

印象に基づく楽曲検索：楽曲印象値の自動付与

熊本 忠彦[†] 太田 公子^{††}

{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

[†]メディアインタラクショングループ、^{††}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、音楽情報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、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。このような楽曲検索を行うためには、ユーザが入力した印象と各楽曲が有する印象との距離を計算することが必要であり、そのためには、印象を数値化して、演算可能な形式で記述することが重要である。さらに、大規模な楽曲データベースに対しても高速な検索処理を行うためには、楽曲印象の数値化を事前に行っておくことも必要となってくる。そこで本稿では、各楽曲（標準 MIDI ファイル）が有する印象を数値化して、ファイルに登録する「楽曲印象値自動付与システム」を提案する。なお、本システムはクラシック系（古典的西洋音楽）の楽曲を用いて設計されているので、その処理の対象もクラシック系となる。

Automatically and Numerically Expressing Impressions of a Music Piece for Music-Retrieval based on User's Impressions

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 are developing 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 music-retrieval system would measure the distance between the impressions of every music piece in the music database and the impressions inputted by the user. As a first step to achieve this, we developed a system of numerically expressing and saving into a file impressions of every music piece in a user-specific music database. In this article, we describe the design, architecture, and implementation of this system.

1 まえがき

我々は、誰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ことで、音楽情報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 99, 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、そのような印象を持つ楽曲が出力されるというものであるが、現段階では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02]を提示し、1対以上の印象尺度に対する7段階評価により動作するシステムを構築している。つまり、明るくて楽しい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、楽しさに関する印象尺度「とても楽しい—楽しい—少し楽しい—どちらとも言えない—少し悲しい—悲しい—とても悲しい」の中から「楽しい」を選択して、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。

さて、以上のようなシステムを構築するためには、ユーザが入力した印象と各楽曲が有する印象との距離を計算することが必要であり、そのためには、印象を数値化して、演算可能な形式で記述することが重要である。さらに、大規模な楽曲データベースに対しても高速な検索処理を行うためには、楽曲印象の数値化を事前に行っておくことも必要となってくる。そこで本稿では、楽曲が有する印象を数値化するための手法として「楽曲印象値自動付与手法」を提案する。本手法は、標準MIDIファイル形式(フォーマット0または1)*¹の楽曲を対象としており、楽曲から抽出される物理的特徴量に基づいてその楽曲の印象値（それぞれの印象尺度に対し1つの実数値が付与される）を算出する。なお、本手法はWindows OS上に実装されており、「楽曲印象値自動付与システム」として近日公開予定である。本稿では当システムの実装について述べるとともに、その公開情報についても触れる。

2 提案手法の設計方針

楽曲印象を数値化したものを「楽曲印象値」と呼び、楽曲から抽出され、楽曲印象値を計算するために用いられる物理的特徴量を「楽曲特徴量」と呼ぶことにすると、「楽曲印象値自動付与手法を設計する」とは、楽曲（標準MIDIファイル）からどのよう

な楽曲特徴量を抽出し、どのような計算式を用いて、どのような楽曲印象値を出力するのか、という問題に帰着する。そこでまず、楽曲印象値や楽曲特徴量、楽曲印象値計算式の設計方針について考察する。

2.1 楽曲印象値の設計方針

文献[池添01]では、SD (Semantic Differential) 法に基づく主観評価実験データに対する因子分析の結果から、音楽感性空間と呼ばれる5次元の因子空間を構成し、ユーザが入力する印象と楽曲が有する印象とをこの空間内の座標値として表している。しかしながら、因子軸の意味の解釈は人手によるので個人差があり、楽曲に付与された座標値が実際にどのような印象を表しているのかを端的に示すことは難しい。また、楽曲の印象を1つの点で表しているため、すべての印象尺度（彼らのシステムでは8個）に何らかの値を入力しなければならず、印象尺度に対する評価として「どちらでもない（楽曲印象値が不定な状態）」を認めていない。そのため、明るい楽曲を検索するつもりで、明るさに関する印象尺度の評価を「明るい」にしても、実際には明るさ以外の印象尺度に対して「どちらとも言えない」に相当する値（1点～7点の7段階評価では4点）を持つ楽曲が検索されることになる[佐藤01]。

