時系列印象変化抽出機構を有する音楽検索データベースシステム

市瀬 綾 倉林 修一 清木 康

慶應義塾大学環境情報学部 〒252-0882 神奈川県藤沢市遠藤 5322

E-mail: {t08099ai, kurabaya, kiyoki}@sfc.keio.ac.jp

あらまし 本稿では、時系列メディアデータである音楽を対象として、楽曲中における時系列に沿った印象変化を分析・検索する音楽データベースシステムを提案する。本システムは、時系列メディアデータの構造分析による部分的時区間の選択を行い、選択した部分的時区間別に感性特徴を分析することにより、印象的な部分的時区間別の印象特徴量を用いた時系列メディアデータのメタデータ生成と検索を実現する。本システムの実現により、利用者は、印象語をクエリとして、それぞれの楽曲において特徴的な印象変化に応じた音楽獲得が可能となる。本稿では、楽曲内における、繰り返し構造分析による印象的なセクション(サビ構造、及び、繰り返し構造)の自動抽出メカニズムを実装したプロトタイプシステムを用いた実験結果を示し、本システムの実現可能性を示す。

キーワード マルチメディア,音楽,検索,個人化

A Music Search Database System with a Selector for Impressive-Sections of Continuous Data

Aya ICHINOSE Shuichi KURABAYASHI and Yasushi KIYOKI

Faculty of Environment and Information Studies, Keio University 5322 Endo, Fujisawa, Kanagawa, 252-0882 Japan E-mail: {t08099ai, kurabaya, kiyoki}@sfc.keio.ac.jp

Abstract This paper presents a music search system that retrieves music by extracting impressive-sections along a timeline. This system realizes an automatic time-duration selector that detects repetitions and bridges by analyzing physical and structural music features, such as pitch and tonality. This system enables users to submit emotive keywords as a query for retrieving music according to the impressive changes in each music piece. This paper shows a prototype system that searches MIDI music files by analyzing them automatically. This paper also shows several experimental results for clarifying the feasibility of the system.

Keyword Multimedia, Music, Search, Personalization

1. はじめに

近年, 音楽メディアのデジタル化とデータベース化 に伴い, 音楽内容の分析による音楽獲得の可能性が高 まっている. 利用者は、PC や携帯端末を用いて、音楽 データを管理・携帯し、様々な場面で鑑賞するように なってきている. 例えば, 我が国において, 社団法人 日本レコード協会の調査[9]によれば、音楽を試聴する 際に利用している機器として, PC が 57.7%, 主要スマ ートフォンが 23.4%のシェアを占めており, 利用者は 平均してこれらの機器に 586.3 曲を保存・管理してい る. 個人が膨大な数の音楽データを保有する状況にお いて, 音楽データベースの利用者は必要とする楽曲を 直ちに見つけ出すことが困難な状況にある [1]. 音楽 のような時系列メディアデータ特有の問題として, 利 用者は、音楽データの内容を把握するためには実際に 試聴する必要があり、ファイル名やアーティスト名な どのシンプルなメタデータによるデータ管理では,デ

一タベース中から目的の音楽データを選択することは容易ではない. 大規模流通環境において, 音楽を提供するためには, 利用者に未知の楽曲を提っし, 新たな音楽の発見の機会を提供する事が重要要である[2]. これまでに,音楽の内容に応じた未知の検索手法として, コード進行を利用した楽曲類似のが大きにないる. 音楽のように, 個人の感性的嗜実においる. 音楽のように, 個人の感性的嗜実であるには,楽曲を対象とした検索・推薦において,音楽データが鑑賞者に与える印象を分析し, コンテスのよりには,楽曲を対象とした検索・推薦においテンスを対したは,楽曲を対象とした検索・推薦においテンスを対したは,楽曲を対象を分析し, コンテスを対した。

本稿では、時系列印象変化抽出機構を有する音楽検索データベースシステムを提案する。本システムは、時系列メディアデータである音楽を対象として、楽曲内における繰り返し構造分析による印象的なセクション(サビ構造、及び、繰り返し構造)の自動

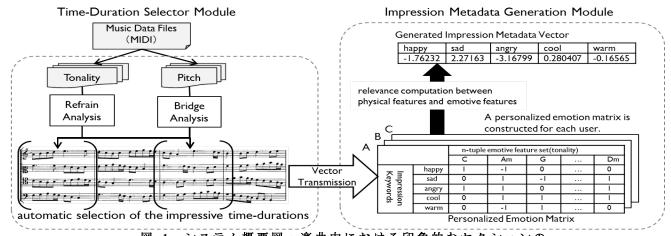


図 1 システム概要図:楽曲内における印象的なセクションの 自動抽出メカニズムを適用し、当該セクションについての感性的特徴量を生成

抽出メカニズムを有している. 本システムでは, 楽曲の印象的なセクションとは, 楽曲内において, 鑑賞者に特に強い印象を与え, かつ, 時間幅をもつ楽曲部分であると定義する.

