楽曲動画印象データセットの作成とその分析

山本 岳洋[†,a]　　中村 聡史[††,b]

† 京都大学大学院情報学研究科　†† 明治大学総合数理学部，JST CREST

a) *tyamamot@dl.kuis.kyoto-u.ac.jp*　b) *satoshi@snakamura.org*

概要　本研究では，“可愛らしい”，“切ない”，“元気がでる” といった印象に基づく楽曲検索を実現するための評価基盤として，動画共有サイト上に投稿された 500 件の楽曲動画について，8 つの印象クラスに対する評価値を評価者から収集した．本稿では，得られたデータセットの統計情報を分析することで，楽曲動画の印象間の関連性や，評価者間による評価値のばらつきを検証する．また，これまで我々が提案してきた，楽曲動画の印象推定技術を得られたデータセットに適用することで，楽曲動画の印象推定手法の現状の精度と課題を明らかにする．

キーワード　印象推定，ユーザ生成メディア，音楽情報検索

1　はじめに

音楽は人々の生活に欠かせない重要な娯楽の 1 つである．我々は日常的に音楽を聞いたり，歌ったりしながら日々を過ごしている．近年のインターネットの発展により，多くの楽曲がウェブ上でアクセス可能となった．特に，初音ミクに代表される，VOCALOID と呼ばれる歌声合成技術 [9] の普及は，これまで楽曲作成とは無縁であったユーザ層にまで創作の場を広く開放することとなった．その結果，現在では多くの人々の手によって膨大な数の楽曲が日々創作，公開されている [10]．

しかし，人々にとって，アクセス可能となる楽曲数が膨大になる一方で，求める楽曲を探すための検索手段は多様であるとは言えない．特に，VOCALOID を利用して創作された楽曲のような，新しい形態の楽曲を視聴しようとする場合，気に入っているアーティストやジャンルといった，新しい楽曲に出会うための手がかりに乏しく，どのような観点から自らの求める楽曲を検索すれば良いのかが不明瞭なことが多いと考えられる．

そのような状況下で楽曲を検索する際，1 つの手段として，楽曲から受ける “印象” が重要な役割を果たすのではと我々は考え研究を進めている．本稿で扱う印象とは，“爽やかな音楽”，“元気がでる音楽”，“切ない音楽” といった，楽曲を視聴して受ける視聴者の主観的な感情のことである．「爽やかな音楽で人気のある楽曲」や「ランキング上位にある楽曲の中で切ない印象を受ける楽曲」といった楽曲の探し方が可能となれば，新しいドメインにおける楽曲を探そうとしている初心者への検索手段になり，また，そうでない検索ユーザに対してもこれまでにない新しい観点からの検索手段を提供することができる．本研究の大きな目的は，楽曲から受けるさまざまな印象に基づいて自由に楽曲を検索可能な仕組みを実現することである．

我々はこれまでに，楽曲動画を視聴中のユーザがその動画に付与したコメントを利用して，楽曲動画の印象を推定する手法を提案してきた [11][12]．これらの研究では，楽曲動画に付与されたタグ集合からボトムアップ的に楽曲動画の印象クラスを用意し，また，楽曲動画が属する印象クラスをタグから自動的に決定することで，評価用のデータセットを構築していた．しかし，タグと視聴者のコメントは独立したものではなく，あるタグが付与されているために視聴者が特定のコメントをしたり，あるコメントが付与されているために特定のタグがその動画に付与されたりといった，因果関係も存在すると考えられる．このような，タグとコメントの関係とは独立に，印象推定手法の有用性を評価するためには，タグから自動的にデータセットを構築するだけではなく，人手の評価を用い，トップダウン的にデータセットを構築する必要があると考えられる．

そこで，本研究では，楽曲動画に対する印象を，タグを用いた自動的なアプローチではなく，ユーザ実験により人手でラベル付けすることで，楽曲動画の印象データセットを構築する．具体的には，ニコニコ動画上に投稿された 500 件の楽曲動画に対して，音楽情報検索で用いられている 8 つの印象クラスそれぞれについて評価者から評価値を収集する．

本稿では，評価データの収集方法について述べるとともに，得られたデータセットを対象に分析を行う．まず，得られたデータセットの統計情報を分析することで，楽曲動画の印象間の関連性や，評価者間による評価値のばらつきを検証する．その後，これまで我々が提案してきた，楽曲動画の印象推定技術を得られたデータセットに適用することで，人手で作成したデータセットに対する，印象推定手法の精度を検証する．本研究で作成した，評価者により構築されたデータセットや，これまでに作成したタグに基づくデータセットなどを構築・整備するこ