現在、我々はユーザが入力できる印象を10対の印象尺度の組み合わせ（ $8^{10} - 1$ 通り）という形で規定している[熊本02]。各印象尺度は楽曲の印象を形容する語句の対から構成されており、ユーザは1対以上の印象尺度に対し7段階評価を行うことにより検索したい楽曲の印象を入力することができる。そこで我々は、各印象尺度における7段階評価結果を1から7の値に対応させ、楽曲印象値を同じ数値軸上（つまり印象尺度上）の実数値として表すことにする。これにより、各印象尺度において楽曲印象値がどのような印象を表現しているのか明確になるし、ユーザが入力する「どちらでもない」という評価結果をその印象尺度に関しては点がない状態だと考えれば、「どちらでもない（無得点）」と「どちらとも言えない（4点）」の区別が可能となる。なお、我々が提案している印象尺度は表1のとおりである。

2.2 楽曲特徴量の設計方針

標準MIDIファイルを対象とする楽曲検索研究では、楽曲特徴量として、音の高さや強さ、長さ、リス

*¹ MIDI用語は文献[田辺01]を参考にしている。

表 1: 我々が提案している印象尺度 [熊本 02]

No.	印象尺度を構成する語句の対
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

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 99, 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ものと考えられる。

そこで我々は、楽曲のダイナミックな変化を表現するために、音の 3 要素である音の「高さ・強さ・長さ」と知覚への影響が大きい「音色」を対象に、N グラム統計量をベースにした楽曲特徴量を設計することにす。但し、N グラム統計量の N の値が大きくなるとデータスパース性という問題が生じてくるので、今回は unigram から 5-gram までの N グラム統計量を利用することにした。従来研究 [辻 97, 佐藤 99, 佐藤 01] でも、このような音の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などが提案されているが、連続する音の数が一定（つまり trigram のみ）であり、限定的な時間推移しか取り扱えなかった。

2.3 楽曲印象値計算式の設計方針

楽曲特徴量から楽曲印象値を計算するための方法としては、重回帰式のような直線近似ベースの方法とニューラルネットワークのような曲線近似ベースの方法とが考えられる。一般に後者の方がより高い検索精度を得られる可能性があるが [池添 01]、ニューラルネットワークでは計算式がブラックボッ

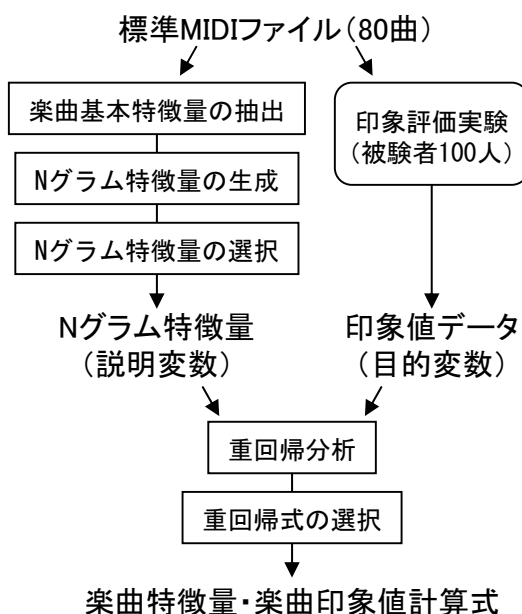


図 1: 楽曲印象値自動付与手法の設計手順

クス化されてしまうため、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）が難しくなると考えられる。

そこで我々は、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）を決定することにす。ニューラルネットワークの利用は今後の課題としたい。

3 提案手法の設計

前章で述べた設計方針に基づいて、楽曲特徴量及び楽曲印象値計算式を設計する。具体的には、図 1 に示した設計手順に従って設計する。以下、この流れに沿って、各手順を説明する。

3.1 印象評価実験による印象値データの獲得

楽曲が有する印象を数値化する際の基準となるデータを得るために、SD 法に基づく主観評価実験として、以下のような印象評価実験を行った。

被験者は、男性 39 名、女性 61 名の計 100 名であり、プロレベル（演奏家としての収入があるような

人)1名、セミプロレベル(音楽大学などで専門的に勉強したような人)7名、アマチュアレベル(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)20名、趣味レベル(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)46名、未経験者(ほとんど演奏できないような人)26名と音楽経験が豊かでない人も多数含まれていた。印象に基づく楽曲検索は、音楽経験の豊富な人というよりも、そうでない人に対して特に有効な検索手段であり、そういう人の音楽感性を反映したデータを利用することは提案手法を設計する上で重要なことと言える。また、実験で用いた楽曲は標準 MIDI ファイル形式のクラシック 80 曲であり、インターネット上で公開されていたものを採用した。但し、実験時間の都合により、楽曲聴取に要する平均試聴時間が 1 分前後となるよう楽曲の長さを調整した。被験者は、各楽曲を 2 回まで試聴することができ、その間にすべての印象尺度に対し 7 段階評価もしくは「どちらでもない」の評価を行うことが求められた。