本システムは、音楽データが鑑賞者に与える印象を 分析し, コンテンツと利用者個人の感性的嗜好の適合 性を評価するために, 物理的特徴量と感性的特徴量の 関係を定義した印象変換マトリクスを用いる. 印象変 換マトリクスは、物理的特徴量(調性)と感性的特徴 量 (印象語) の相関を, -1~1 の値で定義したもので ある. 本システムは、楽曲の印象的なセクションとし て、繰り返し構造を有する時区間、および、"サビ"と 呼ばれる時区間を抽出する. 本システムは, 抽出した それぞれの時区間に印象メタデータを生成し、検索時 に利用者の視点に応じて,適切な時区間のメタデータ を選択し、相関量計量を行う. 相関量計量には、選択 した時区間の物理的特徴量(調性)と,感性的特徴量 (印象語) との相関量を、印象変換マトリクスに定義 された二属性間の関連性定義を用いる. このように, 利用者の視点に応じて時区間を選択することにより, 利用者が楽曲に抱く印象を的確に反映した印象分析を 実現する (図 2).

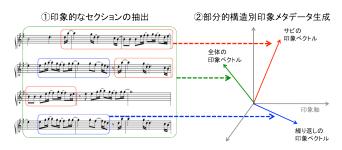


図 2 部分的構造別に印象メタデータを生成

各利用者固有の感性的特性に応じた印象分析を実現するため、本システムは、この印象変換マトリクスを個人別に構築する. この個人別印象変換マトリクス

は、システムが自動的に抽出した楽曲内の印象的なセクションを対象として、セクション内の物理的特徴量 (調性)と利用者個人の感性的特徴量(印象語)の相関を学習するため、利用者が楽曲に抱く印象を的確に 反映した個人化を行うことができる.

本システムの主要な機能は、次の4機能である.

- 1. 印象時区間選択機能:音楽データを対象とし,物理的特徴量,セクション構造の抽出による印象的なセクションの選択
- 2. 印象メタデータ生成機能:印象的なセクションを 対象とし、個人別の印象変換マトリクスを用いた 印象メタデータの生成
- 3. 個人化機能:セクションの印象修正による,音楽 要素と利用者個人の感性的嗜好の関連性の抽出, 印象変換マトリクスの個人化
- 4. 検索機能:印象的なセクション別のメタデータ検索

本システムを音楽データベースに適用することにより、利用者は、キーワード以外の手段として、音楽コンテンツそのものを用いたコンテンツの特徴抽出、及び、個人化された音楽獲得を利用することができるようになる. また、音楽の中身を聴くことなく、音楽の印象変化を視覚的に把握し、個人化を適用することが可能となる.

2. 時系列印象変化抽出機構による音楽検索 データベースシステム

本節では、時系列印象変化抽出機構による音楽獲得 データベースシステムのシステム概要,データ構造, 及び,基本機能を示す.

2.1. システム概要

本システムは、1) 印象時区間選択、2) 印象メタ データ生成、の2つの機構から構成される. 図 1に本 システムの概要図を示す.

本システムは、楽曲 MIDI データを対象として、物 理的特徴量として、ピッチ、および、調性を抽出する. 調性とは、音符の並びから構成される音楽の構造であ り,12の長調と12の短調の,24種類からなる[5].異 なる調性は,異なる印象を聞き手に与えるため,調性 を分析することが, 音楽の感性的分析において重要で ある. 更に, これら2つの物理的特徴量から, セクシ ョン構造として,繰り返し構造,サビ構造を抽出する. 本時系列印象変化抽出機構は, 音楽内容における物理 的特徴量の傾向から、印象的なセクションの選択、及 び, 印象メタデータの生成を行い, 音楽データにおけ る印象の時間的推移を自動的に分析する. 本個人化機 能は分析結果を対象として学習アルゴリズムを適用し, 物理的特徴量と感性的特徴量との関係性を,利用者個 人ごとに計量する. 検索機能は, 個人別に生成された 印象メタデータを用いて計量された印象メタデータを 対象として、音楽データの検索を行う.

