

印象に基づく楽曲検索における NLI と GUI の比較

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

(2004年4月1日, 通信総合研究所(CRL)と通信・放送機構(TAO)は統合され, 情報通信研究機構(NICT)として生まれ変わります)
〒619-0289 京都府相楽郡精華町光台 3-5

1 まえがき

大量にあるマルチメディアコンテンツの中からユーザが持つ何らかの判断基準(嗜好や感性, 気分など)に合致するデータを見つけるための手段として, 印象に基づく検索方式が研究されている[1, 2, 3]. 本方式は, 特定のデータを探し出すという目的には向かないが, ユーザ自身予想しなかったような意外なデータや全く知らないような新奇なデータを提示しうる発見的な検索手段と言える.

印象に基づく検索システムへの印象入力には, 通常, グラフィカルユーザインタフェース(GUI)が採用されており, システムが提示する(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], もしくは(3)複数個の印象尺度(印象語の対)に対する多段階評価[3]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, その中から自分が見つけたい楽曲の印象を表現する語(もしくは語の組合せ)を決定するのに時間と労力を要してしまい, 少ないと, 適切な語(もしくは語の組合せ)を決定できなかつたり, 迷つたりする可能性が高くなってしまふ. このような状況下におい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(NLI)が有効と考えられる[4, 5].

NLIは, ユーザ自身の判断基準(嗜好や感性, 気分など)をそのまま言葉にすればいいので, 誰でも利用できる上, 入力可能な印象語の数にも理論上は制約がなく, 増やしても入力コストは増大しない. 一方, GUIは, 入力可能な印象語が画面上に提示されるので直感性に優れ, 操作も単に入力したい印象語をマウスで選択するだけという単純なものである.

過去, NLIとGUIの比較は数多くなされてきたが, 感性検索シーンにおいての比較はまだ見当たらない. 意味(コマンド)体系の明確なパソコン操作などにおいては, 現在, GUIの方が優勢と考えられるが, 入力項目がユーザの印象という明確な意味体系を構築できない感性検索シーンでは, NLIとGUIのどちらが優れているのか, またその優劣は何を基準に決められるのか明らかでない. そこで, 本稿では, 印象に基づく楽曲検索システムにNLIとGUIを実装し, この2種類の楽曲検索システムの検索性能や使い勝手を比較することによって, NLI設計時に優先的に考慮すべき仕様について考察する.

なお, 我々のスタンスに最も近い比較事例としては, 原田らの文献[4]を挙げることができるが, 彼らが行ったのは, NLIを有した印象に基づく画像検索システム

表 1: GUI が提示する印象尺度

印象尺度	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

と GUI を有した内容(色相, 彩度, 輝度)に基づく画像検索システムとの比較であり, 感性検索シーンにおいて NLI と GUI を比較するものではなかった.

2 2種類の印象に基づく楽曲検索システム

本章では, 本稿において比較される2種類の楽曲検索システム(GUIベース, NLIベース)の概要を述べる. なお, どちらのシステムもユーザインタフェース以外は共通のモジュールで動作する.

2.1 GUIベースの楽曲検索システム[3]

GUIベースのシステムへの印象入力は, システムが提示する10個の印象尺度(表1参照)の中から1個以上を選択し, 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, 楽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「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 *don't care* 項を意味する記号「*nil*」が割り振られ,

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

という10次元の印象ベクトルが検索条件として生成される. 同様に, 暗く悲しい曲を見つけない場合は, 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. なお, 提示されていない印象を入力するためには, 複数個の印象尺度を評価し, 目的とする印象を合成する必要がある.

実際の GUI 画面は図1のとおりであり, ユーザは入力したい印象語をマウスで選択すればよい.



図 1: グラフィカルユーザインタフェース (GUI) 画面



図 2: 自然言語インタフェース (NLI) 画面

2.2 NLIベースの楽曲検索システム [5]

まずはじめに、この楽曲検索システムの NLI 画面を図 2 に示す。画面中段の右側にあるテキスト領域がユーザが印象入力を行うための領域であり、日本語漢字かな混じり文を入力し、改行を行えば、文の解釈、検索条件への変換、検索が行われ、その結果が画面左下のテーブル領域に表示される。

この楽曲検索システムに入力可能な印象語は 164 語であり、印象に基づく楽曲検索時に高頻度で利用される印象語、音楽作品の感情的性格を測るための印象語 (感情価測定尺度 [6])、GUI ベースのシステムにおいて提示される印象語 (表 1 参照) の計 82 語と、その否定語*1 の計 82 語からなっている。したがって、GUI で入力可能な印象語はすべて NLI でも入力できる。

