

## 視点に応じた楽曲印象推移メタデータ生成方式と楽曲形式による実現

伊地智 麻子<sup>†1</sup> 清 木 康<sup>†2</sup>  
佐 藤 聡<sup>†3</sup> 中 神 康 裕<sup>†4</sup>

本稿では、ユーザの楽曲印象生成の注目点として「視点」を定義し、視点に応じた印象推移メタデータ生成方法を示す。ここで、視点とは、ユーザが楽曲の時間的印象推移において、どの部分に、どの程度重きを置いて聞かせるかを表す指標である。楽曲には、いろいろな印象が含まれており、ユーザの視点が定まることにより、楽曲の印象が決定される。本方法は、5つのステップから成り、各ステップには、複数のメソッドがある。本方法の特徴は、視点を決定することにより、全てのステップにおけるメソッドが決定する点にある。視点による印象推移メタデータ生成を実現することにより、ユーザの視点に合致した、印象推移メタデータ生成が実現できる。本稿は、視点に応じた、楽曲の印象推移メタデータを生成する方法を示す。さらに、視点として「楽曲形式」を適用した場合の実現方法を示す。また、実現方式に基づく印象推移メタデータ生成実験を行い、実現可能性を示す。

### A Metadata Generation Method Dealing with User's Viewpoints and Its Application to Musical Structures

ASAKO IJICHI,<sup>†1</sup> YASUSHI KIYOKI,<sup>†2</sup> AKIRA SATO<sup>†3</sup>  
and YASUHIRO NAKAGAMI<sup>†4</sup>

This paper presents a metadata generation method dealing with user's Viewpoints and its application to musical structures. In this paper, we use the term "Viewpoint" as the focussing point in listening to music. This method is used for music-retrieval-by-impression according to user's Viewpoints. In general, music included various impression and its impression is fix only when the user's Viewpoint is specified. The main feature of this method is that Viewpoint is used for fixing impression of music metadata related to various user's Viewpoints. This paper shows several experimental results of the metadata creation to clarify the feasibility and effectiveness of our method.

#### 1. はじめに

近年、MPEG4 などの楽曲データ高品質圧縮や、WWW の発展等により、多くの楽曲が広域ネットワーク上に置かれるようになった。これに伴い、広域ネットワークを介して楽曲を入手可能となってきた。アクセス可能な楽曲データ群の中から、要求に応じて、適切な楽曲を検索するシステムの実現は有効であり、重要な研究課題となっている。

楽曲検索のうち、既知曲に関しては、作曲者、題名といった属性情報によりパターンマッチングが可能である。しかし、膨大な量の楽曲データ群のうち、未知曲は多く存在する。既知曲も未知曲も検索可能な方法として、楽曲の印象を問い合わせとした検索は重要な研究課題である。また、楽曲データのデータベース化において、適切な印象メタデータの生成は、適切な検索結果を得るために重要な要素となる。しかし、連続メディアデータである楽曲データの内容、表現、印象、および、その推移を対象とした検索環境の実現は確立されていない。

これまで、楽曲の内容に関する印象メタデータ生成方法として、我々は、発想標語を用いたクラシック音楽メタデータ生成方式 [5] を提案した。また、音楽データを対象とした印象メタデータ生成方法としては、音楽のメロディパターンから印象メタデータを自動抽出し、その印象メタデータに対して間接的に検索する、楽曲メタデータ自動生成方式 [6,8] を提案した。楽曲

†1 慶應義塾大学大学院政策・メディア研究科  
Graduate School of Media and Governance, Keio University  
†2 慶應義塾大学環境情報学部  
Faculty of Environmental Information, Keio University  
†3 筑波大学学術情報処理センター  
Science Information Processing Center, University of Tsukuba  
†4 慶應義塾大学 SFC 研究所  
Keio Research Institute at SFC, Keio University