2.2. データ構造

本節では、本システムのデータ構造を述べる.

2.2.1. 楽曲 MIDI データ

本システムの分析対象とするデータとして、楽曲データである MIDI データの基本構造を示す。MIDI は、鍵盤楽器操作の自動化を主な目的として開発されたデータ構造であり、鍵盤の操作情報(打鍵、リリース)を時系列に沿って格納している。本システムにおけるMIDI データを、Mp とすると、Mp を次のように定義できる。

$$Mp := \{n_1(t, p, d), n_2, \dots, n_k\}$$

2.2.2. ピッチマトリクス (PMAT)

本システムは、物理的特徴量として、音情報の時系列変化を行列として表す.この行列をピッチマトリクス(PMAT)と呼ぶ.MIDIでは、 $0\sim127$ のピッチの値を定義している為、PMAT は m 行 $\times128$ 列の行列である.次のように PMAT: P を定義する.

$$P := \begin{pmatrix} c_{[0,0]} & \cdots & c_{[0,127]} \\ \vdots & \ddots & \vdots \\ c_{[m,0]} & \cdots & c_{[m,127]} \end{pmatrix}$$

ここで、 $\mathbf{c}_{[i,j]}$ は \mathbf{i} 番目の時間における \mathbf{j} 番目のピッチ情報の有無を意味し、 \mathbf{m} は、楽曲の演奏時間長に対応する.本システムは、 $\mathbf{c}_{[i,j]}$ の値として、ピッチ情報があ

る場合は1を、ない場合には0を設定する.

2.2.3. 調性マトリクス (TMAT)

本システムは、物理的特徴量として、調性の時系列変化を行列として表す.この行列を調性マトリクス (TMAT) と呼ぶ.調性の種類が 24 種類に定義されている為、TMAT は m 行×24 列の行列である.次のように TMAT: T を定義する.

$$T := \begin{pmatrix} v_{[0,0]} & \cdots & v_{[0,24]} \\ \vdots & \ddots & \vdots \\ v_{[m,0]} & \cdots & v_{[m,24]} \end{pmatrix}$$

ここで、 $v_{[i,j]}$ は j 番目の階調の i 番目の時間軸間に対応する相関値を意味し、m は、楽曲の演奏時間長に対応する。例えば、8 小節毎に調性メタデータを生成した場合、m は、全小節数/8 となる。

2.2.4. 繰り返し・サビ構造

本節では、楽曲の印象的なセクションの自動抽出メカニズムにおける、セクション抽出に用いるデータ構造を示す.本システムは、セクション構造として、繰り返し構造とサビ構造の二種類を分析する.繰り返し構造、サビ構造は順序集合である.次のように繰り返し構造:R、サビ構造:Sを定義する.

$$R := \left\{ r_1(b, e, u), r_2, \dots, r_y \right\}$$
$$S := \left\{ s_1(b, e), s_2, \dots, s_l \right\}$$

ここで、 r_y は、y 個目の繰り返し構造を意味し、属性として、開始時間(b)、終了時間(e)、繰り返しの種類識別子(u)を有する. また、 s_1 は、l 個目のサビを意味し、属性として、開始時間(b)、終了時間(e)を有する.

2.2.5. 印象変換マトリクス(EMAT)

本節では、楽曲の印象的なセクションの自動抽出メカニズムにおける、感性的特徴量の計量に用いるデータ構造を示す。本システムは、物理的特徴量と感性的特徴量との関係性を行列として表す。この行列を印象変換マトリクス(EMAT)と呼ぶ。EMAT は、印象メタデータ生成、個人化、検索機能において、楽曲と個人の感性的嗜好の適合性の評価を実現する。EMAT は n 行×m 列の行列である。次のように EMAT: E を定義する。

$$E := \begin{pmatrix} r_{[0,0]} & \cdots & r_{[0,m]} \\ \vdots & \ddots & \vdots \\ r_{[n,0]} & \cdots & r_{[n,m]} \end{pmatrix}$$

ここで、 $r_{[i,j]}$ は i 番目の印象語と j 番目の物理的特徴量の相関値を意味する. n は定義された印象語の数に対

応する. また, m は, 物理的特徴量の属性数に対応する.

