

# 印象に基づく楽曲検索：システムの実装と評価

熊本 忠彦<sup>†</sup>      太田 公子<sup>††</sup>

{kuma, kimiko}@crl.go.jp

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

<sup>†</sup>メディアインタラクショングループ, <sup>††</sup>自然言語グループ

〒619-0289 京都府相楽郡精華町光台 2-2-2

概要：我々は、音楽に詳しくない人でも簡単に利用できる「印象に基づく楽曲検索システム」を構築している。このシステムへの入力は、1対以上の印象尺度（楽曲印象を形容する語句の対からなる評価尺度）に対する7段階評価という形で行われる。つまり、明るい印象の楽曲を検索するには、明るさに関する印象尺度「とても明るい—明るい—少し明るい—どちらとも言えない—少し暗い—暗い—とても暗い」の中から「明るい」を選択し、検索すればよい。本稿では、本システムの構成と Windows OS への実装を示すとともに、被験者 25 名による主観評価実験の結果について考察する。なお、本システムが検索対象とする楽曲は、標準 MIDI ファイル形式のクラシック系（古典的西洋音楽）である。

## Implementation and Evaluation of our System of Retrieving Music Pieces with Similar Impressions by Inputting the Impressions of a Target Music Piece

Kumamoto, Tadahiko and Ohta, Kimiko

Keihanna Human Info-Communication Research Center, Communications Research Laboratory

2-2-2, Hikari-dai, Seika-cho, Sohraku-gun, Kyoto 619-0289, Japan

Abstract: We developed a music-retrieval system based on user's impressions. Users select one out of seven values from each of one or more impression scales to input their impressions of a target music piece into our music-retrieval system. For example, a user who wants to listen to sad music in a music database is asked to select "sad" from the impression scale "very funny — funny — a little funny — either funny or sad — a little sad — sad — very sad." Our system measures the distance between the impressions of every music piece in the music database and the impressions inputted by the user, and shows the results of retrieval according to distance. In this article, we describe the architecture and implementation of this system, and show the results of performance evaluation experiments.

### 1 まえがき

我々は、音楽に詳しくない人でも簡単に利用できるユーザフレンドリーな楽曲検索システムとして、「印象に基づく楽曲検索システム」を構築している。印象に基づく楽曲検索では、楽曲の印象という主観的な情報を入力とするので、音楽に詳しくない人でも利用できるし、また他の検索手段との統合によって「モーツァルトの悲しい曲」といったより高度な

検索も可能となる [辻 97, 佐藤 01, 池添 01]。

我々が目指すシステムは、検索したい楽曲の印象を自然言語で入力すれば、その人の音楽感性（楽曲印象の受け取り方）や言語感覚（印象を表現する言葉の選び方）を考慮した上で、そのような印象を持つ楽曲を出力するというものであるが、現段階では提示された 10 対の印象尺度（楽曲印象を形容する語句の対からなる評価尺度）[熊本 02a] の中から 1 対以上の印象尺度を選択し、7 段階評価することにより

検索を行うシステムを構築している。つまり、明るい印象の楽曲を検索したいときは、明るさに関する印象尺度「とても明るい—明るい—少し明るい—どちらとも言えない—少し暗い—暗い—とても暗い」の中から「明るい」を選択し、検索すれば、目的とする楽曲が得られるというものである。

本システムの最大の特徴は、楽曲から抽出される物理的特徴量（本稿では楽曲特徴量と呼ぶ）として、音の高さや強さ、長さ、音色に関する N グラム統計量（unigram, bigram, trigram, 4-gram）を採用している点にあり、これにより楽曲のダイナミックな変化を表現することが可能となっている。先行研究では、楽曲特徴量として音の高さや強さ、長さ、リズムやテンポ、拍子、調性（短調/長調）等の音楽構成要素に対する平均や分散、時間的割合といった静的な特徴量を用いていることが多い [佐藤 01, 池添 01]。しかしながら、本来時系列データである楽曲を静的な特徴量だけで表現するのは本質的に限界があるのは明らかである。そのため、文献 [辻 97] や [佐藤 01] では、時間的推移を考慮した特徴量として、連続する 3 音の音の高さや長さの推移をパターン化したものが試されているが、連続する音の数が一定 (= 3) であり、限定的な時間的推移しか取り扱えなかった。

