10$ ), その印象語を修飾している程度語の定数を  $a$  とすると, この印象ベクトルは

$$\vec{v}^{\prime} = [a(v_i - 4) + 4] \quad (i = 1, 2, \dots, 10) \quad (1)$$

と変換される . この操作は,  $a$  が 1 より大きければ拡大であり, 小さければ縮小を意味する . また,  $v_i$  の値が 4 点より大きければ, その印象尺度  $i$  の左側にある印象語の意味に対する操作であり, 小さければ, 右側にある印象語の意味に対する操作となる . 4 点 (すなわち「どちらとも言えない」を示す点数) のときは,  $a$  の値に関係なく, 元の値が保持され, 「どちらとも言えない」のままとなる .

表 5: 程度語定数表 (一部)

最高に (静かな)	1.75	比較的 (静かな)	0.54
本当に (静かな)	1.57	少し (静かな)	0.5
とても (静かな)	1.5	やや (静かな)	0.47
割と (静かな)	0.63	かすかに (静かな)	0.18

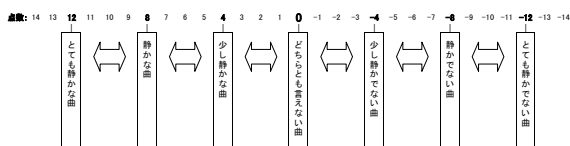


図 1: 点数の基準表 (29 段階評価)

この程度語定数表は以下のようにして作成された。まず、被験者 100 人 (男性 50 人, 女性 50 人)<sup>\*3</sup> に程度語 (全部で 119 語) によって修飾された印象語 (実際には印象尺度 1 の左側の印象語「静かな」を用いた) と図 1 のような点数の基準表を紙ベースで提示し, その程度語と印象語の組み合わせによって表現される曲 (例えば「わりと静かな曲」) が点数の基準表に照らし合わせてどのあたりに位置するか, ということに回答してもらった。この点数の基準表は、「どちらとも言えない」を原点 (0 点) とする数直線であり、「とても静かな曲」より静かな曲と「とても静かでない曲」より静かでない曲を想定することにより, 29 段階の評価が可能となっている。

以上の実験の結果, それぞれの程度語に対し, 図 2 に示したようなヒストグラムが得られた。このヒストグラムは「比較的静かな曲」に対する評価結果であるが, モードは 4, メディアンは 4.5, 平均は 4.57 (標準偏差は 1.93) と被験者の評価結果にばらつきが見られた。このようなデータはほかにも数多く見られたことから, 図 3 に示すアルゴリズムを導入し, 各程度語の定数を決定することにした。すなわち, このアルゴリズムによって出力された値を印象尺度の評定スケール (7 段階評価値) に換算し, 式 (1) を用いて  $a$  の値を逆算した結果が表 5 に示された程度語定数表である。ここで参考のために, 本アルゴリズムによる出力値を表 6 に示す。なお, 移動平均を用いたのは, 程度語の多義性に由来すると思われる評価結果のばらつきに対し, 最も多くの人に支持された解釈を選び出すためであり, 被験者数が過半数を超える最小範囲内のデータだけで平均を求めたのは, 少なくとも過半数の人が支持するような解釈に対し, その平均的な解釈を決定するためである。ちなみに, 「比較的静かな曲」の場合は, 移動平均の最大値は 16.2 (4 点) であり, 幅 5 のデータ区間 (2 点 ~ 6 点) で過半数を超え, そのときの範囲内のデータの平均値は 4.28 であった。この値は, 全データに対する平均値 (4.57) よりも若干小さく, モードの値 (4) に近づいた値となっている。

#### 4. 提案手法の有効性の検証

我々は, 程度語によって修飾された印象語から印象ベクトルを生成する手法を楽曲検索システム [熊本 02b] に実装し, 被験者 40 名 (男性 20 名, 女性 20 名) による評価実験を行った。このとき, 作成された程度語定数表の妥当性を検証するため