2.3. 印象時区間選択機能

本節では、MIDIデータの物理的特徴量、及び、セクション構造を分析する方法について説明する. 本システムは、次の手順を実行する.

● Step-1: [PMAT 分析]

本システムは、MIDI データを対象から、音情報の時系列変化を表す行列 (PMAT) を生成する.本システムは、MIDI データの打鍵番号、及び、打鍵信号、リリース信号を対象として、時系列に沿って行列構造にピッチの有無を 0 か 1 の値で格納し、PMAT を生成する.

● Step-2:[TMAT分析]

Step-2 では、Step-1 により分析した PMAT を対象として、ピッチ情報から調性を推測する Krumhansl-Schmuckler 調性自動判定アルゴリズム [3]を適用し、調性の時系列変化を表す行列(TMAT)を生成する. 本システムは、Krumhansl-Schmuckler アルゴリズムを用い、各調性のプロファイルを統計により算出し、各調性のプロファイルと対象とする時区間のピッチのヒストグラムとの相関値を格納した TMAT を生成する.

● Step-3: [繰り返し構造分析]

Step-3では、Step-2において分析したTMATを対象として、繰り返し構造分析アルゴリズムを適用する。本アルゴリズムは、TMATから調性の列を生成し、生成された調性の列を対象として、調性の変化を抽出する為に、前後の調性と同値の時区間を削除し、調性の列の長さを短縮する。続いて、短縮された調性の列を対象として、図3の処理を再帰的に行い、最長一致のパターンを検出する。短縮前の調性の列を対象として、検出された最長一致のパターンを用い、類似する調性の列の時区間を同じ繰り返し構造として定義する。

● Step-4:[サビ構造分析]

Step-4 では、Step-1 において分析した PMAT を対象として、サビ構造分析アルゴリズムを適用する. 現代のポップミュージックのサビの時区間には、高いピッチが多く含まれていることが知られている. 本アルゴリズムは、PMAT を対象として、ウィンドウを用いてピッチ情報を区切り、時系列に沿ってピッチ情報の密度を計算する(図 4). ここで、密度が定義済みのしきい値 t (25%など)を超えている場合、ウィンドウを用いて区切った

時区間をサビ構造として定義する.

● Step-5:[時区間の最適化]

Step-5 では、Step-3,4 において分析した繰り返し構造とサビ構造を対象として、各構造の時区間の最適化を行う. 具体的には、繰り返し構造分析アルゴリズムを用いて抽出した繰り返し構造のセクションは、サビ構造のセクションと時区間が重複するため、繰り返し構造の時区間の開始時間で分割し、分割した時区間を繰り返し構造として再定義し、繰り返し構造とサビ構造の最適化を行う.

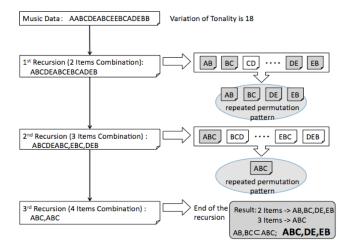


図 3 繰り返し構造分析アルゴリズム:調性の列を対象として,最長一致のパターンを検出

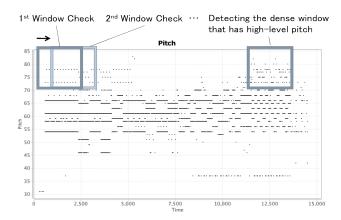


図 4 サビ構造分析アルゴリズム:ウィンドウ を用い時系列に沿ってピッチの密度を計算

2.4. 印象メタデータ生成機能

本節では、抽出した印象的なセクションを対象として、印象メタデータを生成する方法を示す。本システムは、印象的なセクション部分の物理的特徴量を対象として、印象変換マトリクス(EMAT)を用いて、当該セクションについての感性的特徴量を生成し、印象メタデータとしてそれぞれのセクションのメタデータ

を保存する.

本システムは、EMAT の物理的特徴量として、24 階調の調性を採用し、感性的特徴量として、5 ワードの印象語(happy, sad, warm, cool, angry)を実験的に採用する. EMAT における印象語と調性の関係性の定義を行うため、文献[5]に記述されている調性と印象の関連性の定義文を用いる. 調性と印象語が、同様、または、近似する意味を有する場合、行列中の対応するセルに1 を格納し、否定的な意味を有する場合は、対応するセルに-1 を入力した. 調性と印象が無関係のセルについては、0 を格納している.