ここで、各印象尺度の 7 段階評価結果に対し点数を割り振った。例えば、明るさに関する印象尺度では、「とても明るい」を 7 点、「明るい」を 6 点、「少し明るい」を 5 点、「どちらとも言えない」を 4 点、「少し暗い」を 3 点、「暗い」を 2 点、「とても暗い」を 1 点とし、「どちらでもない」は無得点とした。

以上の結果得られた 80,000 個(100 人 × 80 曲 × 10 印象尺度)のデータから各楽曲毎の平均を求め、印象値データ(800 個 = 80 曲 × 10 印象尺度)とした。但し、無得点のデータは事前に除外し、計算には用いなかった。

3.2 楽曲基本特徴量の抽出

標準 MIDI ファイル形式(フォーマット 0 または 1)の楽曲から各トラックチャンク各チャンネル毎に楽曲基本特徴量を抽出し、それぞれを 1 つのストリームデータとする。例えば、1 トラックチャンク・3 チャンネルの楽曲からは 3 つのストリームデータが生成される。抽出される楽曲基本特徴量は、音の高さ、音の強さ、音の長さ、音色情報の 4 種類であり、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(ミリ秒)、GM 規格に基づく音色番号に対応している。

214 81 110 48
428 74 110 48
0 s s s
214 62 110 48 71 110 48 79 110 48
0 s s s
642 79 116 48

図 2: 楽曲基本特徴量の抽出例

ここで、楽曲基本特徴量の抽出例としてストリームデータの一例を図 2 に示す。ストリームデータにおいて、各行の第 1 列が音の長さ、第 2 列が音の高さ、第 3 列が音の強さ、第 4 列が音色情報に対応している。また、同一トラックチャンク同一チャンネルにおいて、2 音以上が同時に発音している場合を「和音」と定義し、和音がある場合は、2 音目以降の楽曲基本特徴量(音の長さを除く)を第 5 列以降に繰り返し記述する。また、各チャンネルにおいて、そのチャンネルの無音状態を休符と定義し、音の長さを 0、音の長さ以外を記号「s」で表す。

3.3 N グラム特徴量の生成

N グラム特徴量は、楽曲特徴量の候補となる特徴量であり、以下の手順で楽曲基本特徴量から生成される。

[N グラムの生成]

まず、各ストリームデータから 4 種類の楽曲基本特徴量を分離し、音色情報からは unigram($N = 1$)を、それ以外の特徴量からは N グラム($N = 1, 2, 3, 4, 5$)を生成する。例えば、図 2 に示された抽出例の音の高さからは図 3 のような N グラムが生成される。なお、和音は、値の大きい順に並べ替えられ、リスト形式の入れ子として記述される。

[N グラムの抽象化]

次に、音色情報以外の楽曲基本特徴量から生成された N グラムの各要素($x_1 x_2 \dots x_N$)を表 2、表 3 の抽象化ルールに基づいて置換する。表 2 のルールは、各 N グラムの第 1 要素 x_1 に適用され、楽曲基本特徴量の種類に応じてその要素を置換する。このとき、リスト形式の入れ子を 1 つの記号(例えば 79-71-62)で記述するとともに、楽曲基本特徴量の種類を示すためのタグとして、音の高さなら h、音の強さなら v、音の長さなら d を付加する(例えば

unigram:	(81)(74)(s)((79 71 62))(s)(79)
bigram:	(81 74)(74 s)(s (79 71 62))((79 71 62) s)(s 79)
trigram:	(81 74 s)(74 s (79 71 62))(s (79 71 62) s)((79 71 62) s 79)
4-gram:	(81 74 s (79 71 62))(74 s (79 71 62) s)(s (79 71 62) s 79)
5-gram:	(81 74 s (79 71 62) s)(74 s (79 71 62) s 79)

図 3: N グラム生成の例 (音の高さの場合)