とで，印象に基づく音楽情報検索に関する研究を進めるための基盤データとなると我々は考えている．

2 関連研究

音楽情報処理の分野では，楽曲のジャンル，作者，そして印象などの推定に関する研究が，ユーザの検索を支援するために行われている．特に，楽曲の印象（*mood* あるいは *emotion* と呼ばれる）推定は，近年注目を集めており，たとえば，音楽情報検索の評価に関するワークショップである MIREX [5] では 2007 年から楽曲の印象推定に関するタスクが行われている．

2.1 楽曲の印象モデル

楽曲の印象の表現方法については，さまざまなアプローチが提案されている．楽曲の印象のモデル化に関する最も古いものとしては Hevner の研究 [3] がある．Hevner は楽曲に対する印象を，8 グループの印象語群としてモデル化している．また，MIREX では，印象を表す形容詞をクラスタリングすることで，印象を 5 つのクラスに分割し，印象推定のタスクに用いている．

また，楽曲のみを対象としたものではないが，楽曲の印象推定にも広く用いられているモデルとして，Russel が提案した Valence-Arousal 空間がある [7]．Valence は快-不快を表す次元，Arousal は覚醒-鎮静を表す次元であり，印象をこの 2 つの軸で張られる空間上で表現するという考え方である．本稿では，MIREX の 5 つの印象クラスと，Russel による Valence-Arousal 空間を基にして，印象データセットを構築する．

2.2 楽曲の印象推定

楽曲の印象推定に関する研究は，音楽情報検索の分野において，近年特に取り組まれてきている [6]．それらの研究では，楽曲の音響信号から得られる音響信号を利用したものが多い．また，近年では音響特徴量だけでなく，楽曲の歌詞を利用した手法も提案されている [4]．

このように，楽曲の印象を推定する手法がいくつか提案されているものの，楽曲のアーティスト名やジャンルの推定などと比較して，印象推定の精度は低い．我々が作成したデータセットは，そうした楽曲の印象推定技術のための評価基盤の 1 つとなると考えられる．

3 印象データセットの構築

本章では．本研究で作成した，楽曲動画印象データセットの構築方法を述べる．まず，本研究で対象とする印象モデルと楽曲動画について説明し，実際の評価データ収集方法について述べる．

3.1 印象クラス

本研究では，楽曲動画に対する印象として，2.1 節で述べた，MIREX で用いられている 5 つの印象クラスと，Russel らの Valence-Arousal 空間を参考にした．MIREX では，5 つの印象クラスが用いられていたが，我々のこれまでの研究により，ニコニコ動画上では，「可愛らしい」と感じる楽曲やそれに関するタグが多く存在することが分かっている．そのため，本研究では MIREX の 5 クラスに加えて，可愛らしさ表す印象クラスを加えることで，6 つの印象クラスと，Valence と Arousal にを表す 2 クラスの計 8 つの印象クラスを評価対象とした．

表 1 本実験で対象とした印象クラス．

印象クラス名	印象を表す形容詞・形容動詞
C1（堂々）	堂々とした，どっしりとした， 心躍る，にぎやかな，
C2（元気が出る）	元気が出る，楽しい気持ちにさせる， 陽気な，心地よい
C3（切ない）	切ない，悲痛な，ほろ苦い， 気が滅入る，哀愁の
C4（激しい）	アグレッシブな，激しい，興奮させる， 熱情的な，感情あらわな
C5（滑稽）	滑稽な，ユーモラスな，面白げな，奇抜な， 気まぐれな，いたずらっぽい
C6（可愛い）	可愛らしい，愛くるしげ，愛おしい，かわいい
Valence	明るい気持ちになる，楽しい， 暗い気持ちになる，悲しい
Arousal	激しい，積極的な，強気な， 穏やか，消極的な，弱気な

本研究で用いた 8 つの印象クラスを表 2 に示す．表中の “印象クラス名” は，著者らが便宜上付与した，印象クラスを表すラベル名である．また，“印象を表す形容詞” は，評価実験において評価者から評価値を収集する際に，その印象クラスを表現するために用いた表現を表している．クラス C1 から C5 については，MIREX で用いられていた形容詞を著者らが日本語に直したもの，C6 については，“可愛い” の類義語を集めた．また Valence-Arousal についても，既存研究を参考に著者らが日本語に直したものを，印象クラスを表す語群として用いた．

3.2 楽曲動画

評価対象の楽曲動画として，動画共有サイト「ニコニコ動画」上に投稿された楽曲動画を用いた．実際には，「VOCALOID」タグの付与された動画のうち，2012 年 8 月時点で再生数の多い動画上位 500 件を抽出し，評価対象の動画とした．

3.3 楽曲評価インタフェース

図 1 に評価データ収集に用いたインタフェースを示す．評価者は楽曲動画を試聴し，その楽曲動画に対する印象を， 以下に示す形で付与する．

- **C1-C6 の印象クラス:** 表 2 に示した形容詞・形容動詞群に対する，5 段階（1:まったくそう思わない

図 1 評価用インタフェース.