印象語 164 語と検索条件との対応付けは、以下のような被験者実験 (男性 50 名、女性 50 名) の結果に基づいて行われた。まず、各被験者にそれぞれの印象語と印象尺度 (10 個 × 7 段階評価) を紙ベースで提示し、その

*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。例えば、「静かな」と「静かでない」はお互いに否定語の関係にある。

表 2: 印象語と検索条件の対応関係 [5] の例

印象尺度	1	2	3	4	5
衰れな	<i>nil</i>	<i>nil</i>	2.48	2.13	<i>nil</i>
懐かしい	5.01	5.40	<i>nil</i>	4.47	<i>nil</i>
優しい	5.49	5.79	5.62	5.27	<i>nil</i>
印象尺度	6	7	8	9	10
衰れな	<i>nil</i>	<i>nil</i>	1.75	<i>nil</i>	<i>nil</i>
懐かしい	5.37	<i>nil</i>	<i>nil</i>	<i>nil</i>	5.72
優しい	5.62	6.01	5.10	5.85	6.16

印象語で表現される曲 (例えば「穏やかな曲」) はどのようなイメージの曲か、ということをも 1 個以上の印象尺度に対する 7 段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わけだが、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、それぞれの印象尺度において、データの平均値 \bar{x} と標準偏差 σ 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、この数が被験者の過半数 (51 名) 以上であるデータだけを採用することにした。他のデータは除去し、記号「*nil*」を付与した。以上の操作の結果が印象語と検索条件との対応関係 (表 2 参照) を示しており、ある印象語が入力されると対応する検索条件が出力される仕組みとなっている。また、複数の印象語が入力された場合は、それぞれの印象語から検索条件が生成され、各要素毎に平均値を計算するという方法で合成される。例えば、印象語「楽しい」と「優しい」からは、それぞれ

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

という 2 つの検索条件が生成され、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。

3 NLI と GUI の比較および考察

本章では、NLI ベースのシステムと GUI ベースのシステムを検索性能や使い勝手の面から比較し、その結果について考察する。

3.1 検索性能評価実験の概要

まず、被験者 35 名に対し、「数多くのクラシック曲の中から、ホームページの BGM にしたいと思う曲を検索しようとするとき、あなたならどのような曲を検索しますか? 曲名や作曲者名といった具体的な情報ではなく、どんな印象の曲か、どんな雰囲気の曲か、どんな感じの曲か、といったことについて教えてください。ただし、楽器名や演奏形式、ジャンル、テンポについては触れないでください。ホームページの種類については、適当に仮定してください。」というアンケート調査を行い、検索したい楽曲の印象を 4 文書き出してもらった。

加えて、その印象を印象尺度（表1参照）を用いて入力するとしたら、どの印象尺度に対しどのような評価を行えばよいと思うか、ということにも回答してもらった。NLIに入力する文やGUIに入力する印象尺度評価を先に決めてもらったのは、検索欲求が乏しい状況下で、なるべく自由かつ自然な印象入力文を考え出してもらうとともに、システムの性能や操作性に影響されないようにするための工夫であり、安易な印象入力文や印象尺度評価を考え出すのを防ぐ狙いもあった。

次に、この4文のうちの第1文と第2文を用いて、以下の手順で楽曲検索*2を行ってもらった。すなわち、あらかじめ決定しておいた印象入力文や印象尺度評価をそれぞれNLIもしくはGUIを用いて入力してもらい、検索結果があれば、第1位の楽曲を聴取し、印象の適合度を5段階（適合している（5点）、少し適合している（4点）、どちらとも言えない（3点）、あまり適合していない（2点）、適合していない（1点））で評価することとし、なければ（例えば、意味解釈に失敗した場合や検索条件に合う楽曲が存在しなかった場合）、第3文以降を用いて再度楽曲検索を行うこととした。なお、実験に用いる楽曲検索システムの利用順序を被験者によって変え、利用順序が被験者に与える影響を分散させた。被験者は男性18名、女性17名の計35名であり、年齢層は20代がもっとも多く（24名）、他10代から50代までが数名ずつであった。

3.2 実験の結果と考察

まず、2種類の楽曲検索システムの検索性能として、成功率（検索に失敗しなかった割合）と検索に成功した場合の平均適合度（被験者毎の適合度の平均値をさらに平均した値）を調べた。その結果を表3に示す。