本システムは、EMAT に格納されている感性的特徴 量の各ベクトルと、印象的なセクションの各ベクトル 間の内積値を算出する関数を提供する.この関数を次 のように定義する.

$$trans(E, Mp) \rightarrow \bigcup_{k=0}^{l} \left(\sum_{i=0}^{n} \sum_{j=0}^{m} E_{[i,k]} \cdot Mp_{[i,j]} \right)$$

ここで、E は EMAT の i 番目の印象語の k 番目の属性値であり、Mp は当該セクションの i 番目の時区間の j 番目の属性値である.

2.5. 個人化機能

本システムは、印象変換マトリクス(EMAT)を個人別に変更することにより、音楽検索における個人化を実現している. 具体的には、印象的なセクションがもつ物理的特徴量を、利用者が当該セクションに対して抱く印象の物理的特徴量として定義し、EMATに追加する.

2.6. 検索機能

本システムは、2.5 節において分析した印象メタデータを対象として、楽曲の検索機能を提供している. 本システムは、クエリとして与えられた印象語を用いて、音楽データベース内の楽曲をランキング化し、検索結果を出力する.また、各楽曲おける印象メタデータの生成結果と、抽出した印象的なセクション構造を可視化する機構を有している.

3. プロトタイプシステム

本節は、プロトタイプシステムの実装について述べる. 実装には Java 言語を用いた. 図 5 に、音楽データを対象とした印象メタデータの生成結果(図中右のレーダーチャート)と印象的なセクション構造の可視化結果(図中左の線グラフ)を示す.

利用者は、本システムの検索入力画面にクエリを入力することにより、クエリと適合した印象をもつ楽曲の検索一覧を得る. 更に、検索一覧の楽曲名を選択す

ることにより、楽曲の印象メタデータの生成結果と印象的なセクション構造の分析結果を閲覧することが可能である.

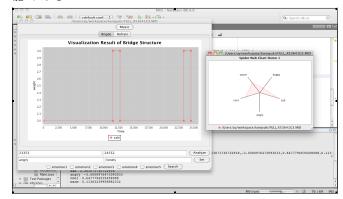


図 5 プロトタイプシステム:印象メタデータ の生成結果と、抽出した印象的なセクション構造 を可視化

4. 実験

本節は、楽曲内における、繰り返し構造分析による 印象的なセクション(サビ構造、及び、繰り返し構造) の自動抽出メカニズムの実現可能性、及び、有効性の 検証を行った2つの実験結果を示す.

4.1. 実験1:自動抽出精度の検証

実験1では、10楽曲を対象として、印象的なセクョンの自動抽出を行った結果を示し、本プロトタイプシステムの自動抽出精度について検証する.実験対象データとして、曲調にばらつきがあった、1999年にミリオンセラーの売上を記録したシングル CD に格納されている10楽曲の音楽データを採用した.実験対象データの一覧を次に示す.

作品名	アーティスト	ジャンル
だんご3兄弟	速水けんたろう,茂森あ	童謡
	ゆみ/ひまわりキッズ	
Energy flow	坂本龍一	ピアノ(クラシック調)
Addicted To You	宇多田ヒカル	R & B
MOVIN' ON	宇多田ヒカル	R & B
WITHOUT YOU		
フラワー	KinKi Kids	J-POP
Precious Time	SPEED	J-POP
LOVEマシーン	モーニング娘。	J-POP(ダンス&ボーカル)
Winter, again	GLAY	ロック(バラード)
ここではない,	GLAY	ロック
どこかへ		
HEAVEN'S DRIVE	L'Arc-en-Ciel	ロック

実験1では、あらかじめ対象とする音楽データを試聴して、サビ構造の開始時間、及び、終了時間を測定し、開始時間の早いものから順に番号を振り、正解データとした。繰り返し構造についても同様の手順を用い正解データを生成した。続いて、システムが音楽データから自動抽出を行ったサビ構造、及び、繰り返し構造の開始時間、及び、終了時間(自動抽出結果)を対象として、正解データとの比較を行い、次の4つの誤差の基準を定義し抽出時間の誤差の総和を算出した。