本稿では、本システムの構成と Windows OS への実装を示すとともに、被験者 25 名による主観評価実験の結果について述べる。なお、本システムが取り扱える楽曲の形式は標準 MIDI ファイル（フォーマット 0 または 1）<sup>\*1</sup> であり、楽曲ジャンルとしてはクラシック系（古典的西洋音楽）を対象としている。

## 2 システムの構成

本システムは主に、検索対象としたい楽曲からその印象値（印象尺度毎に 1 個の実数値）を算出し、既定のファイルに登録するための「楽曲印象値自動付与部」と、ユーザが入力する印象値からその印象値に相当する楽曲を検索するための「楽曲検索部」から構成される。以下、この 2 つの処理部について述べる。

表 1: 3 種類の重み付け方法

重みタイプ	特徴量の種類	重み
$w_1$	全種類	1
$w_2$	音の長さ 上記以外	1 $\sum$ 音の長さ
$w_3$	音の長さ 音の強さ 上記以外	1 $\sum$ 音の長さ $\sum$ (音の強さ <sup>2</sup> × 音の長さ)

### 2.1 楽曲印象値自動付与部

まず、楽曲印象値の計算方法であるが、我々は、標準 MIDI ファイルから抽出される物理的特徴量を説明変数、印象尺度  $m$  ( $m = 1, 2, \dots, 10$ ) における楽曲印象値を目的変数とする重回帰分析（変数増加法）[管 00] を行い、その結果に基づいて、楽曲印象値計算式（重回帰式）と楽曲特徴量（重回帰式を構成する説明変数）を決定した [熊本 02b]。但し、目的変数である楽曲印象値には、SD 法に基づく印象評価実験（被験者：男性 39 名・女性 61 名の計 100 名、聴取曲：標準 MIDI ファイル形式のクラシック全 80 曲）において、被験者が各楽曲に対し印象尺度毎に付与した印象値（1 点から 7 点、もしくは無得点）の平均値を用い、説明変数となる特徴量の値には、その特徴量の相対的な存在割合 (= その特徴量と同じ特徴量に与えられた重みの和/その特徴量と同じ種類、同じ N グラム である特徴量に与えられた重みの和) を用いた。なお、各特徴量に与えられる重みの計算方法としては表 1 に示す 3 タイプを用意した。

ここで、印象尺度 1 の場合を例に、標準 MIDI ファイルから抽出される楽曲特徴量と、対応する楽曲印象値計算式の係数及び定数項を表 2 に示す。

表 2 において、「s」及び「42, 43, 46, 48」は音色に関する特徴量であり、「s」は無音を表し、他の数値は GM 規格に基づく音色番号に対応している。第 1 要素に「h, v, d」のいずれかを付された特徴量は、それぞれ連続した  $N$  個（印象尺度 1 の場合は  $N = 2$ ）の音における音の高さ、強さ、長さの時間的推移を表す特徴量であり、それぞれノートナンバー値、オンベロシティ値、ノートオンメッセージからノートオフメッセージが到着するまでの時間（ミリ

<sup>\*1</sup> MIDI 用語は文献 [田辺 01] を参考にしている。

表 2: 印象尺度 1 の楽曲印象値計算式 (重回帰式の偏回帰係数と定数項)

楽曲特徴量	係数	楽曲特徴量	係数
s	-3.835	v4 xs	17.226
42	-3.775	v7-5-4 d	97.405
43	-5.844	v10 e	0.559
46	1.334	v11 u	-5.663
48	-1.306	v12 u	-4.869
h44 d	-20.908	d3 u	-7.272
h62-58 u	59.918	d3 e	-1.877
h73-64 e	-61.359	d4 e	-2.671
h78-66 u	-83.845	d51 d	83.808
h81-78 u	99.110	d61 d	21.226
		定数項	5.075

表 3: 第 1 音  $x_1$  の抽象化ルール

特徴量の種類	抽象化ルール
音の高さ/音色	何もしない
音の強さ	$x_1/10$ の商に 1 を足した値で置換する。但し, $x_1 = s$ のときは何もしない。
音の長さ	$x_1/50$ の商に 1 を足した値で置換する。但し, $x_1 = 0$ のときは何もしない。

秒換算<sup>\*2</sup>に対応している。「-」でつながれた音は同時に発音された音を表しており、例えば「h62-58」は音の高さ 62 の音と 58 の音が同時に発音されたことを示している。なお、各特徴量の第 1 音  $x_1$  と第 2 音以降  $x_i (i = 2, \dots, N)$  にはそれぞれ表 3, 表 4 のルールが適用され、抽象化されている。また、連続した 2 音  $x_{i-1}$  と  $x_i (i = 2, \dots, N)$  の比較は、同時発音された音が対象の場合はその最大値と最小値を用いて行われ、単音が対象の場合は最大値 = 最小値として行われる。

