

印象に基づく楽曲検索における印象語の意味理解

熊本 忠彦 太田 公子

独立行政法人 通信総合研究所 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 2-2-2

1 まえがき

たくさんの楽曲データの中からある特定の楽曲を探し出すための手段としては、曲名や作曲家名といった書誌的な情報を用いる従来からの方法だけでなく、ハミング(鼻歌)や歌声などの音響情報や歌詞などのテキスト情報を用いる方法 [1][2] も提案されている。これらの方法は、具体的な情報を入力できる場合には良いが、漠然とした要求しかないような場合には、不向きと言える。ある特定の楽曲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きたいという場合には、印象に基づく楽曲検索 [3]-[6] が有効であり、他の検索手段との統合によって「モーツァルトの悲しくて、ゆったりとした曲」といったより高度な検索も可能となる。

従来の印象に基づく楽曲検索において、印象の入力は、(1) 複数個の印象語の中から一つあるいは複数個を選択する [3]、(2) 複数個の印象語に対する多段階評価(例えば5段階評価) [4]、(3) 複数個の印象尺度(印象語の対)に対する多段階評価 [5][6] といった形式で行われる。そのため、入力可能な印象語の数が多いと、その中から適切な語を探し出すのに時間と労力を要してしまうし、逆に印象語の数が少ないと、適切な語が見つからない可能性が出てくる。従って、ユーザフレンドリーな検索を実現するためには、ユーザの自由かつ自然な印象入力を可能にする自然言語インタフェースの実現が必須と言える。

ある印象語を入力できるようにするためには、その印象語を検索条件に変換するためのルール(本稿では印象語解釈ルールと呼ぶ)が必要とされるが、そのようなルールの作成は、通常、楽曲の聴取、印象語や印象尺度による楽曲の評価、評価結果に対する多変量解析(主に因子分析)といった手順で行われるため、多くの印象語を対象にルールを作成することは実質的に困難であった。すなわち、印象語同士の意味的關係(生成される検索

条件の違い)を明確にするためには、印象語の数に比例して聴取すべき楽曲数を増やしていく必要があるが、楽曲の聴取には時間を要するため、高コストであり、経済的ではない。また、1回の楽曲聴取において、楽曲を評価する印象語の数が多いと、評価に時間を要するだけでなく、聴取した楽曲の印象がぼやけてくる恐れもある。印象語をグループ分けし、1回の楽曲聴取で評価する印象語の数を減らすという方法も考えられるが、そうすると同じ楽曲をグループの数だけ繰り返し聴取することになり、被験者がどのような影響を受けるか定かではない。

そこで、本稿では、入力できるようにしたい印象語を我々が提案している印象尺度(10対20語の印象語で構成されている)で評価することにより、印象語解釈ルールを作成するという間接的な方式を提案する [7]。この方式は、印象語同士の意味的關係を楽曲を媒介としないで決定しようという試みであり、楽曲を媒介とする従来の方式 [3]-[6] に比べ、誤差は大きくなるものと予想されるが、その反面、より多くの印象語を入力可能にすることができ、自由かつ自然な印象表現(印象語を含む文表現)を受理できる自然言語インタフェースの実現に資するものと考えられる。

以下、2で我々が開発した印象尺度ベースの楽曲検索システムについて述べ、印象語解釈ルール作成の際の前提条件について考察する。3でシステムが受理すべき印象語を選定し、4で印象語解釈ルールを作成する。5でこのルールを用いた意味理解(印象表現の検索条件への変換)の流れを示し、最後に6で本稿のまとめと今後の課題について述べる。

2 印象語解釈ルール作成に関する考察

印象語解釈ルールは、印象語と検索条件との対応関係を定義するものであり、我々が開発した印象尺度ベースの楽曲検索システム [6] に実装される。従って、この印象語解釈ルールに基づいて生成される検索条件は、我々のシステムで用いられ