- 誤差1)自動抽出結果が正解データと重複し,自動抽出結果の開始時間が正解データの開始時間よりも早いもしくは遅かった場合,自動抽出結果,及び、正解データ各々の開始時間の時間差
- 誤差2)自動抽出結果が正解データと重複し、自動抽出結果の終了時間が正解データの終了時間よりも早いもしくは遅かった場合、自動抽出結果、及び、正解データ各々の終了時間の時間差
- 誤差3)自動抽出結果が正解データと重複せず、 自動抽出結果の開始時間、あるいは、終了時間の 前後 t 秒以内に正解データが存在した場合、自動 抽出結果、及び、正解データ各々の中心点の時間 差(t=20)
- 誤差4)自動抽出結果が正解データと重複せず、 自動抽出結果の開始時間、あるいは、終了時間の 前後 t 秒以内に正解データが存在しない場合、自 動抽出結果の時間(t = 20)

また、本方式は、時系列メディアデータの構造分析による部分的時区間の選択を行うため、システムの自動抽出結果が、正解データとより近似することを想定している。そのため、本方式よって検出されたサビ構造、及び、繰り返し構造の適合率を次のように算出した。

検出適合率=(自動抽出結果と正解データの結果が適合した数×2/(自動抽出によるセクション構造の検出数+正解データの数))×100【単位:パーセンテージ】

実験1の検証結果として、図6にサビ構造の抽出における誤差の総和を算出した結果、及び、検出適合率の計測した結果のグラフを、図7に繰り返し構造の抽出における誤差の総和を算出した結果、及び、検出適合率を計測した結果のグラフを示す.

サビ構造の抽出における検証結果から、本方式によるサビ構造の検出は、楽曲構造が単純であるり、童郎や現代音楽はないては、サビ構造の検出は難しい。また、ロック楽曲において講しい。また、ロックであるにおいて、サビ構造の検出は難しい。また、ロックであるにおいても、間奏中にギターの高音域が特徴的た。このも発出適合率が低くなった。のも分析のより、音楽のジャンルごとに、サビ構造のがいることが必要であることが必要である。特性、共通性を抽出し、それらを反映可能なサビ構造、共通性を抽出し、それらを反映可能なサビ構造に、共通性を抽出し、それらを反映可能なサビ構造に、共通性を抽出し、それらを反映可能なサビ構出アルゴリズムを実現することが必要である。

続いて、繰り返し構造の抽出における検証結果から、本方式による繰り返し構造の検出は、楽曲の大きな変化においては有効であり、1番と2番以降の検出には

十分な精度が確認できた.本方式は,24 階調の調性における楽曲中の変化を用いているため、楽曲の短い時区間を対象とした場合は、前後で調性における大きな変化をしない.そのため、楽曲の短い時区間を対象とした変化を抽出するには、新たな方式を試す必要がある.

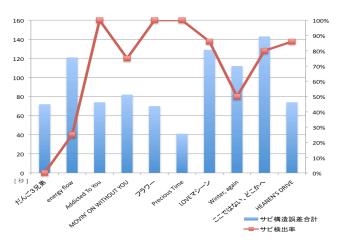


図 6 サビ構造の自動抽出における誤差の総和 (左縦軸),検出適合率(右縦軸)

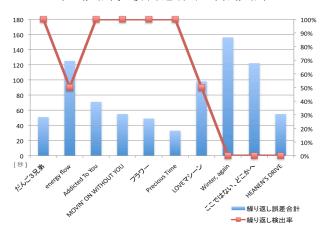


図 7 繰り返し構造の自動抽出における誤差の総和(左縦軸),検出適合率(右縦軸)

4.1. 実験2:印象時区間選択機能の検証

実験2では、実験1において用いた10楽曲の音楽データを対象として、音楽データの全時区間、及び、印象的なセクションの部分的時区間における印象メタデータ生成結果を示し、印象時区間選択機能による楽曲検索の実現可能性と有効性について検証する.

実験2では、物理的特徴量(調性)と感性的特徴量(印象語)の相関関係を定義した印象変換マトリクス(EMAT)を用い、印象メタデータの生成を行った.本システムは、EMATの物理的特徴量として24階調の調性を採用し、感性的特徴量として、5ワードの印象語(happy, sad, warm, cool, angry)を実験的に採用した.

図 8 に定義した EMAT の概略図を示す.

本実験では、印象的なセクションのセクション構造として、サビ構造を採用した。図 9 に対象音楽データの全時区間における印象メタデータの生成結果を、図 10 に印象的なセクションの部分的時区間における印象メタデータの生成結果を示す.