表3より、NLIベースのシステムでは、全88回の検索に対し、成功した（何らかの検索結果が提示された）のは68回（77.3%）であり、残りの20回は失敗であった（19回：検索条件の生成に失敗した、1回：検索条件に合う楽曲がなかった）。一方、GUIベースのシステムでは、全74回中70回（94.6%）が成功であり、失敗は4回だけであった。この割合の差は、統計的に有意であり（有意水準1%）、NLIベースのシステムの成功率がGUIベースのシステムに比べ、低いことを示している。また、検索に成功した場合の平均適合度は、NLIベースのシステムで4.12、GUIベースのシステムで4.01と、若干ながらNLIベースのシステムの方が高い値を示しているが、統計的に有意な差はなかった（有意水準5%）。つまり、検索に成功した場合の適合度という点では、両システムはほぼ同等と言える。

ここで、検索性能に関するアンケート調査の結果を図3に示す。この図より、適合度という点では、NLIベースのシステムもGUIベースのシステムもほぼ同等

表3: 2種類の楽曲検索システムの検索性能

	検索回数	成功率（回数）	平均適合度
NLI	88回	77.3%（68回）	4.12
GUI	74回	94.6%（70回）	4.01

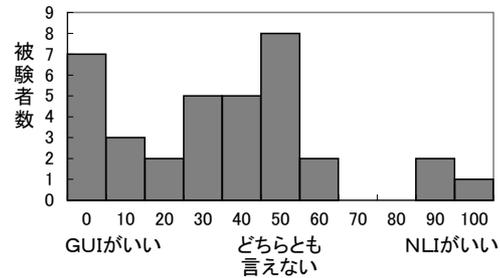


図3: 検索性能に関するアンケート調査の結果

表4: 検索性能に関するランク別適合度と成功率

ランク	平均適合度			成功率		
	NLI	GUI	差	NLI	GUI	差
1	3.96	4.00	-0.04	0.76	0.97	-0.21
2	4.21	4.14	0.07	0.87	1.00	-0.13
3	4.00	3.67	0.33	0.89	0.83	0.06
相関係数	0.15	-0.68	0.97	0.93	-0.78	0.97

であったにもかかわらず、ユーザの印象としては、GUIベースのシステムの方が高く評価されていることがわかる。そこで、この検索性能に関するアンケート調査において、低い評価得点（0点～20点）を与えた人をランク1、中位の評価得点（40点～60点）を与えた人をランク2、高い評価得点（80点～100点）を与えた人をランク3と分け、ランク毎に平均適合度と成功率を調べてみた。すると、表4のような結果が得られた。表中、相関係数はランクとの単回帰分析の結果であり、各システムの絶対的な検索精度（つまり平均適合度）が検索性能に関する印象の良し悪しにはあまり影響を与えず、その一方で、両システムの相対的な検索精度（つまり平均適合度の差）が大きく影響していることがわかる。また、表4は、NLIベースのシステムの成功率および両システムの成功率の差も重要な要素であることを示している。

以上のことから、検索性能に関する印象を良くするためには、絶対的な検索精度（平均適合度）を向上させることよりも、検索に失敗せず、常に何らかの結果を提示できるようロバスト性（成功率）を高めることの方が重要であると言える。

次に、システムの使い勝手に関して検討する。まず、NLIベースのシステムとGUIベースのシステムの使い勝手を100点満点で評価してもらったところ、図4に示すような結果が得られた。平均値は、NLIベースのシステムで82.9、GUIベースのシステムで83.0とほぼ同じであった。そこで、両システムの使い勝手を相対

*2 検索の対象となる楽曲データベースは、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曲からなり、インターネット上で公開され、二次利用が認められている楽曲（入手先：<http://nocturne.vis.ne.jp>、<http://k2works.com/nerve/>）をそのまま再利用した。検索結果としては、ユーザの印象入力との距離[3]が小さい3曲（ただし、距離100以下）が提示された。