メタデータ自動生成方式は、楽曲を最小単位として、その楽曲に対応する印象メタデータ生成を実現した。楽曲が最小単位 (楽曲が始まってから終わるまでがアトミック) であるため、楽曲全体の総合的な印象メタデータを生成する方式であり、単一曲内での時間的変化に応じた印象抽出は対象としなかった。時間的変化に応じた印象メタデータ抽出方式としては、音楽データの印象の時間的推移を扱う印象メタデータ自動生成方式 [4] を提案した。この方式は、楽曲の印象の時間的変化を対象とした印象抽出の枠組みを示している。また、統計値を用いた研究として、感情価を対象とした方法 [7] が示されている。

本稿では、ユーザの楽曲印象生成の注目点として「視点」を定義し、視点に応じた印象推移メタデータ生成方法を示す。ここで、視点とは、ユーザが楽曲の時間的印象推移において、どの部分に、どの程度重きを置いて聞かせる指標である。楽曲には、いろいろな印象が含まれており、ユーザの視点が定まることにより、楽曲の印象が決定される。本方法は、5つのステップから成り、各ステップには、複数のメソッドがある。本方法の特徴は、視点を決定することにより、全てのステップにおけるメソッドが決定する点にある。視点による印象推移メタデータ生成を実現することにより、ユーザの視点に合致した、印象推移メタデータの生成が実現できる。

本稿は、視点に応じた、楽曲の印象推移メタデータを生成する方法を示す。さらに、視点として「楽曲形式」を適用した場合の実現方法を示す。また、実現方式に基づく印象推移メタデータ生成実験を行い、実現可能性を示す。

## 2. 視点に応じた楽曲データの印象推移メタデータ生成方法

本節では、視点に応じた楽曲印象推移メタデータ生成方法を示す。

本方法は、5つのステップから成り、各ステップには、複数のメソッドがある。ユーザの視点が定まることにより、全てのステップにおけるメソッドが決定する。視点による印象推移メタデータ生成を実現することにより、ユーザの視点に合致した、印象推移メタデータ生成が実現できる。

5ステップによる印象推移メタデータ抽出方式の枠組みは、音楽データの時間的推移を扱う印象メタデータ自動生成方式 [4] にもとづいている。この方式は楽曲を時間軸上でセクション (時間軸に沿って区切られた部分) に区切り、その各セクション毎の印象をもと

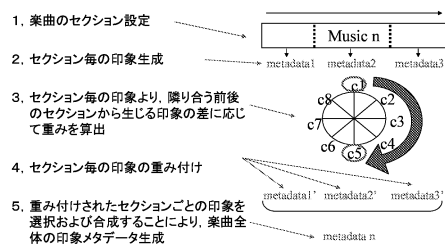


図 1 印象推移メタデータ生成の 5 ステップ

に、楽曲全体の総合的な印象メタデータを生成する方式である。各ステップによる処理の流れを図 1 に示す。

- Step1** 楽曲のセクション設定
- Step2** セクション毎の印象生成
- Step3** 隣り合う前後のセクションから生じる印象の差に応じて重みを算出
- Step4** セクション毎の印象の重み付け
- Step5** 選択および合成し、楽曲全体の総合的な印象メタデータを生成

### 2.1 視点の設定 (適用メソッドの設定)

視点の設定は、提案方式の 5つのステップにおける適用メソッドを設定し、印象推移メタデータ生成に反映させる。視点  $V_y (y = 1, z)$  ( $y$ : 視点の識別子,  $z$ : 視点数の最大値) から、各 Step における視点に対応する以下のメソッドを設定する。メソッドの詳細は、各ステップにおいて述べられている。

- Step1 のセクション設定ポリシー  $P_f (f = 1, n)$
- Step2 の印象生成方式  $A_h (h = 1, q)$
- Step3 の印象の差算出方法  $C_{[h,e]} (i = 1, t)$
- Step4 の印象重み付け方法  $W_j (j = 1, q)$
- Step5 の印象合成、選択方法  $X_k (k = 1, r)$