ここで、全時区間における印象メタデータの生成結果と、サビ構造である部分的時区間における印象メタデータの生成結果を、図 11 に示した一般的な J-POP 楽曲である「フラワー」と「Precious Time」において比較すると、「Precious Time」は印象メタデータの生成結果が近似している(図 11 中左)が、「フラワー」は印象メタデータの生成結果に差が生じている(図 11 中右).「フラワー」のように、全時区間における印象メタデータの生成結果と、サビ構造である部分的時区間における印象メタデータの生成結果に差が生じるような音楽データの場合、本方式による部分的時区間の選択は有効である.

次に、印象語による対象音楽データの検索を行い、 全時区間を対象とした印象メタデータ生成と、印象的 なセクションの部分的時区間を対象とした印象メタデータ生成において、検索結果のランキングの変化を検 証した、検証結果として、図 12 に音楽データの全時 区間、及び、印象的なセクションであるサビ構造を対 象とした楽曲検索におけるランキングの変化を示す・ 検証結果から、検索におけるランキングの変化を示す・ で、楽曲検索において楽曲順位が入れ替わり、検索結 果は変化している。本方式による時区間の選択によって、特定の時区間が鑑賞者に与える印象に基づいた検索や、時区間の間の印象の違いを考慮した検索の実現 可能性を示した。

		物理的特徵量					
		C	Am	G		Dm	
	happy	1	-1	0		0	
感	sad	0	1	-1		1	
性的	angry	1	1	0		1	
特.	cool	0	1	1		1	
徴	warm	0	-1	1		0	
量							

図 8 物理的特徴量と感性的特徴量の相関関係 を定義した印象変換マトリクス (EMAT)

対象音楽データ	happy	sad	angry	cool	warm
だんご3兄弟	-0.319529708	-0.653648164	-0.248418313	-0.758756623	-0.988117494
energy flow	-1.568467817	0.737095243	-3.652698557	-0.757500472	-1.270191702
Addicted To You	-0.094544294	0.067927124	-0.733126835	-1.07381759	-2.917039601
MOVIN' ON WITHOUT YOU	-0.847027397	2.027729442	-0.326555466	1.092706479	1.952883616
フラワー	-1.543090731	2.689577468	-2.284431465	1.38634745	0.600861363
Precious Time	-1.375967622	0.744203061	-3.206041485	-1.029110327	-1.876937702
LOVEマシーン	0.226958543	0.5120399	1.243735313	1.254095852	3.114082105
Winter, again	-1.585932039	2.558316159	-1.834320731	1.801021355	1.415778982
ここではない、どこかへ	-1.396654442	2.334403156	-1.893690824	1.427470817	1.37525968
HEAVEN'S DRIVE	-1.294482082	1.614244595	-1.828345105	-0.747451007	-0.656219184

図 9 定義した EMATと,音楽データの全時区 間の印象メタデータを生成

対象音楽データ	happy	sad	angry	cool	warm
だんご3兄弟	-0.659664471	-0.150415575	-0.618215282	-0.669123938	-1.105812913
energy flow	-1.3078731	0.445262432	-3.521986283	-1.076869074	-1.585934471
Addicted To You	-0.119578709	0.013648828	-0.666541896	-1.101804518	-2.958881563
MOVIN' ON WITHOUT YOU	-0.904208641	1.98771948	-0.283831317	1.265550084	2.159133658
フラワー	-1.488410041	2.489999421	-2.470604084	1.591715068	0.771325909
Precious Time	-1.695203975	0.777826183	-3.632040367	-1.084114306	-1.177302465
LOVEマシーン	0.04535938	0.705783003	0.940797695	1.417500817	3.12379077
Winter, again	-1.548726045	2.795582801	-1.580157116	1.621141411	0.885988243
ここではない、どこかへ	-0.923091911	2.281117919	-1.10344924	1.642287998	1.233939967
HEAVEN'S DRIVE	-1.406124434	1.049227348	-2.505743466	-1.07810393	-2.249694027

図 10 定義した EMATと,音楽データのサビ構造における部分的時区間の印象メタデータを生成

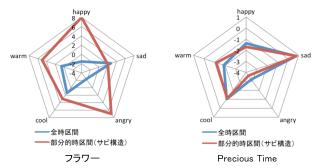


図 11 全時区間,及び,印象的なセクションであるサビ構造の部分的時区間の印象メタデータの可視化結果(楽曲"フラワー"では,全時区間から抽出した印象メタデータと部分的時区間から抽出した印象メタデータが大きく異なり,部分時区間を選択することが有効である)