表 2: 第 1 音 x_1 の抽象化ルール

楽曲基本特徴量	抽象化ルール
音の高さ	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。

h79-71-62 } 一方、表 3 のルールは、各 N グラムの第 2 要素以降 x_i ($i = 2, \dots, N$) に適用され、その直前の要素 x_{i-1} との比較結果に応じて x_i を対応する記号で置換する。このとき、 x_{i-1} と x_i の比較は、それぞれの最大値同士、最小値同士で行われるが、和音以外では最大値 = 最小値として扱われる。

以上の処理の結果、例えば、図 3 の N グラムは抽象化され、図 4 のようになる。ここで、音の高さの場合を例に、抽象化による異なり N グラム数の変化を表 4 に示す。

[N グラム特徴量の生成]

以上のようにして抽象化された N グラムの異なり N グラムを、本稿では「N グラム特徴量」と呼ぶ。そして、それぞれの N グラム特徴量は、その相対出現頻度に重み w を掛けたものを値として持つ。但し、相対出現頻度は、楽曲基本特徴量の種類毎、N グラム統計量の N 値毎に計算され、小数点第 4 位で四捨五入される。例えば、図 4 の bigram からは 4 つの N グラム特徴量が生成され、(hs sx) の相対出現頻

表 3: 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 (つまり s 以外) を表している。

unigram:	(h81)(h74)(hs)(h79-71-62)(hs)(h79)
bigram:	(h81 d)(h74 xs)(hs sx)(h79-71-62 xs)(hs sx)
trigram:	(h81 d xs)(h74 xs sx)(hs sx xs)(h79-71-62 xs sx)
4-gram:	(h81 d xs sx)(h74 xs sx xs)(hs sx xs sx)
5-gram:	(h81 d xs sx xs)(h74 xs sx xs sx)

図 4: N グラム抽象化の例 (音の高さの場合)

度は 0.400、それ以外の相対出現頻度は 0.200 となる。一方、重み w には表 5 に示すような 3 種類の重み付け方法を用意した。

3.4 N グラム特徴量の選択

表 4 に示されたように、抽象化により異なり N グラム (すなわち N グラム特徴量) の数は約半分に減少しているが、それでもまだ 1,000 のオーダーである。重回帰分析の性質上、説明変数となる N グラム特徴量の数は、目的変数である印象値データのサンプル数 (= 80) よりも 2 個以上 (3 個以上が推奨されている) 少なくなければならない [管 00]。そこで我々は、N グラム特徴量の数が多くても 77 個を超

表 4: N グラム抽象化による異なり N グラム数の変化 (音の高さの場合)

	抽象化前	抽象化後
unigram	2,177	2,177
bigram	6,120	3,346
trigram	10,640	4,507
4-gram	13,566	5,928
5-gram	14,848	7,393

表 6: スクリーニングによる N グラム特徴量数の変化 (音の高さの場合)

	処理前	処理後
unigram	2,177	289
bigram	3,346	436
trigram	4,507	513
4-gram	5,928	663
5-gram	7,393	825

表 5: 3 種類の重み付け方法

重みタイプ	楽曲基本特徴量	重み (算出方法)
w_1	全種類	1
w_2	音の長さ	1
	上記以外	\sum 音の長さ
w_3	音の長さ	1
	音の強さ	\sum 音の長さ
	上記以外	\sum 音の強さ ² \times \sum 音の長さ

えないよう、以下のような方法で N グラム特徴量の取捨選択を行った。

[相対出現頻度によるスクリーニング]

まず、各楽曲における N グラム特徴量の相対出現頻度がいずれの楽曲においても 0.01 未満であった N グラム特徴量を除外した。この操作により、N グラム特徴量の数は表 6 のように変化した。但し、この操作は音色情報に対しては行っていない。

[相関係数による N グラム特徴量の選択]

次に、N グラム特徴量と印象値データとの相関係数を求め、その絶対値が大きかった特徴量 (最大 77 個) を重回帰分析のための説明変数として選択した。このとき、N グラム特徴量の N 値の組み合わせとして、unigram のみ、bigram のみ、bigram と trigram、bigram から 4-gram まで、bigram から 5-gram まで、の 5 通りを用意したので、この 5 グループのそれぞれにおいて N グラム特徴量の選択を行った。