視点の設定方法として、以下の式 (1) を定義する。

$$f_{viewpoint}(V_y) \rightarrow \{P_f, A_h, C_{[h,e]}, W_j, X_k\} \quad (1)$$

### 2.2 Step1: 楽曲のセクション設定

Step1 は、楽曲  $M_d (d = 1, l)$  ( $d$ : 楽曲の識別子,  $l$ : 楽曲数の最大値) を、時間軸上で意味のある長さ毎のセクション  $S_{[d,j]} (j = 1, m)$  ( $j$ : セクションの識別子,  $m$ : セクション数の最大値) に区切る。セクション設定ポリシー (セクションの境界を決定する要素) を  $P_f (f = 1, n)$  ( $f$ : セクション設定ポリシー識別子,  $n$ : セクション設定ポリシー数の最大値) とし、以下の式 (2) を定義する。視点は、セクション設定ポリシー  $P_f$  に対応する。

$$f_{divide}(M_d, P_f) \rightarrow \{S_{[d,1]}, \dots, S_{[d,m]}\} \quad (2)$$

### 2.3 Step2: セクション毎の印象生成

Step2 は、セクション毎に区切られたセクション  $S_{[d,j]}$  から、セクション毎の印象  $I_{[d,p]}(p = 1, m)$  ( $p$ : 印象の識別子,  $h$ : 印象生成方式の識別子,  $q$ : 印象生成方式数の最大値) を生成する。印象を生成する方式を  $A_h(h = 1, q)$  とし、以下の式 (3) で定義する。視点は印象を生成する方式  $A_h$  に対応する。

$$f_{impression}(S_{[d,1]}, \dots, S_{[d,m]}, A_h) \rightarrow \{I_{[d,1]}, \dots, I_{[d,m]}\} \quad (3)$$

### 2.4 Step3: セクション毎の印象より、隣り合う前後のセクションから生じる印象の差に応じて重みを算出

Step3 は、セクション毎の印象  $I$  から、隣り合う前後のセクションから生じる印象の差に応じて重み  $D_{[d,s]}(s = 1, m - 1)$  ( $s$ : 印象の差の識別子) を算出する。Step-2 で生成された、隣り合う前後のセクションの印象の意味的な距離 (意味的な近さ) を計量する。隣り合う前後のセクションから生じる印象の差に応じた重みの算出メソッドを  $C_{[h,i]}(i = 1, t)$  ( $i$ : 印象の差の算出メソッドの識別子,  $t$ : 印象の差の算出メソッド数の最大値) として、以下の式 (4) を定義する。視点は隣り合う前後のセクションから生じる印象の差に応じた重みの算出メソッド  $C_{[h,i]}$  に対応する。

$$f_{degree}(I_{[d,1]}, \dots, I_{[d,m]}, A_h, C_{[h,i]}) \rightarrow \{D_{[d,1]}, \dots, D_{[d,m-1]}\} \quad (4)$$

### 2.5 Step4: セクション毎の印象の重み付け

Step4 は、Step3 で算出された、隣り合う前後のセクションから生じる印象の差に応じた重み  $D$  をセクション毎の重みとして印象へ反映させ、重み付き印象  $I'_{[d,p]}(p = 1, m)$  ( $p$ : 重み付き印象の識別子) を生成する。セクション毎の重み付け方法を  $W_j(j = 1, q)$  ( $j$ : 重み付け方法の識別子,  $q$ : 重み付け方法数の最大値) とし、以下の式 (5) を定義する。視点は重み付け方法  $W_j$  に対応する。

$$f_{weight}(I_{[d,1]}, \dots, I_{[d,m]}, D_{[d,1]}, \dots, D_{[d,m-1]}, W_j) \rightarrow \{I'_{[d,1]}, \dots, I'_{[d,m]}\} \quad (5)$$

### 2.6 Step5: 重み付けされたセクション毎の印象より、選択および合成し、楽曲全体の総合的な印象メタデータを生成

Step5 は、重み付けされたセクション毎の印象  $I'$  から、選択および合成し、楽曲全体の総合的な印象メタ

データ  $I_d(d = 1, m)$  を生成する。印象選択、合成方法を  $X_k(k = 1, r)$  ( $k$ : 印象選択、合成方法の識別子,  $r$ : 印象の選択、合成方法数の最大値) とし、以下の式 (6) を定義する。視点は印象選択、合成方法  $X_k$  に対応する。

$$f_{aggregate}(I'_{[d,1]}, \dots, I'_{[d,m]}, X_k) \rightarrow \{I_d\} \quad (6)$$