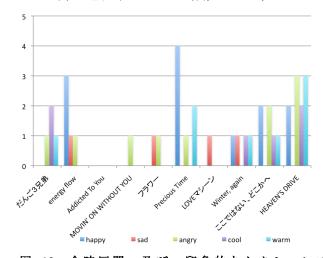


図 12 全時区間,及び,印象的なセクションであるサビ構造を対象とし,印象語による楽曲検索におけるランキングの変動量(すべての印象語においてランキングに有意な変化が見られた)

5. 結論

本稿では、時系列メディアデータである音楽を対象 として、楽曲中における時系列に沿った印象変化を分析・検索する音楽データベースシステムを提案した. 本システムは、時系列メディアデータの構造分析に よる部分的時区間の選択を行い、選択した部分的時区 間別に感性特徴を分析することにより, 印象的な部分的時区間別の印象特徴量を用いた時系列メディアデータの印象メタデータ生成と検索を実現した.

本システムの実現方式として,次の 2 機能により, 検索時に利用者の視点に応じた,適切な時区間の印象 メタデータの選択を行い,利用者が楽曲に抱く印象を 的確に反映した印象分析・検索を実現するシステム構 成を示した.

- 機能 1)音楽データを対象として、楽曲内における繰り返し構造分析による印象的なセクション (サビ構造、及び、繰り返し構造)を自動的に抽出する機能
- 機能2)楽曲の物理的特徴量(調性)と感性的特徴量(印象語)の相関関係を定義した、個人別の印象変換マトリクスを用いて、音楽データが鑑賞者に与える印象を分析し、コンテンツと利用者個人の感性的嗜好の適合性を評価する機能

本システムの実現により、利用者は、印象語をクエリとして、それぞれの楽曲において特徴的な印象変化に応じた音楽獲得の可能性を示すと同時に、キーワード以外の手段として、音楽コンテンツそのものを用いた音楽獲得が可能となった。また、音楽の中身を聴くことなく、音楽の印象変化の視覚的に把握し、個人別の印象変換マトリクスを用いた個人化を適用することが可能となった。

本稿では、本システムの自動抽出精度の検証、及び、 印象時区間選択機能による楽曲検索の実現可能性と有 効性の検証を行い、楽曲内における、繰り返し構造分 析による印象的なセクション(サビ構造、及び、繰り 返し構造)の自動抽出メカニズムの実現可能性、及び、 有効性について検証を行った.

本提案方式は、音楽データに留まらず、様々な時系列メディアデータに対して応用可能である。今後、個人の感性的嗜好に強く関連するメディアデータを対象として、時系列メディアデータが鑑賞者に与える印象を分析し、コンテンツと利用者個人の感性的嗜好の適合性を評価する検索支援の応用を行っていく。

参考文献

- [1] 後藤 真孝, 平田 圭二: "音楽情報処理の最近の研究", 日本音響学会誌 60 巻 11 号, (2004), pp.675-681.
- [2] 吉井和佳 and 後藤真孝. "音楽情報処理技術の最前線: 7.音楽推薦システム". 情報処理学会誌, Vol.50, No.08, pp.751-755, 2009.
- [3] Krumhansl, C. L. "The Cognition of Tonality as We Know it Today". Journal of New Music Research, 33, 3, pp.253 268, 2004.

- [4] D.Temperley: "Music and Probability", MIT Press, ISBN 0262201666, 2007.
- [5] 貴島清彦: "音楽の形式と分析",音楽之友社, 1980.
- [6] Hanna, P., Robine, M. and Rocher, T. "An alignment based system for chord sequence retrieval". In Proceedings of the Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries. ACM, pp.101-104, 2009.
- [7] 長澤槙子,渡辺知恵美,伊藤貴之."Web から入手したデータに基づくコード進行を利用した楽曲類似度の提案と楽曲視聴支援システムの開発". 電子情報通信学会 第19回データ工学ワークショップ (DEWS2008)
- [8] Sahori Imai, Shuichi Kurabayashi and Yasushi Kiyoki: "A Music Database System with Content Analysis and Visualization Mechanisms," In Proceedings of the IASTED International Symposium on Distributed and Intelligent Multimedia Systems, pp.455-460, 2008.
- [9] 社団法人日本レコード協会: 2009 年度音楽メディアユーザー実態調査報告書, 2010.