

E-038

## 印象に基づく楽曲検索のための対話処理手法

A Dialogue Processing Method for Impression-based Music-Retrieval System

熊本 忠彦

太田 公子

Tadahiko Kumamoto Kimiko Ohta

## 1. まえがき

ある特定の楽曲を探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,気分など)に合致するものを見つけたいという場合には,印象に基づく楽曲検索が有効であると考えられ,いくつかのグループによって研究されている[1]-[4].

先行研究の楽曲検索システムにおいて,印象の入力は,(1)複数個の印象語の中から1つあるいは複数個を選択する[1],(2)複数個の印象語に対する多段階評価[2],(3)複数個の印象尺度(印象語の対)に対する多段階評価[3][4]といった方式で行われている.そのため,入力可能な印象語の数が多いと入力に要する時間と労力が増大し,少ないと検索可能な印象のバリエーションが限られてしまう.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(以下,NLIと略す)の実現が必須と言える.

我々は,これまでの研究成果として,印象語164語(「穏やかな」や「優しい」など),程度語119語(「かなり」や「比較的」など)を解釈できるNLIを提案している[5][6].しかしながら,本NLIにおいてユーザの印象入力は1回限りであり,絞り込み検索を行うことができなかった.そこで本稿では,直前に用いられた検索条件を修正するための印象入力を解釈し,絞り込み検索を可能にする対話処理手法を提案する.

## 2. ベースとなる楽曲検索システムで用いられている検索条件

NLIの役割は,ユーザの印象入力から楽曲検索システムへの入力となる検索条件を生成することである.そこで本章では,提案NLIが実装される印象尺度ベースの楽曲検索システム[4]において用いられている検索条件の仕様を示す.

本システムへの印象入力は,システムが提示する10対の印象尺度(表1参照)の中から1対以上を選択し,7段階評価するという形で行われる.すなわち,楽しい楽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「楽しい」を選択すればよい.このとき,選択されなかった印象尺度には記号「nil」が割り振られ,

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

という10次元のベクトルが検索条件として生成される.同様に,暗く悲しい楽曲を見つけない場合には,印象尺度4の「暗い」と印象尺度8の「悲しい」を評価すればよく,その結果,

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

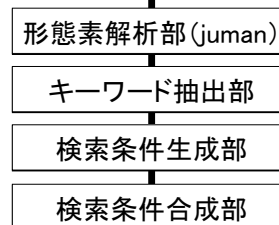
という検索条件が生成される.

以上のことから,NLIによって生成される検索条件には以下の3点を満たすことが要求される.(1)検索条件は10次元のベクトルで記述される,(2)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する.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(7段階評価値)に準じたものとなる.但し,評価されなかった印象尺度には記号「nil」が割り振

表1: 楽曲検索システムで用いられている印象尺度

番号	印象尺度を構成する印象語の対
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

ユーザの自由かつ自然な印象入力



検索条件(10次元のベクトル)

図1: 自然言語インタフェース(NLI)における処理の流れ

られる.(3)2つ以上の検索条件は合成され,1つにまとめられる.

## 3. 自然言語インタフェースにおける意味解釈

ユーザの印象入力は図1の流れに沿って処理され,最終的には検索条件として10次元のベクトルが生成される.以下,各処理について述べる.

## 3.1 形態素解析部

形態素解析部には汎用の日本語形態素解析システムであるjuman[7]を採用している.jumanは,入力された文を形態素の列に分解し,それぞれの形態素に基本形,品詞名,品詞細分類名といった情報を付与する.

## 3.2 キーワード抽出部

検索要求が同じでも「優雅な曲がいい」や「優雅な感じの曲を聴きたい」のように言い回しが異なる場合がある.そこで,このような言い回しの多様性を吸収するために,印象語や程度語をキーワードスポット的に抽出する方式を採用している.印象語には印象に基づく楽曲検索時に高頻度で利用される語,音楽作品の感情的性格を測るための語(感情価測定尺度[8]),ベースとなる楽曲検索システムにおいて採用されている語(表1)の82語(異なり数)とその否定語<sup>\*1</sup>の計164語を定義し,程度語にはシソーラス[9]から抽出した程度を表す副詞等の119語を定義する.