ている検索条件と同じ仕様でなければならない。そこで、まず、このシステムにおいて、どのような検索条件が用いられているかを示す。

さて、我々のシステムへの印象入力は、表 1 に示された 10 対の印象尺度の中から 1 対以上を選択し、それぞれを 7 段階評価するという形で行われる。すなわち、「悲しい」曲を見つけない場合は、印象尺度 8 の「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)」の中から「悲しい」を選択すればよく、選択されなかった印象尺度に記号「nil」を付与することにより、検索条件は

(nil nil nil nil nil nil nil 2 nil nil)

というベクトル形式で記述される。同様に、「静かでゆったりとした」曲を見つけない場合には、印象尺度 1 の「静かな」と印象尺度 6 の「ゆったりとした」を評価すればよく、その結果、

(6 nil nil nil nil 6 nil nil nil nil)

という検索条件が生成される。

以上のことから、生成される検索条件への要求仕様としては、(1) 検索条件は 10 次元のベクトルとして記述される、(2) このベクトルの各要素はそれぞれ表 1 に示された印象尺度に対応している (すなわち、各要素の意味はその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、値は印象尺度の評定スケール (1 点~7 点) に準じる。但し、評価対象外の印象尺度には記号「nil」が付与される)、(3) 2 つ以上の検索条件を合成し、1 つにすることができる、の 3 つが挙げられる。

3 システムが受理すべき印象語

ユーザフレンドリーな自然言語インタフェースを実現するためには、ありとあらゆる印象語を受理できるようにしたいところであるが、実際問題としては時間と労力の制約からその数を限定する必要がある。そこで、受理すべき印象語を決定するために、以下の 3 つの基準を導入する。

- 基準 1) 印象に基づく楽曲検索において高頻度で利用される語 (62 語)
- 基準 2) 感情価測定尺度を構成する語 (24 語)
- 基準 3) 我々の楽曲検索システムにおいて利用されている語 (20 語)

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

表 2: 印象表現データベース [8] において出現頻度 5 以上であった印象語

静かな (51), 落ち着いた (30), 明るい (29), 気持ちが落ち着く (26), 心が癒される (22), ゆったりとした (16), 爽やかな (14), 荘厳な (12), 穏やかな (12), 楽しい (12), 軽快な (10), リラックスできる (9), 気持ちが安らぐ (7), 壮大な (7), 優しい (7), 綺麗な (6), 重厚な (6), 単純な (6), 軽やかな (6), 透明な (5), 心地よい (5), 柔らかい (5)

(括弧内の数字は出現頻度を表す)

表 3: 感情価測定尺度 [9] を構成する印象語

沈んだ, 哀れな, 悲しい, 暗い, 陽気な, うれしい, 楽しい, 明るい, 優しい, いとしい, 恋しい, おだやかな, 強い, 猛烈な, 刺激的な, 断固とした, きまぐれな, 浮かれた, 軽い, 落ち着きのない, 厳粛な, おごそかな, 崇高な, 気高い

基準 1 の高頻度で利用される語とは楽曲検索時に入力される文を集めた印象表現データベース [8] において 2 回以上出現した印象語のことであり (出現頻度 5 以上の印象語を表 2 に示す)、基準 2 の感情価測定尺度 [9] とは認知心理学の分野において提案されている音楽作品の感情的性格を測るための評価尺度のことであり (表 3 に示す)。また、基準 3 によって選択される語は表 1 に示された 20 語のことであり。

以上の 3 つの基準に基づいて、本稿では、印象語 82 語 (異なり数) とその否定語*1 82 語の計 164 語を受理すべき印象語として決定する。

*1 本稿では、肯定形 (例えば「うるさい」) に対してはその否定形 (「うるさくない」) を、否定形 (例えば「うるさくない」) に対してはその肯定形 (「うるさい」) を否定語と呼ぶ。