3.5 重回帰分析による N グラム特徴量と印象値データの対応付け

楽曲特徴量及び楽曲印象値計算式を決定するために、3.4 で選択された N グラム特徴量を説明変数、印象尺度 m ($m = 1, 2, \dots, 10$) における楽曲印象値 (SD 法に基づく印象評価実験の結果) を目的変数とする重回帰分析 (変数増加法) を行った。このとき、説明変数に用いる N グラム特徴量の N 値の組み合わせは、3.4 で述べたように、5 通りあり、重みタイプには w_1 、 w_2 、 w_3 の 3 種類を用いたので、結局、各印象尺度毎に 15 回の重回帰分析を行った。

3.6 楽曲特徴量及び楽曲印象値計算式の選択

さて、各印象尺度毎に 15 回の重回帰分析を行ったわけだが、その中で自由度修正済み決定係数 $R^{2'}$ *2 [管 00] が最も大きかった重回帰式を楽曲印象値計算式として採用し、その重回帰式を構成する説明変数 (N グラム特徴量) を楽曲特徴量と定義した。ここで、各印象尺度において $R^{2'}$ が最大となる N 値の組み合わせ及び重みタイプを、そのときの $R^{2'}$ とともに表 7 に示す。なお、表 7 は、 $N = 5$ の N グラム特徴量 (5-gram) が用いられなかったことを示しており、N グラム特徴量における N 値としては 4 までで十分なことを示唆している。

ここで、印象尺度 1 の場合を例に、設計された楽曲特徴量と楽曲印象値計算式の偏回帰係数及び定数

*2 サンプル数と説明変数の数との差が小さいと、決定係数は大きくなる傾向がある。この不具合を修正したのが自由度修正済み決定係数であり、次の式で計算される。

$$R^{2'} = 1 - \frac{S_e / (n - q - 1)}{S_{yy} / (n - 1)}$$

但し、 S_e : 残差平方和、 S_{yy} : 偏差平方和、 n : サンプル数、 q : 説明変数の数

表 7: 自由度修正済み決定係数 $R^{2'}$ が最大となる N 値の組み合わせと重みタイプ、及びそのときの自由度修正済み決定係数

印象尺度	N 値	重みタイプ	$R^{2'}$
1	2	w_1	0.852
2	2, 3	w_1	0.859
3	1	w_2	0.662
4	2, 3, 4	w_3	0.748
5	1	w_3	0.701
6	2, 3, 4	w_1	0.846
7	2	w_1	0.668
8	2, 3, 4	w_2	0.664
9	2, 3	w_1	0.836
10	2	w_3	0.714

表 8: 印象尺度 1 における楽曲特徴量と楽曲印象値計算式の偏回帰係数及び定数項

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

項を表 8 に示す。印象尺度 1 の場合の重みタイプは w_1 なので、楽曲から抽出される楽曲特徴量の相対出現頻度に重み 1 を掛けた値が楽曲印象値計算式（重回帰式）に代入され、その楽曲の印象尺度 1 における楽曲印象値が算出される。

4 システムの実装

我々は提案手法を Windows OS 上に実装し、楽曲印象値自動付与システムとした（Windows 98 Second Edition、NT、2000、XP 上で動作を確認し