## 3. 視点に「楽曲形式」を適用した場合の印象推移メタデータ生成実現方法

2章に示した印象推移メタデータ生成方法にもとづき、視点に「楽曲形式」を適用した場合の印象推移メタデータ生成実現方法を示す。

### 3.1 視点「楽曲形式」の実現方法

本方法は、楽曲形式という視点による適用メソッドを、2章で示した Step1 から Step5 に適用することにより実現する。視点「楽曲形式」の場合の適用メソッドを、表 1 に示すように設定する。Step には、ステップ名を示し、適用メソッド ID には、提案方法で示した視点に対応するメソッドを ID とした。適用メソッド名には、具体的に適用されるメソッドを示した。

表 1 視点「楽曲形式」の設定 (適用メソッドの設定)

Step	適用メソッド ID	適用メソッド名
Step1	セクション設定方法 $\{P_f\}$	楽曲形式による設定メソッド
Step2	印象生成方式 $\{A_h\}$	楽曲メタデータ自動生成方式
Step3	印象の差算出方法 $\{C_i\}$	Hevner 形容詞群による算出メソッド
Step4	印象重み付け方法 $\{W_j\}$	全ての印象の差を重みとするメソッド
Step5	印象合成、選択方法 $\{X_k\}$	最大値を選択するメソッド

構造を楽曲形式として示している楽曲の多くは、クラシック楽曲である。そのため本実現においては、楽曲形式という視点は、クラシック楽曲という要素を含んでいる。これまで、[4] における、クラシック楽曲を対象とした印象メタデータ生成実験結果と、クラシック楽曲および楽曲形式の特性を考慮した知識として、適用メソッドを設定した。

Step3 に適用される、Hevner 音楽形容詞群は、図 2 に示すように Hevner が心理学実験をもとに、音楽を表現するために必要な形容詞セットとして 66 語を設定し、さらにそれらを意味的に 8 形容詞群  $c_1, c_2, \dots, c_8$  へ分類した形容詞群である。各形容詞群は、6-11 語から成り立つ。8 つの形容詞群は、環状に位置し、それぞれ円上での距離が近いほど意味的に近いと定義されている。さらに、円の対称にあるグループ同士 (例えば、 $c_1$  と  $c_5$ ) は、逆の印象であると定義されている。例として、 $c_1$  を中心に考えると、 $c_2$   $c_3$   $c_4$  の順に意味的に遠くなり、同様に  $c_8$   $c_7$   $c_6$  の順に意

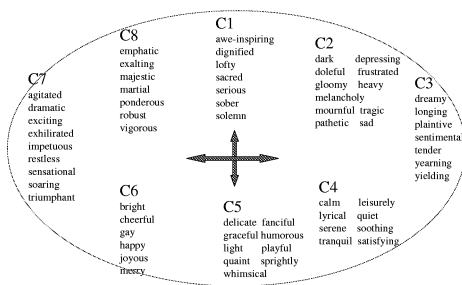


図 2 Hevner 音楽形容詞群

味的に遠くなり、c5 が意味的に c1 の逆となる。

Hevner は心理学者であり、全ての形容詞から、音楽を顕著に表現し効果的に特徴づける形容詞を示している。これは被験者にクラシック楽曲を聞かせ音楽を表現するのにふさわしい語を選択する方式で行われ、最終的に 66 語を選択し、その 66 語は 8 つの群を形成し、群に含まれる形容詞同士は、互いに近い表現で使われていることを示したものである。

詳しくは [1,2,3] に述べられている。

### 3.2 楽曲形式の特性

本実現では、楽曲形式を、視点の 1 つとして、印象推移メタデータ生成を行う。多種の視点のうち、特にクラシックに関しては、楽曲形式が、楽曲理解や楽曲の聴き方として、確立された知識とされ、客観的な指標の 1 つである。具体的には、楽曲を時間軸に沿った意味のあるまとまりであるセクション設定を行い、それらが時間的に、どのような順で、どの程度の頻度で出現するかを分類した知識としての役割が大きい。その分類は、例えば「ソナタ形式」「ロンド形式」「三部形式」などと呼ばれているものである。セクションは A, B などアルファベットで表現されることが多く、形式毎に、その推移として「ABA」などと表される。