印象尺度 1 では、楽曲特徴量として連続した 2 音から抽出された特徴量 (すなわち bigram) が用いられたが、本システムでは印象尺度によって異なった N グラム (もしくはその組み合わせ) が用いられている。同様に、重み付けの方法も印象尺度によって異なる。ここで、各印象尺度において用いられる N グラムと重みタイプを表 5 に示す。なお、音色に関する特徴量には、すべての印象尺度において unigram

\*2 無音の場合、音の長さは便宜的に 0 として計算される。

表 4: 第 2 音以降  $x_i (i = 2, \dots, N)$  の抽象化ルール

直前音 $x_{i-1}$	対象音 $x_i$	記号
s	s	se
s	*	sx
*	s	xs
(*...*)	(大きい...大きい)	u
(*...*)	(大きい...同じ)	u
(*...*)	(大きい...小さい)	ud
(*...*)	(同じ...大きい)	u
(*...*)	(同じ...同じ)	e
(*...*)	(同じ...小さい)	d
(*...*)	(小さい...大きい)	du
(*...*)	(小さい...同じ)	d
(*...*)	(小さい...小さい)	d

記号「\*」は任意の有音 (s 以外) を表している。

表 5: 各印象尺度において用いられる N グラム (楽曲特徴量) と重みタイプ

印象尺度	N グラムの組み合わせ	重みタイプ
1	bigram	$w_1$
2	bigram, trigram	$w_1$
3	unigram	$w_2$
4	bigram, trigram, 4-gram	$w_3$
5	unigram	$w_3$
6	bigram, trigram, 4-gram	$w_1$
7	bigram	$w_1$
8	bigram, trigram, 4-gram	$w_2$
9	bigram, trigram	$w_1$
10	bigram	$w_3$

が用いられているが、重みタイプは表 5 のとおりである。

## 2.2 楽曲検索部

さて、ユーザが入力できる印象は、表 6 に示した 10 対の印象尺度に対する 7 段階評価 (1 点から 7 点) もしくは無得点の組み合わせ ( $8^{10} - 1$  通り) という形で規定されている [熊本 02a]。印象尺度  $i (i = 1, 2, \dots, 10)$  に対するユーザの入力を  $x_i$ 、楽曲  $m$  の印象尺度  $j (j = 1, 2, \dots, 10)$  に対する値 (楽曲印象値) を  $v_{m,j}$  とすると、ユーザ入力と楽曲  $m$  の距離  $d_m$  は次式で計算される。

$$d_m = \sum_{k=1}^{10} \frac{(x_k - v_{m,k})^2}{\sigma_k^2} / \sum_{n=1}^{10} l_n$$

但し、 $x_k$  の値が無得点のときは、 $x_k = v_{m,k}$ 、 $l_k = 0$ 、そうでないときは、 $l_k = 1$  とする。また、 $\sigma_k^2$  は、前

表 6: 我々が提案している印象尺度

印象尺度	印象を形容する語句の対
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

表 7: 各印象尺度における楽曲分散の平均値

印象尺度	$\sigma_k^2$	印象尺度	$\sigma_k^2$
1	1.0623	6	1.2165
2	1.1212	7	1.2549
3	1.1532	8	0.9221
4	0.9792	9	1.3917
5	1.1760	10	1.2188

述の印象評価実験で得られた被験者データ (80 曲 × 100 人分) の楽曲  $m$ , 印象尺度  $k$  における分散を  $\sigma_{m,k}^2$  とすると,

$$\overline{\sigma_k^2} = \sum_{m=1}^{80} \sigma_{m,k}^2 / 80$$

と表され, 実際に用いられた値は表 7 のようになっている.

楽曲検索部は, 以上の式を用いて, ユーザ入力と各楽曲との距離を計算し, 距離が小さい楽曲を最大 20 曲提示する.

### 3 システムの実装

我々は本システムを Windows OS (Windows 98 Second Edition, NT, 2000, XP 上で動作を確認した) 上に実装した. そのメイン画面を図 1 に示す. この画面において, 「楽曲データベースの作成 (楽曲印象値の自動付与)」ボタンを押下すると, 検索対象としたい楽曲の印象値を計算し, 楽曲データベースに登録するための画面に進み, 「印象に基づく楽曲検索」ボタンを押下すると, 検索したい楽曲の印象 (すなわち検索条件) を入力するための画面 (図 2) に進む.