\*3 印象語解釈ルール作成のための実験に参加した被験者とは重複しない。

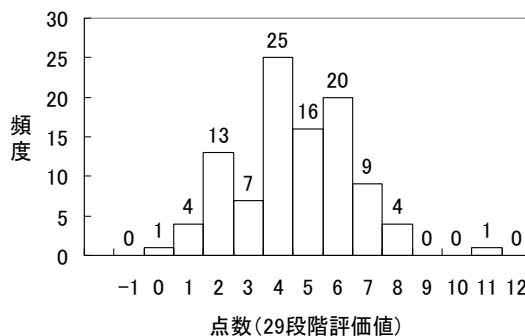


図 2: 「比較的静かな曲」の評価結果

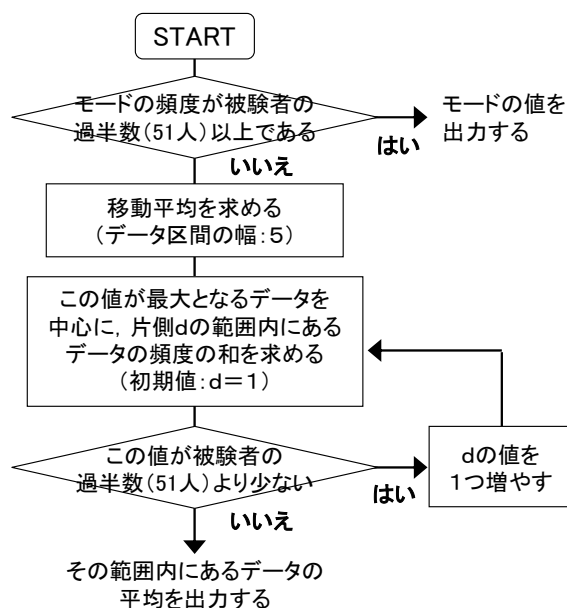


図 3: 程度語定数を決定するためのアルゴリズム

に, 程度語補正係数  $b$  を導入し, 式 (1) を,

$$\vec{v}^i = [ba(v_i - 4) + 4] \quad (i = 1, 2, \dots, 10) \quad (2)$$

と再定義した。

このような程度語補正係数を導入したのは (i) 程度語定数表を作成するための実験では, 29 段階の評価が可能であったが, 楽曲 (から抽出される物理的特徴量) と印象尺度とを対応付けるための実験では, 7 段階評価であったため, 同じ程度語 (「とても」や「少し」) でも意味の粒度が異なる (ii) 楽曲と印象尺度との対応付けは, 楽曲を聴取しながらの評価であり, 楽曲の印象と言葉 (程度語 + 印象語) が直接的に対応付けられたのに対し, 程度語定数表の作成では, 言葉の意味を印象尺度で意味づけており, 程度語どうしの相対的な意味関係が求められたにすぎない, といったことから, 程度語どうしの相対的な意味関係を保ちつつ, 程度語定数によって示される程度語の意味と楽曲印象との対応付けの妥当性を調べるためである。

さて, 実験の手順であるが, 各被験者は, まず (1) 入力可能な印象語 (164 語) と程度語 (119 語) のリストを見ながら, システムへの入力となる文を作った。次に (2) この文をシステム a ( $b = 0.60$ ), システム b ( $b = 0.88$ ), システム c

表 6: 程度語評価実験の結果 (一部)

最高に (静かな)	14	適度に (静かな)	6.53
極端に (静かな)	13.61	割と (静かな)	5.04
超 (静かな)	12.93	比較的 (静かな)	4.28
本当に (静かな)	12.58	少し (静かな)	4
とても (静かな)	12	やや (静かな)	3.76
いやに (静かな)	10.98	心持ち (静かな)	2.09
十分 (静かな)	8.54	かすかに (静かな)	1.46

表 7: ユーザが入力した印象語と程度語の組み合わせ

結構心地よくてかすかに寂しい曲, 割とカッコいい曲, 実にほっとする曲, えらく主張のある曲, 少々コミカル な曲, うんと透き通った曲, かなり暗い曲, 極めてうる さい曲, とても単純な曲
---

( $b = 1.00$ ) のそれぞれに入力し, 楽曲検索を行った. そして (3) それぞれの第 1 候補楽曲を聴取し, 入力した印象との適合度 (5 点満点) を 5 段階で評価した. 但し, 同じ楽曲が第 1 候補として提示された場合は, 最初の 1 回だけ楽曲聴取・適合度評価を行い, 2 回目以降は楽曲を聴取せずに, 同じ適合度を与えることとした. 以上の手順を約 10 回繰り返した.

表 7 にユーザが作った文の一部を示し, 表 8 に各システムにおける適合度の平均値, 標準偏差を示す.