4 印象語解釈ルールの作成

印象語解釈ルールを作成するに当たり、それぞれの印象語の意味をどう解釈し、どのような検索条件に変換すべきか、ということ調べるために、以下のような被験者実験を行った。被験者は、20代、30代を中心とした男性50名、女性50名の計100名であり、各被験者に印象語と我々の印象尺度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、ということ各印象尺度に対する7段階評価という形で記述してもらった。但し、印象語の提示順序として10パターンを用意し、提示順序と所要時間による被験者への影響を分散させた。また、該当するイメージがない場合は無理に評価しなくてもよいと教示した。

この被験者実験の結果、それぞれの印象語に対し、印象尺度毎に評価結果の平均値 (\bar{x}) と標準偏差 (σ)、データ数 (N) が得られた。その一部を表4に示す。なお、データ数とは当該印象語を当該印象尺度で7段階評価した被験者の数である。

さて、この実験結果は、100名の被験者がそれぞれの印象語の意味をどう解釈したかを示しており、印象語と検索条件との対応関係を表している。しかしながら、データの中には標準偏差が大きいものやデータ数が少ないものもある。そこで、被験者 X 人中何人がその平均値 \bar{x} を支持しているかを示す指標として、支持率 p という関数を定義し、 $p \geq a$ を満足するデータだけを採用することにす。

まず、被験者による得点分布を正規分布と仮定すると、区間 $(\bar{x} - \sigma, \bar{x} + \sigma)$ の間には約 $0.683X$ 人の被験者が含まれており、同様に区間 $(\bar{x} - 2\sigma, \bar{x} + 2\sigma)$ の間には約 $0.954X$ 人の被験者が含まれていることになる [10]。そこで、区間 $(\bar{x} - 1, \bar{x} + 1)$ の間に含まれている被験者の数 n を計算し、

$$p = n/X$$

と定義する。なお、区間 $(\bar{x} - 1, \bar{x} + 1)$ とは、印象尺度の評定スケール上では、「静かな (6点)」を平均値 (\bar{x}) とすると、「少し静かな (5点)」より高得点で、「とても静かな (7点)」より低得点であることを意味している。

さて、 n の値の計算方法であるが、区間 $(\bar{x} - 1, \bar{x} + 1)$ は区間 $(\bar{x} - (1/\sigma)\sigma, \bar{x} + (1/\sigma)\sigma)$ と変形できるので、 n の値は、標準正規分布の面積 [10]

表4: 被験者実験の結果 (一部)

印象尺度		1	2	3	4	5
哀れな	\bar{x}	4.99	4.99	2.48	2.13	4.81
	σ	1.23	1.09	0.90	0.95	1.08
	N	81	72	86	94	67
心地よい	\bar{x}	5.30	5.61	5.52	5.28	4.57
	σ	1.07	1.08	1.05	0.82	1.12
	N	92	94	95	90	74
優しい	\bar{x}	5.49	5.79	5.62	5.27	4.45
	σ	0.93	0.77	0.83	0.80	1.13
	N	95	95	85	88	67
印象尺度		6	7	8	9	10
哀れな	\bar{x}	4.21	4.18	1.75	3.77	2.55
	σ	1.53	1.60	0.83	1.48	1.37
	N	78	73	93	56	86
心地よい	\bar{x}	5.66	6.01	5.21	6.03	6.13
	σ	0.82	0.99	0.93	1.18	0.99
	N	89	92	85	95	92
優しい	\bar{x}	5.62	6.01	5.10	5.85	6.16
	σ	0.77	0.67	0.86	1.10	0.76
	N	92	93	81	93	96

を用いて、計算することができる。すなわち、区間 $(\bar{x} - k\sigma, \bar{x} + k\sigma)$ の面積を $f(k)$ 、標本数 (データ数) を N とすると、

$$n = f(k)N$$

という式が与えられる。従って、 $X = 100$ 、 $k = 1/\sigma$ であることから、

$$\begin{aligned} p &= n/X \\ &= f(k)N/100 \\ &= f(1/\sigma)N/100 \end{aligned}$$