次に, 検索結果を表示するための画面を図 3 に

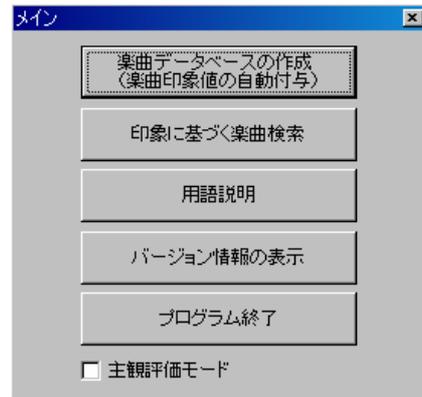


図 1: 印象に基づく楽曲検索システム (メイン画面)



図 2: 楽曲印象 (検索条件) を入力するための画面

示す. 画面の上半分に入力された検索条件が表示され, 下半分に検索結果として検索順位, 楽曲ファイル名 (標準 MIDI ファイル名), ファイルサイズ (KB), 距離, 評価が表示される. 楽曲を聴取するには, 聴取したい楽曲をマウスで選択し, 「再生」ボタンを押せばよい. なお, 本システムには, 主観評価実験を行うための機能が付加されており, メイン画面 (図 1) で主観評価モード (チェックボックス) をチェックすると, 被験者 ID (数字のみ) の入力を求められ, 検索結果画面 (図 3 の右下) で聴取した楽曲の適合度評価 (「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)」の 5 段階評価) が可能となる. 評価結果は被験者 ID に対応して自動的に生成されるテキストファイルに書き込まれる. 例えば, 被験者 ID を 007 とすると, 「log007.txt」というテキストファイルが生成され, 検索条件 (10 個の整数. 無得点に対しては



図 3: 検索結果が表示される画面

便宜的に 0 点を与えている)と検索結果(検索順位, 楽曲ファイル名, ファイルサイズ, 距離, 評価得点)が書き込まれる。

#### 4 主観評価実験

被験者 25 名(男性 12 名, 女性 13 名)に対し, 本システムの主観評価モードを用いて, 楽曲聴取による主観評価実験を実施した。被験者は検索条件を自由に入力でき, その結果に対し, 適合度評価(5 段階評価)を行った。被験者に関する情報を表 8 にまとめるとともに, 適合度評価の結果を表 9 に示す。但し, 表 8(c) の音楽経験については, 演奏家としての収入があるような人を「プロレベル」, 音楽大学などで専門的に勉強したような人を「セミプロレベル」, バンドやオーケストラ, 合唱団などに入っているような人を「アマチュアレベル」, 以上の条件には該当しないけれども一応演奏できるような人を「趣味レベル」, ほとんど演奏できないような人を「未経験」と定義している。

表 9 において, 学習データとは楽曲印象値自動付与部の設計(重回帰式の決定)の際に用いられた楽曲 80 曲を検索対象とした場合の結果であり, 未知データとはそれ以外の楽曲 80 曲を検索対象とした場合の結果である。また, 「適合している」を 5 点, 「少し適合している」を 4 点, 「どちらとも言えない」を 3 点, 「あまり適合していない」を 2 点, 「適合していない」を 1 点として計算し, 検索順位毎に平均値を求めている。

表 8: 主観評価実験に参加した被験者(25 名)に関する情報

(a) 年齢分布

20 歳未満	2	(8.0%)
20 ~ 29 歳	15	(60.0%)
30 ~ 39 歳	6	(24.0%)
40 ~ 49 歳	2	(8.0%)

(b) 職業に関する情報

有職	3	(12.0%)
生徒・学生	15	(60.0%)
専業主婦	2	(8.0%)
無職	5	(20.0%)

(c) 音楽経験

プロレベル	1	(4.0%)
セミプロレベル	1	(4.0%)
アマチュアレベル	2	(8.0%)
趣味レベル	11	(44.0%)
未経験	10	(40.0%)