表 8 は, いずれのシステムもほぼ同等の性能 (有意水準 5% で適合度に統計的な有意差はなかった) であったことを示しているが, 検索対象となった楽曲が 160 曲と少なかつたため, 多少印象ベクトルが異なっても, 同じ楽曲が検索されることがしばしば観測された. そこで, 同じ検索結果 (第 1 位のみ) が提示されたデータをノイズとして除去した上で, 2 つのシステムどうし ( $a$  と  $b$ ,  $a$  と  $c$ ,  $b$  と  $c$ ) を比較することにした. 結果を表 9 に示す. なお, 被験者のタイプミスによってシステムによって異なる印象が入力されてしまった場合のデータも取り除いた.

表 9 に示された結果からは, システム  $b$  が多少ながら高得点のように見えるが, いずれのシステム間においても統計的な有意差はなかった (有意水準 5%). 従って, いずれのシステムも同程度の検索性能を有していることになる.

もともと, 同じ楽曲を聴いても受ける印象が異なる, 受けた印象が同じでも言葉による表現の仕方が異なる, といった個人差が存在するわけだが, 程度語によって表現される意味の違いがこのような個人差よりも小さかつたため, 検索性能に差が生じなかつたものと考えられる.

## 5. まとめ

本稿では, 「やや」や「比較的」などの程度語によって修飾された印象語から印象に基づく楽曲検索システム用いられる検索条件 (印象ベクトルと呼ぶ) を生成するための手法を提案した. 具体的には, 印象語から生成される 10 次元の印象ベクトルを, その印象語を修飾している程度語に応じて, 拡大/縮小することを考え, そのための定数 (程度語定数と呼ぶ) を, 被験者実験により求めた.

結論としては, 程度語の使用を認めることにより, より柔軟

表 8: 各システムの性能評価の結果

システム	平均値	標準偏差	データ数
a (0.60)	3.76	1.24	394
b (0.88)	3.87	1.20	393
c (1.00)	3.80	1.26	393

表 9: 実験結果

システム	平均値	標準偏差	データ数
a (0.60)	3.80	1.24	287
c (1.00)	3.83	1.29	287
b (0.88)	3.85	1.17	118
c (1.00)	3.58	1.36	118
a (0.60)	3.80	1.25	267
b (0.88)	3.94	1.20	267

な検索が可能となる自然言語インタフェースを実現できるめどがあつたが, その一方で, 楽曲からの印象の受け方や受けた印象の言葉での表現の仕方における個人差に比べ, 程度語によって表現される印象の違いは小さく, 今回の提案手法が有効であるかどうかを検証するためには, 被験者間の個人差を考慮した評価実験を行うべきであることが示唆された.

## 参考文献

- [池添 01] 池添剛, 梶川嘉延, 野村康雄, “音楽感性空間を用いた感性語による音楽データベース検索システム”, 情処学論, vol.42, no.12, pp.3201-3212, Dec. 2001.
- [熊本 02a] 熊本忠彦, 太田公子, “印象に基づく楽曲検索: 検索ニーズに合った印象尺度の設計”, 情処学自然言語処理研報, vol.2002, no.4, pp.35-40, Jan. 2002.
- [熊本 02b] 熊本忠彦, 太田公子, “印象に基づく楽曲検索システム: Nグラム統計量の利用”, 情報技術レターズ, vol.1, no.LD-6, pp.63-64, Sept. 2002.
- [熊本 02c] 熊本忠彦, 太田公子, “印象に基づく楽曲検索: 自由かつ自然な印象表現の意味理解”, 第 4 回音声言語シンポジウム, vol.NLC-2002-57/SP2002-134, pp.81-86, Dec. 2002.
- [熊本 03] 熊本忠彦, 太田公子, “印象に基づく楽曲検索: 自然言語インタフェースの実現”, 情処学自然言語処理研報, vol.2003-NL-153, no.13, pp.97-104, Jan. 2003.
- [大野 85] 大野晋, 浜西正人, “類語国語辞典,” 角川書店, 東京, 1985.
- [佐藤 01] 佐藤聡, 小川潤, 堀野義博, 北上始, “感情に基づく音楽作品検索システムの実現に向けての検討”, 信学技報, vol.SP2000-137, pp.51-56, Feb. 2001.
- [辻 97] 辻康博, 星守, 大森匡, “曲の局所パターン特徴量を用いた類似曲検索・感性語による検索”, 信学技報, vol.SP96-124, pp.17-24, Mar. 1997.