～ 5:とてもそう思う）のリッカート尺度.

- **Valence:** −2（暗い気持ちになる，悲しい）～ +2（明るい気持ちになる，楽しい）の 5 段階のリッカート尺度.
- **Arousal:** −2（穏やか，消極的な，弱気な）～ +2（激しい，積極的な，強気な）の 5 段階のリッカート尺度.

なお，楽曲を試聴せずに評価してしまうことがないように，楽曲動画を全て視聴し終えるまで，評価ボタンは押下できないようにした.

3.4 評価データ収集

3.3 節で述べた評価インタフェースを用い，2013 年 4 月から 2013 年 10 月にかけて，楽曲動画の印象に対する評価データを収集した．評価には，明治大学の学部生と著者らを含む計 14 名が参加した．また，1 つの動画につき少なくとも 3 名の評価者から評価が得られるよう，評価対象動画を割り当てた.

実験期間中に，500 件の楽曲動画それぞれに対して少なくとも 3 名の評価者から，延べ 1,537 件の動画に対する評価を収集した．また，評価者 1 人あたりの評価動画数は平均約 110 件（最大 302 件，最小 13 件）であった.

4 統計情報

本章では．3 章の実験から得られたデータセットの基礎的な分析を行うことで，印象クラスの分布，クラス間の関連，評価者間の差異といった情報を明らかにする.

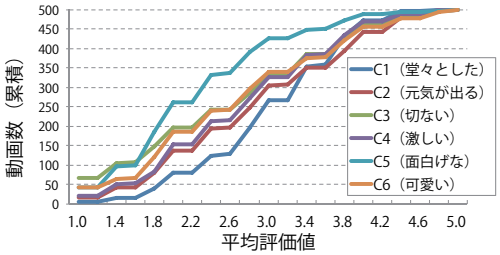

図 2 印象クラスごとの累積度数分布（C1-C6）.

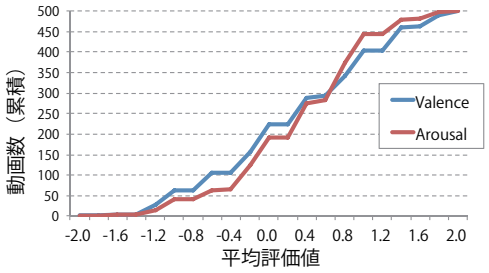

図 3 印象クラスごとの累積度数分布（Valence-Arousal）.

4.1 印象の分布

図 2，図 3 は，印象クラス C1-C6，Valence-Arousal それぞれについて，評価者の評価値の平均値の分布を表した図である．図 2 を見ると，C2-C5 および C6 に関しては，類似した分布となっており，評価値 3.0 以下の動画数が 250 件程度となっており，各評価値に均等に動画が分布していることがわかる．一方で，C5（面白げな）クラスに関しては，多くの動画が低い評価値を得ており，高い評価値を得ている動画が他の印象クラスと比較して少ないことが分かる．これは，C5 クラスは表 2 にあるように，他の楽曲とは異なる，変わった楽曲に関するクラスであるため，このような分布になったと考えられる.

4.2 印象間の相関

本研究で用いた 8 つの印象クラスは，それぞれに関連があるクラスもあれば，互いに独立したクラスもあると考えられる．そこで，印象クラス間の評価値の相関を評価することで，印象クラス間の関連について調査した.

表 2 は，評価者から得られた評価値の平均値について，印象クラス間ごとにピアソンの積率相関係数をまとめた表である．表 2 から，たとえば，C1（堂々とした）クラスを見てみると，C2（元気が出る）や Valance と強い正の相関を持っていることが分かる．また，C3（切ない）クラスでは，Valence と強い負の相関を持っていることなどが分かる．一方で，C5（面白げな）クラスは，他のクラスと強い相関を持っておらず，他のクラスとは独立した印象である傾向が強いことを示している.

表 2　印象クラス間のピアソンの積率相関係数．表中の太字は，データ間に有意に（$p < .01$）相関があることを示す．また，表中の網掛で表されたセルは，相関係数が.500 より大きい，あるいは-.500 より小さいことを示す．

	C2	C3	C4	C5	C6	Valence	Arousal
C1（堂々）	**.527**	**-.482**	**.295**	**.163**	**.300**	**.530**	**.387**
C2（元気）	–	**-.724**	**-.142**	**.213**	**.679**	**.835**	-.015
C3（切ない）	–	–	**.180**	**-.258**	**-.465**	**-.792**	-.077
C4（激しい）	–	–	–	-.029	**-.241**	**-.149**	**.674**
C5（面白げな）	–	–	–	–	**.153**	**.197**	.045
C6（可愛い）	–	–	–	–	–	**.628**	**-.209**
Valence	–	–	–	–	–	–	.053
Arousal	–	–	–	–	–	–	–