となる。ここで、被験者の過半数 (51人) が支持しているか否かを閾値として採用することにし、 $a = 51/100 = 0.51$ とする。すなわち、

$$p = f(1/\sigma)N/100 \geq 0.51$$

を満たすデータだけを採用し、他のデータは記号「nil」で置換した。その結果の一部を表5に示す。また、 $p \geq 0.51$ という条件では、すべてのデータが除去されてしまう印象語が5語あった。このような場合には、最も支持率の高いデータだけを採用することにした。表6にその5語を支持率とともに示す。

以上の操作の結果得られた印象語と印象尺度の評価結果との対応表は、印象語と検索条件との対応関係を示したものとなっており、この対応表が印象語解釈ルールとしてシステムに実装される。

表 5: 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
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
印象尺度	6	7	8	9	10
哀れな	nil	nil	1.75	nil	nil
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

表 6: 支持率の低い印象語

印象語	支持率
壮大でない	0.502
かっこよくない	0.495
幻想的でない	0.473
メリハリのない	0.470
ダイナミックでない	0.449

5 自由かつ自然な印象表現の意味理解

自由かつ自然な印象表現の意味理解, すなわち検索条件への変換は, 以下のような流れで行われる。まずはじめに, 入力された印象表現は日本語形態素解析システム juman[11] により形態素の列に分解される。このとき, 各形態素には基本形情報, 品詞情報, 品詞細分類情報といった情報が付与される。次に, この形態素解析の結果から印象語が抽出され, 印象語解釈ルール (表 5 を参照) を用いて検索条件に変換される。複数の検索条件が生成された場合には, それぞれの印象尺度において平均値が計算され, 1 つの検索条件に合成される。

6 まとめ

現在, 我々は, 印象に基づく楽曲検索のための自然言語インタフェースを開発している。多くの印象語を入力できるようにするためには, その印象語の数に見合った楽曲を聴取し, それぞれの印象語で評価する必要があるため, 多くの印象語を対象にすることは実質的に困難であった。そこで, 本稿では, 入力できるようにしたい印象語を既存の印象尺度で評価するという間接的な方式を提案した。この方式は, 楽曲の聴取を伴わないので, 比較的 low コストであり, 164 語の印象語を入力可能にすることができた。

今後の課題としては, 程度語 (「断然」や「比較

的」など) や比較語 (「もっと」や「もう少し」など) への対処, 未知語処理, ユーザフレンドリー性の検証, ユーザ個人の音楽感性 (印象の受け方) や言語感覚 (言葉の選び方) への適応などが挙げられる。

参考文献

- [1] 園田智也, 後藤真孝, 村岡洋一, “WWW 上での歌声による曲検索システム,” 信学論 (D-II), vol.J82-D-II, no.4, pp.721-731, Apr. 1999.
- [2] 小杉尚子, 小島明, 片岡良治, 串間和彦, “大規模音楽データベースのハミング検索システム,” 情処学論, vol.43, no.2, pp.287-298, Feb. 1999.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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,” 信学技報, vol.NLC2002-57, pp.81-86, Dec. 2002.
- [8] 熊本忠彦, 太田公子, “印象に基づく楽曲検索研究のための印象表現の収集,” 情処学論, vol.43, no.10, pp.3231-3234, Oct. 2002.
- [9] 谷口高士, 音楽と感情, 北大路書房, 京都, p.75, Jan. 1998.
- [10] P.G. ホーエル, 初等統計学, 浅井晃, 村上正康 (訳), 培風館, 東京, Feb. 1989.
- [11] 黒橋禎夫, 長尾真, 日本語形態素解析システム JUMAN version 3.61, <http://pine.kuee.kyoto-u.ac.jp/nl-resource/juman.html>, May 1999.