本実現では、楽曲形式のうち、セクション設定が形式において一定の構造である「一部形式、二部形式、三部形式」の 3 形式を用いる。ここで一定の構造としているのは、形式が定まることにより、セクションの構造が一意に決定する形式を指している。一部形式は「A」、二部形式は「AB」、三部形式は「ABA」と表現できる。

視点は個人の習慣的なものから、楽曲形式のように知識として、教育によって得られるものまで多様である。視点に沿って聴くこと、つまり本実現の場合は、楽曲形式に沿って聴くことがユーザにとって必須なのではない。ユーザの視点の 1 つとして、ここでは、楽

曲形式による実現方法を示している。

## 4. 視点に「楽曲形式」を適用した、印象推移メタデータ生成実験

視点に「楽曲形式」を適用した、印象推移メタデータ生成実現方法による、印象推移メタデータ生成実験を示す。実験に用いた楽曲と、その楽曲の楽曲形式の関係を表 2 に示す。Music ID は、楽曲の識別 ID を示し、形式には、楽曲の形式を示している。さらに、各楽曲形式のセクションについて、Section1,2,3 に示す。

表 2 楽曲とその楽曲の楽曲形式

Music ID	楽曲形式名	Section1	Section2	Section3
1	一部形式	A	null	null
2	二部形式	A	B	null
3	三部形式	A	B	A

対象楽曲に対し、印象推移メタデータ生成方法を適用した結果を表 3 に示す。Music ID は、楽曲の識別 ID を示し、Section1,2,3 にはそれぞれの印象と、括弧内はその印象の重みを示す。重みは、小数点以下第一位までをデータとして用いている。各セクションの印象は、3 章で示した Hevner 形容詞群の群名 (c<sub>i</sub>) として表す。

参考として、楽曲形式という視点で聴いた場合のユーザと、特定の視点を与えずに聴いたユーザの印象を示す。

楽曲形式という視点で聴いたユーザには、対象楽曲の楽曲形式をあらかじめ提示し、対象楽曲を複数回聴かせた後、印象とその強さを、推移に沿って記述する方法で行った。楽曲形式という視点で聴いたユーザの各楽曲の印象推移を、表 4 に示す。Music ID は、楽曲の識別 ID を示し、Section1,2,3 にはそれぞれの印象と、括弧内はその印象の重みを示す。

特定の視点を与えずに聴いたユーザも同様に、対象楽曲を聴かせ、印象とその強さを、記述する方法で行っ

表 3 楽曲の印象推移メタデータ (Step1 から Step4 の適用結果)

Music ID	Section1	Section2	Section3
1	C6(0.7)	null	null
2	C3(0.3)	C6(0.4)	null
3	C2(0.4)	C6(0.6)	C3(0.8)

表 4 楽曲形式の視点で聴いたユーザの印象推移

Music ID	Section1	Section2	Section3
1	C6(0.5)	null	null
2	C3(0.3)	C6(0.5)	null
3	C2(0.2)	C6(0.3)	C3(0.6)

表 5 視点を与えずに聴いた場合のユーザの印象推移

Music ID	Section1	Section2	Section3	Section4
1	C6(0.3)	C5(0.3)	C6(0.4)	null
2	C3(0.2)	C3(0.3)	C5(0.4)	C6(0.5)
3	C2(0.2)	C3(0.5)	null	null

た。この場合は、あらかじめ対象楽曲の楽曲形式、および、セクション数は提示せず、このユーザの印象によるセクション数の設定である。視点を与えずに聴いたユーザの印象推移を、表 5 に示す。Music ID は、楽曲の識別 ID を示し、Section1,2,3,4 にはそれぞれの印象と、括弧内はその印象の重みを示す。

実験により、視点に「楽曲形式」を適用した印象推移メタデータ生成を実現を示した。一部形式では、セクションは1つであるため、印象推移メタデータは1つである。二部形式、三部形式では、セクションが複数あるため、印象の差、重み付けを反映した印象推移メタデータを生成した。