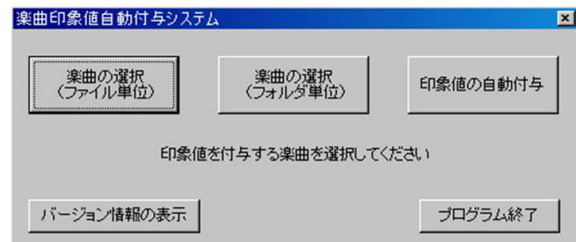


図 5: 楽曲印象値自動付与システム（メイン画面）

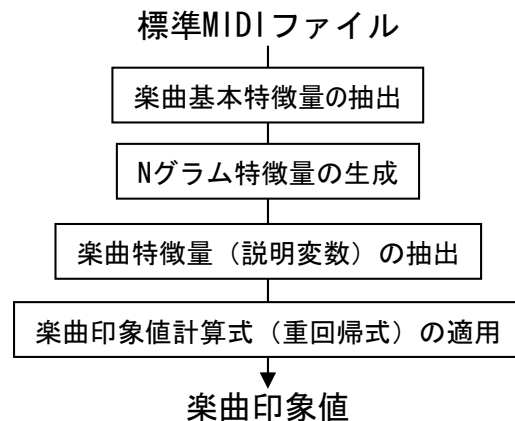


図 6: 楽曲印象値自動付与手法の処理手順

た)。本システムのメイン画面のスナップショットを図 5 に示す。図中、「楽曲の選択」ボタンは、楽曲印象値を付与したい楽曲をファイル単位もしくはフォルダ単位で選択するためのコマンドであり、このコマンドが実行されなかった場合は、デフォルトとして本システムが実装されているフォルダが選択される。「印象値の自動付与」ボタンは、選択された楽曲の楽曲印象値を計算し、既定のファイル (midi.iwt) に登録するためのコマンドである。このファイルは、csv (カンマ区切り) 形式のファイルであり、1 行 1 楽曲で、各行の第 1 要素に標準 MIDI ファイル名 (拡張子は含まない) 第 $m+1$ 要素に印象尺度 m に対する楽曲印象値という並びで登録される。

ここで、本システムに実装されている楽曲印象値自動付与手法の処理手順を図 6 に示す。入力には標準 MIDI ファイル形式 (フォーマット 0 または 1) の楽曲であり、各印象尺度毎に 1 個の楽曲印象値 (実数値) が出力される。処理手順中、「楽曲基本特徴量の抽出」と「N グラム特徴量の生成」は図 1 で示された処理と同じであり、標準 MIDI ファイルから N

グラム特徴量が生成される。この N グラム特徴量の中から楽曲特徴量が抽出され、その相対出現頻度に重み w を掛けた値が楽曲印象値計算式に代入されることによって楽曲印象値が算出される。但し、抽出される楽曲特徴量、用いられる重み w 、適用される楽曲印象値計算式は、それぞれ印象尺度毎に定義されており、それぞれの値・式を用いて、各印象尺度における楽曲印象値が算出される。

5 システムの公開

我々は、「印象に基づく楽曲検索」研究に必要なツールやデータ、資料の公開を進めており [熊本 01]、その一環として、今回実装した「楽曲印象値自動付与システム」も近日中に公開する予定である。公開場所は我々のホームページ (<http://www2.crl.go.jp/jt/a133/resource.html>) 上であり、システム利用に際しての主な条件は、(1) 原則として性能評価・比較のための学術的利用に限る、(2) ソースコードの解読及び改変を禁止する、(3) システムの第三者への配布を禁止する、の 3 点を予定している。なお、システムは実行ファイル形式 (*.exe) で配布され、そのファイルサイズは 256KB 程度になる。

6 まとめ

本稿では、標準 MIDI ファイル形式 (フォーマット 0 または 1) の楽曲 (特にクラシック系) を対象に、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手法を提案し、楽曲印象値自動付与システムとして Windows OS 上に実装した。提案手法は、標準 MIDI ファイルから楽曲特徴量 (N グラム統計量をベースに設計された特徴量) を抽出し、その相対出現頻度に重み w を掛けた値を楽曲印象値計算式に代入することによって、楽曲印象値 (10 対の印象尺度に対し各々 1 個の実数値) を算出する。提案手法が用いている楽曲特徴量及び楽曲印象値計算式は、SD 法に基づく印象評価実験 (被験者 100 名がクラシック 80 曲を聴取し、楽曲印象値を付与した) の結果に基づいて各印象尺度毎に設計されたものであり、被験者 100 名による楽曲印象値の平均値を目的変数、標準 MIDI ファイルから抽出可能な N グラム統計量ベースの特徴量を説明変数とする重回

帰分析の結果に基づいている。すなわち、決定された重回帰式が楽曲印象値計算式として採用され、その重回帰式を構成する説明変数が楽曲特徴量として定義されている。

今後の課題としては、提案手法の性能評価を行うとともに、現在構築している印象尺度ベースの楽曲検索システムを完成させたい。更には、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、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) における個人差を考慮した楽曲検索システムの開発にも取り組みたい。

参考文献

- [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).
- [佐藤 99] 佐藤 聡, 菊地幸平, 北上 始: 音楽データを対象としたイメージ検索のための感情価の自動生成, 情処研報, データベースシステム 118-8, 情報学基礎 54-8, pp. 57-64 (1999).
- [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).
- [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).
- [熊本 02] 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).
- [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).
- [菅 00] 菅 民郎: 多変量統計分析, 現代数学社, 京都 (2000).
- [熊本 01] 熊本忠彦, 太田 公子: 印象に基づく楽曲検索: 検索表現の収集と分析, 情処研報 (自然言語処理), 146-16, pp. 101-106 (2001).