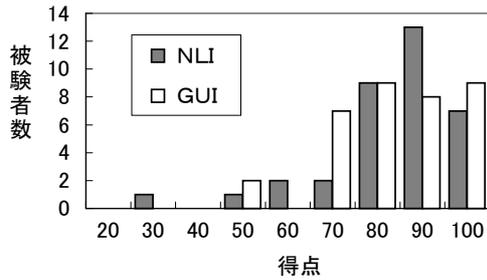


図 4: NLI と GUI の使い勝手

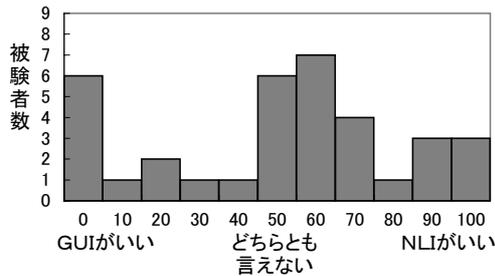


図 5: 使い勝手に関するアンケート調査の結果

的に評価してもらったところ、図 5 に示すような結果となった。「どちらも言えない」から若干「NLI がいい」の方に偏っているのがわかるが、その一方で「GUI がいい」と強く主張している被験者も多い。ここで、検索性能に関する考察のときと同様、評価得点に応じて 3 つのランクを用意し、ランク毎に平均適合度と成功率を調べてみた。その結果を表 5 に示す。表 5 より、平均適合度との間には弱い相関があり、成功率との間には強い相関があるのがわかる。したがって、使い勝手に関しても、システムのロバスト性（成功率）が最も重要であると言えるが、検索性能に関する考察のときと異なり、検索精度（平均適合度）に関しては、良いに越したことはないという傾向が見られる。

使い勝手に関しては、評価得点をつけてもらうと同時に、その理由も記述してもらった。GUI の使い勝手に関しては、長所として、印象語が提示されているという点と入力可能な印象の多様性（10 印象尺度 × 7 段階評価 = $8^{10} - 1$ パターン）という点が挙げられており、逆に、欠点としては、入力したい印象を印象尺度に対する評価という形で表現することの難しさや、印象尺度の数や種類、評価段階数への不満、そしてマウス操作に関する不満などが挙げられている。一方、NLI の使い勝手に関しては、長所として、自由かつ自然な文形式で印象を入力できる点や、単に文章を入力するだけという操作性が評価され、逆に、文解釈におけるロバスト性の低さや頭の中のイメージを言葉にすることの難しさなどが欠点として挙げられている。

なお、我々が用意した楽曲検索システムは「音楽作品の感情的性格」と「楽曲聴取者の感情状態の変化」にタイプ分類されるような印象のみを対象としており、「心に感じ入る曲」のような楽曲聴取者の感想や評論や「凍

表 5: 使い勝手に関するランク別適合度と成功率

ランク	平均適合度			成功率		
	NLI	GUI	差	NLI	GUI	差
1	3.83	4.22	-0.39	0.74	1.00	-0.26
2	4.21	3.89	0.32	0.89	0.98	-0.08
3	4.14	3.93	0.21	0.90	0.93	-0.02
相関係数	0.76	-0.81	0.79	0.90	-0.98	0.96

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などは対象としていない [5]。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教示しなかった。そのため、文解釈に失敗するケースが比較的多かったものと思われる。実際、文解釈に失敗した 19 文を調べてみると、うち 10 文が対象外の印象入力であり、残りの 9 文のうち、3 文がプログラムのバグによる解析失敗、6 文が未登録語（入力可能な印象語を決定する際の基準から外れた語）の出現による解析失敗であった。

以上のことから、使い勝手のよりよい検索システムを構築するためには、検索に失敗しないロバストなシステムであることが第一で、この前提が満たされるなら、NLI の方が好ましいと考えられる。また、印象の言語化を補助するような発想支援的メカニズムの必要性も示唆されている。

4 まとめ

本稿では、2 種類の印象に基づく楽曲検索システム（NLI ベースと GUI ベース）を用意し、被験者 35 名による検索性能評価実験を行った。その結果、我々は、検索に失敗しないロバストな検索システムの実現が第一で、このロバスト性が保証されるなら、NLI の方が検索性能、使い勝手の両面において、より良い印象を得ることができる、という結論に達した。

参考文献

- [1] 栗田多喜夫, 加藤俊一, 福田郁美, 坂倉あゆみ, “印象語による絵画データベースの検索”, 情処学論, Vol.33, No.11, pp.1373-1383, 1992.
- [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.SP2000-137, pp.51-56, 2001.
- [3] 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, 2002.
- [4] Syouji Harada, Yukihiro Itoh, and Hiromasa Nakatani, “Interactive Image Retrieval by Natural Language,” Opt. Eng., Vol.36, No.12, pp.3281-3287, 1997.
- [5] 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, 2003.
- [6] 谷口高士, “音楽と感情”, 北大路書房, 京都, 1998.