連絡先: 〒619-0289 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,熊本忠彦, kuma@crl.go.jp

\*1 本稿では,肯定形に対してはその否定形を,否定形に対してはその肯定形を否定語と呼ぶ.

表 2: 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 3: 程度語定数表 (一部)

程度語	定数 $a$	程度語	定数 $a$
最高に (静かな)	1.75	比較的 (静かな)	0.54
めちゃ (静かな)	1.57	少し (静かな)	0.50
とても (静かな)	1.50	やや (静かな)	0.47
割と (静かな)	0.63	かすかに (静かな)	0.18

表 4: 比較語定数表 (一部)

比較語	定数 $b$	定数 $c$	比較語	定数 $b$	定数 $c$
もっと	0.89	1.36	より	0.91	1.15
ますます	0.88	1.53	も少し	0.95	0.60
いっそう	0.78	2.22	ほんの	0.96	0.44
一段と	0.80	2.28	もう	0.92	0.92

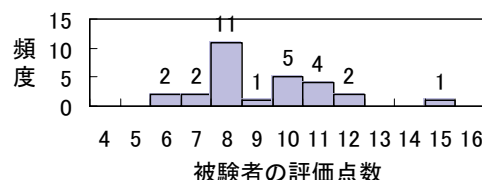


図 2: 「少し静かな曲 (4 点)」という仮定の下での「一段と静かな曲」の評価結果

### 3.3 検索条件生成部

検索条件生成部は、印象語解釈ルール (一部を表 2 に示す) [5] を用いて、抽出された印象語から検索条件を生成する。

程度語が抽出されている場合は、程度語定数表 (一部を表 3 に示す) [6] を用いて、印象語から生成された検索条件の大きさを拡大/縮小する。すなわち、ある印象語から生成された検索条件の第  $i$  要素を  $v_i$ 、その印象語を修飾している程度語の定数を  $a$  とすると、この検索条件の第  $i$  要素は

$$v_i' = a(v_i - 4) + 4 \quad (1)$$

と変換される。但し、 $v_i = nil$  のときは  $v_i' = nil$  とする。

### 3.4 検索条件合成部

複数の検索条件が生成された場合には検索条件の合成を行う。この操作は、印象尺度毎に平均値を計算することにより行われるが、「nil」は計算の対象外として扱われる。例えば、印象語「楽しい」と「優しい」からは、それぞれ

$$(nil \ nil \ nil \ nil \ nil \ nil \ nil \ nil \ 6 \ nil \ nil) \\ (5.49 \ 5.79 \ 5.62 \ 5.27 \ nil \ 5.62 \ 6.01 \ 5.10 \ 5.85 \ 6.16)$$

という検索条件が生成され、この 2 つが合成されると、

$$(5.49 \ 5.79 \ 5.62 \ 5.27 \ nil \ 5.62 \ 6.01 \ 5.55 \ 5.85 \ 6.16)$$

となる。

## 4. 対話的な検索条件修正のための対話処理手法

直前に検索が行われていて、かつユーザの印象入力に比較語\*2 が含まれている場合を絞り込み検索と考え、直前の検索において用いられた検索条件の第  $i$  要素  $v_i$  を

$$v_i' = \{(bx_i - x_i - 4b + 6)v_i + (x_i - 4)c\}/2 \quad (2)$$

と変換する。但し、 $x_i$  は比較語が修飾している印象語から生成された検索条件の第  $i$  要素であり、 $b$  と  $c$  は表 4 に示された比較語定数表中の値である。 $x_i = nil$  の場合は変換しないで、 $v_i' = v_i$  とする。 $x_i \neq nil$  の場合で、かつ  $v_i = nil$  の場合は、 $v_i$  を直前の検索において第 1 位となった楽曲の印象値で代用する。