表 9: 楽曲検索・聴取による適合度評価実験の結果

	1 位	2 位	3 位
学習データ	4.3	4.1	3.9
未知データ	3.8	3.8	3.7

表 9 から, 学習データに対しては一定の評価を得ているが, 未知データに対しては十分な性能が得られていない。この原因として, 以下の 3 点が考えられる。

(1) 同じ楽曲を聴取しても, 受ける印象には個人差がある。

表 7 に示した, 印象尺度  $k$  における分散の平均値  $\overline{\sigma_k^2}$  という概念は, 数学的には厳密ではないが, ユーザによる 7 段階評価における個人差の程度を示す指標と考えられ, その値が各印象尺度において 1 前後であることから, 個人差が比較的大きいことを示している。

(2) 線形近似となる重回帰式は楽曲印象値計算式には適さない。

先行研究において, 重回帰式を用いた楽曲検索システムとニューラルネットワークを用いた楽曲検索システムが比較されており, 総じてニューラルネットワークの方がより高い検索精度を得ている[池添 01]。今回, 我々は個人適応への応用(例えば, 楽曲印象の受け取り方や受け取った印象の表現の仕

方における個人差を楽曲印象値計算式等に反映させること)を考え、計算式がブラックボックス化されてしまうニューラルネットワークを避けたが、考え直す必要があるかもしれない。

(3)N グラム統計量だけでは楽曲の特徴を十分に表現できない。

N グラム統計量を用いた楽曲検索研究は今までなかったため、N グラム統計量だけで検索精度がどの程度得られるのかを確かめるために、今回は、敢えてN グラム統計量だけを用いた。今後は、従来から用いられている静的な特徴量との組み合わせも考えていく必要がある。

なお、被験者に対して実施したアンケートにおいて、システムの使いやすさ(100点満点)に対しては、86.8点を得ており、画面・操作手順等の設計に関しては合格点と考えられる。

## 5 まとめ

本稿では、我々が開発した印象尺度ベースの楽曲検索システムに関し、その構成と実装を示すとともに、被験者25名による主観評価実験の結果について報告した。本システムは、1対以上10対以下の印象尺度を7段階評価することにより、その印象に相当する楽曲を検索し、提示するというものであり、検索対象としたい楽曲の印象を数値化して、既定のファイルに登録するための「楽曲印象値自動付与部」とユーザが行った印象尺度の評価結果(検索条件)と各楽曲が有する印象との距離を計算し、距離の小さいものを検索結果として提示する「楽曲検索部」とから構成される。なお、本システムは、標準MIDIファイル形式(フォーマット0または1)を処理対象とし、楽曲ジャンルとしてはクラシック系を対象としている。

我々は、本システムのような「印象に基づく楽曲検索」システムの利用者を音楽に詳しくない人と仮定している。そのため、楽曲印象値計算式を設計するための印象評価実験([熊本02b]参照)や今回の主観評価実験(表8(c)参照)における被験者として、音楽経験の乏しい趣味レベル・未経験の人を優先的に採用している。したがって、主観評価実験の結果の良し悪しが必ずしもシステムの性能の良し悪しを意味するものではない。当該研究分野では、システム

同士の性能比較に資する音楽データベース[後藤02]が作成され、配布され始めているが、より正確な性能比較を行うためには、同じ被験者による評価実験が不可欠と考えられる。この目的のため、本システムを公開する予定であり、自由に試していただければ、と考えている。

今後の課題として、任意の印象表現(テキスト形式)を入力可能な楽曲検索システムや、対話的に新規検索や絞り込み検索を行える楽曲検索システム、音楽感性と言語感覚における個人差を考慮した楽曲検索システムの開発を順次行う予定である。

## 参考文献

- [後藤02] 後藤真孝, 橋口博樹, 西村拓一, 岡 隆一: RWC 研究用音楽データベース: クラシック音楽データベースとジャズ音楽データベース, 情処研報(音楽情報科学), 2002-MUS-44, 5, pp. 25-32 (2002) .
- [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, 42, 12, pp. 3201-3212 (2001) .
- [菅00] 菅 民郎: 多変量統計分析, 現代数学社, 京都 (2000) .
- [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, 情処研報(自然言語処理), 2001-NL-147, 6, pp. 35-40 (2002) .
- [熊本02b] 熊本忠彦, 太田公子: 印象に基づく楽曲検索: 楽曲印象値の自動付与, 情処研報(データベースシステム), 2002-DBS-127, 12, pp. 89-96 (2002) .
- [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(音声), SP2000-137, pp. 51-56 (2001) .
- [田辺01] 田辺義和: Windows サウンドプログラミング, 翔泳社, 東京 (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(音声), SP96-124, pp. 17-24 (1997) .