比較の参考として、提案方法による印象推移メタデータは、「楽曲形式」という視点で聴いたユーザの印象推移に、近い結果が得られたことが分かる。ユーザは、特定の視点によって楽曲を聴く場合に、その視点としての適切なセクション設定から、セクション毎の印象を生成し、これによって楽曲印象推移を生成していると考えられる。この結果は、本方法が、ユーザの視点と合致した印象メタデータ生成方法として、提案方法の実現可能性を示している。

また、提案方法による印象推移メタデータは、特定の視点を与えずに聴いたユーザの印象推移とは、異なる結果が得られたことが分かる。特に、セクションの数が一致していない点で、印象生成の相違を顕著に示している。これは、特定の視点を与えず楽曲を聴く場合には、ユーザは、独自の視点によって聴くが、その視点が示されていない場合には、システムとして、そのユーザと合致する印象推移メタデータを生成することが、難しい事を示している。つまり、ユーザに合致する印象推移メタデータを生成するためには、ユーザの視点と、メタデータ生成方法の視点を一致させることが重要であることを示している。

実験により、本稿で示した印象推移メタデータ生成方法の実現可能性を示した。より多くの対象楽曲と、対象楽曲形式による有効性検証は今後の課題とする。

## 5. おわりに

本稿では、ユーザの楽曲印象生成の注目点として視点を定義した。本方法の 5 つのステップにおいて、視

点に対応するメソッドを設定することにより、視点に応じた印象推移メタデータ生成方法を示した。さらに、視点として「楽曲形式」を適用した場合の実現方法を示し、実現方法に基づく印象推移メタデータ生成実験を示し、実現可能性を示した。

提案方法は、さまざまな時間的変化をするメディアに適用可能な方法であり、メディアデータに依存してメソッドを変更することにより、他連続メディアデータ(映像、文章)に応用することが考えられる。さらに、時間軸以外の要素、例えば同時に演奏している複数の楽器に対する視点と、その視点に応じた印象推移メタデータ生成方法が考えられ、これらを今後の研究の課題とする。

## 謝 辞

本論文の方式設計、実験にあたり、貴重な御助言を頂いた石橋直樹氏(慶應義塾大学 政策・メディア研究科)に感謝致します。また、本研究の一部は、日本学術振興会 学術創成研究プロジェクト「人文社会科学と自然科学を連携するメタレベル知識ベースシステムの開発」によるものである。ここに記して謝意を表します。

## 参 考 文 献

- 1) Hevner, K.: "Expression in Music: a Discussion of Experimental Studies and Theories," *Psychological Review*, Vol.42, pp.186-204 (1935).
- 2) Hevner, K.: "Experimental Studies of the Elements of Expression in Music," *American Journal of Psychology*, Vol.48, pp.246-268 (1936).
- 3) Hevner, K.: "The Affective Value of Pitch and Tempo in Music," *American Journal of Psychology*, Vol.49, pp.621-639 (1937).
- 4) 伊地智 麻子, 石橋 直樹, 清木 康: "音楽データの印象の変化を扱うメタデータ自動生成方式," 情報処理学会 データベースと Web 情報システムに関するシンポジウム (DBWeb2001) 論文集, pp.241-248(2001)
- 5) 伊地智 麻子, 清木 康: "発想標語を用いたクラシック音楽メタデータ生成による意味的連想検索方式," 第 12 回データ工学ワークショップ (DEWS2001) 論文集, 4B-4 (2001).
- 6) Kitagawa, T. and Kiyoki, Y.: Fundamental Framework for Media Data Retrieval Systems Using Media-lexico Transformation Operator - In the Case of Musical MIDI Data, *Proceedings of THE TENTH EUROPEAN-JAPANESE CONFERENCE ON INFORMATION MODELING AND KNOWLEDGE BASES* (2000).

- 7) 小川 潤, 佐藤 聡, 北上 始: “感性に基づく音楽作品のための類似度計算方式,” 情報処理学会 データベースと Web 情報システムに関するシンポジウム (DBWeb2000) 論文集, pp.229-234 (2000).
- 8) 吉野 太智, 高木 秀幸, 清木 康, 北川 高嗣: “楽曲データを対象としたメタデータ自動生成方式とその意味的連想検索への適用,” データベースシステム, No.7 , pp.109-116 (1998).