\*2 「もっと」や「も少し」など程度の増加を表す副詞等 (14 語) をシソーラス [9] から抽出し、定義した。

この比較語定数表は以下のような被験者実験 (男性 50 名、女性 50 名) に基づいて作成された。まず、各被験者に対し、直前の検索条件を仮定 (例えば「静かな曲 (8 点)」を検索したと仮定) した上で、比較語によって修飾された印象語 (例えば「もっと静かな曲」と点数の基準表を提示し、その印象語で表現される楽曲が点数の基準表の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。但し、点数の基準表は、「静かな曲」を 8 点、「どちらとも言えない曲」を原点 (0 点)、「静かでない曲」を -8 点とする数直線であり、「とても静かな曲 (12 点)」より静かな曲と「とても静かでない曲 (-12 点)」より静かでない曲も想定された。また、仮定のパターンとして「とても静かでない曲 (-12 点)」から「とても静かな曲 (12 点)」までの 7 種類を用意し、1 つの仮定を 30 名程度の被験者が回答するよう被験者の割り振りを行った。

以上の実験の結果得られた被験者の評価結果には図 2 に示すようなばらつきが見られた。そこで、図 3 に示すアルゴリズムを導入し、被験者の過半数以上の人々が支持する最小範囲を求め、その中のデータの局所平均値を評価結果の代表値とした。そして、横軸を変換前の点数 (仮定の点数  $w_i$ )、縦軸を変換後の点数 (評価結果の代表値  $w_i'$ ) とする回帰分析を行い回帰式 (1 次式) を求めた。この回帰式を印象尺度の評定スケール (7 段階評価値) に変換し ( $v_i' = bv_i + c$ )、その係数  $b$  と定数項  $c$  を比較語定数表に登録した。さて、この 1 次式を用いれば、直前の検索条件を比較語と印象語の組み合わせに応じて変換できるわけだが、実験では比較語によって修飾される印象語として「静かな (6 点; 7 段階評価値)」を用いたため、他の点数の印象語 (例えば「暗い (2 点)」など) に対してこの式を適用することはできない。そこで、検索条件の変化分  $v_i' - v_i$  は、比較語が修飾している印象語から生成される検索条件  $x_i$  に比例すると仮定し、 $x_i = 6$  のとき  $v_i' = bv_i + c$ 、 $x_i = 4$  のとき  $v_i' = v_i$ 、 $x_i = 2$  のとき  $v_i' = v_i - (v_i' - v_i) = 2v_i - v_i' = (2 - b)v_i - c$ 、という連立式を立て、これを解くことにより、式 (2) を生成した。

## 5. 性能評価実験

我々は、提案手法を取り入れた NLI を印象尺度ベースの楽曲検索システム [4] に実装し、以下の手順で性能評価実験 (男性 16 名、女性 19 名) を行った。各被験者は、検索したい楽曲の印象を 4 文書いた上で、各文をキーボード入力し、楽曲検索\*3 を

\*3 検索対象は、数十秒から数分程度に短くアレンジされたクラシック系の 160 曲であり、ユーザの印象入力との距離 [4] が小さい 3 曲 (距離 100 以下) が提示された。

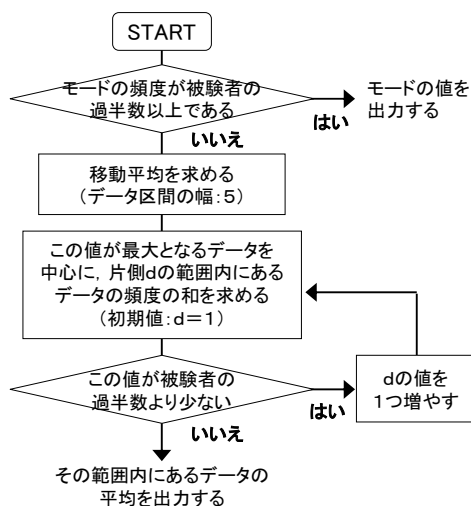


図3: 点数分布の代表値を決定するためのアルゴリズム

表5: 性能評価実験の結果

絞込み検索	得点 (得点の変化)	回数
なし	5点	36
	4点	1
	検索結果0件	2
	意味解釈失敗 (未知語)	17
	意味解釈失敗 (対象外)	31
あり	上がった	16
	変わらなかった	8
	下がった	2
	検索結果0件	13
	意味解釈失敗 (未知語)	5
	意味解釈失敗 (対象外)	9
合計		140

行った。検索結果がある場合は、第1位の楽曲を聴取し、印象の適合度(5点満点)を評価した。その結果が5点でない場合は、検索条件修正のための文を適宜入力し、絞込み検索を行った。但し、利用可能な比較語については別途提示した。また、検索の終了条件は、検索結果第1位の楽曲に対する適合度が5点になるか検索結果がなくなるまでとした。以上の実験の結果を表5にまとめる。

実験の結果、最終的に5点を得ることができたのは、全140回中、1回目の検索で5点を得た36回と絞込み検索により評価が「上がった」15回(残りの1回は3点止まりだった)の計51回(36.4%)であった。そこで、楽曲検索がうまくいかなかった原因を分析してみた。まず、絞込み検索「なし」において、「4点」というのは、被験者のミスタイプにより続く印象入力が絞込み検索のためのものと判断されず、対話が打ち切られたためである。「検索結果0件」の2回は「~すず」という接尾辞の解釈が適切でなかったためである。例えば、「明るすぎない曲」という印象入力は、明るすぎる楽曲以外の楽曲を指し、明るさに幅が存在する。しかしながら、現行のシステムでは印象を点で表しているため、このような幅に対処できず、とても暗い楽曲を提示してしまった。この点については今後の課題とする。「意味解釈失敗(対象外)」というものは全体で40回(28.6%)観測されているが、これは本システムの対象としないタイプの印象が入力されたためである。本システムは「音楽作品の感情的性格」と「楽曲聴取者の感情状態の変化」にタ

イブ分類されるような印象を対象としており[5]、「心に感じ入る曲」のような楽曲聴取者の感想や評論(11回)や「凍てついた厳冬の空のような曲」のような情景的なイメージ(7回)などは対象としていない。しかしながら、自由かつ自然な印象入力を行ってもらおうという観点から、入力可能な印象についての制約条件を詳しくは教示しなかった。この点、何らかの対策を考えたい。さて、本NLIでは164語の印象語を入力できるわけだが、未知語を解釈できずに対話を終了した事例が全体で22回(15.7%)観測されている。すなわち、未知語処理導入の必要性が示唆されており、今後の課題とする。一方、絞込み検索により適合度が「下がった」事例が2回あったが、同じ楽曲が提示されたにもかかわらず、評価が下がった場合と被験者が楽曲途中での印象の変化を検索条件としたため、適切な検索結果を提示できなかった場合のみであった。また、「変わらなかった」などにおいて、検索終了条件が満たされていないにもかかわらず、絞込み検索が行われていない事例が観測された。事前の教示方法を再検討したい。

以上のことから、我々のNLIは、意味解釈に成功した場合は、比較的良好な検索結果を提示しているが、その反面、解決すべき課題も多いと言える。なお、被験者へのアンケート調査では、検索される楽曲の種類が少ないという不満があった。実際、絞込み検索107回中31回において同じ楽曲が検索されており、何らかの改善策を検討したい。

## 6. まとめ

本稿では、印象尺度ベースの楽曲検索システムに絞込み検索機能を付加した自然言語インタフェース(NLI)を実装した。その結果、従来の楽曲検索システムでは見つけたい楽曲の印象をシステムが提示する少数の印象語の中から探す必要があったが、我々のシステムでは語彙的にも言い回し的にも自分の言葉で入力できるようになった。

今後は、音楽感性や言葉の意味解釈における個人差、未知語の出現、印象の時間的な変化、といった問題に取り組みたい。

## 参考文献

- [1] 大前寛子, 石橋直樹, 清木康, 安西祐一郎, “複数の楽器を用いた楽曲を対象とするメタデータ自動生成方式とその実現,” 情処研報, vol.2001-DBS-125, no.84, pp.145-152, Jul. 2001.
- [2] 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, no.137, pp.51-56, Feb. 2001.
- [3] 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.
- [4] 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, no.LD-6, pp.63-64, Sept. 2002.
- [5] 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.SP2002-134, pp.81-86, Dec. 2002.
- [6] 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,” 情処研報, vol.2003-NL-153, no.13, pp.97-104, Jan. 2003.
- [7] 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.
- [8] 谷口高士, “音楽と感情,” 北大路書房, 京都, 1998.
- [9] 大野晋, 浜西正人, “類語国語辞典,” 角川書店, 東京, 1985.