4.3　評価者間の相関

4.1 節および 4.2 節では，評価者の評価値の平均値を用いて，データセットの分析を行った．また，5 章でも，評価者の評価値の平均値を用いて，印象推定手法の精度検証を行う．

しかし，楽曲動画に対する印象は，評価者の主観的に強く依存すると考えられ，アーティスト名やジャンルといった情報と比較すると，被験者間の評価のばらつきが生じやすい情報であると考えられる．そこで，本節では，評価者間の評価がどの程度一致しているのかを確認し，被験者の評価値の平均値も用いることの妥当性を検証する．

図 4 は，各印象クラス間で，評価者間の評価値のピアソンの積率相関係数を求め，被験者間ごとに平均値を求めたものである．図中の “合計” は全ての印象クラスの評価値についての相関値を表す．なお，被験者間の共通となる動画数が少なすぎる場合は相関値に信頼性がないと考え，共通に評価した動画が 10 件以上存在する被験者間でのみ，相関係数を求め，平均値を算出した．

図 4 を見ると，全ての印象クラスに対する評価値における相関係数をみると，0.6 程度となっていることが分かる．このことから，楽曲の印象は人々の主観に依存するものの，人々間に一定の傾向が存在することを示している．また，個々の印象クラスを見てみると，C2（元気が出る），C3（切ない）や C6（可愛い）などのクラスにおいて被験者間の相関係数が高くなっていることが分かる．一方で，C1（堂々とした）や C5（面白げな）といった印象クラスについては，相関係数の平均値が 0.2 から 0.25 程度となっており，正の相関があるものの，他の印象クラスと比べて弱い相関となっていることが分かる．

4.4　動画共有サイト上のメタデータとの関連

本研究で対象とした楽曲動画は，動画共有サイト上に投稿された動画であった．そこで，楽曲動画に対する評価値と，動画共有サイトに特有の，再生数，コメント数といったメタデータとの相関を調査した．

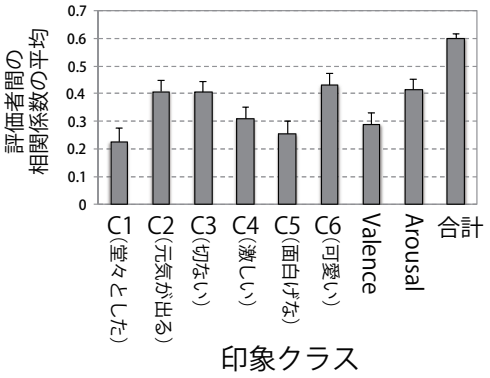

図 4　評価者間の相関．図中のエラーバーは標準誤差を表す．

表 3　印象クラスと動画サービス上のメタデータとのスピアマンの順位相関係数．表中の太字は，データ間に有意に（$p < .01$）相関があることを示す．

	再生回数	コメント数	マイリスト数
C1（堂々）	**.154**	**.133**	**.149**
C2（元気が出る）	.064	.072	.056
C3（切ない）	**.119**	**.169**	**.128**
C4（激しい）	**.212**	**.271**	**.250**
C5（滑稽）	**.184**	**.212**	**.136**
C6（可愛い）	**.150**	**.170**	**.162**
Valence	.001	-.011	.002
Arousal	.110	**.140**	**.123**

表 4 は，各印象クラスにおける評価値と，再生数，コメント数，マイリスト数それぞれとのスピアマンの順位相関係数をまとめた表である．表にあるように，多くの印象クラスにおいて，強い相関は得られなかった．これは，今回評価の対象とした動画共有サイト上には，特定の印象に関連した楽曲動画だけでなく，多様な楽曲動画が，多様な質で投稿されていることを示しているのではと考えられる．

5　視聴者コメントを用いた印象推定

本章では，3 章で得られた印象データセットに対して，これまで我々が提案してきた，視聴者コメントを用いた楽曲動画の印象推定手法 [11][12] を適用した結果を分析する．まず，手法の概要を説明する．その後，3 章で得られた印象データセットに対して，手法を適用した結果について述べ，結果を考察する．

5.1　手法の概要

図 5 に手法の概要を示す．提案手法では，視聴者コメントから（1）形容詞,（2）正規化されたコメント,（3）楽曲のサビ区間に出現する形容詞および正規化されたコメントを素性として抽出し，分類器を学習する．以下，それぞれの素性について簡単に説明する（手法の詳細については，文献 [12] を参照）．

（1）形容詞: 形態素解析器を用いて，視聴者コメントを単語に分割し，得られた形容詞および形容動詞を

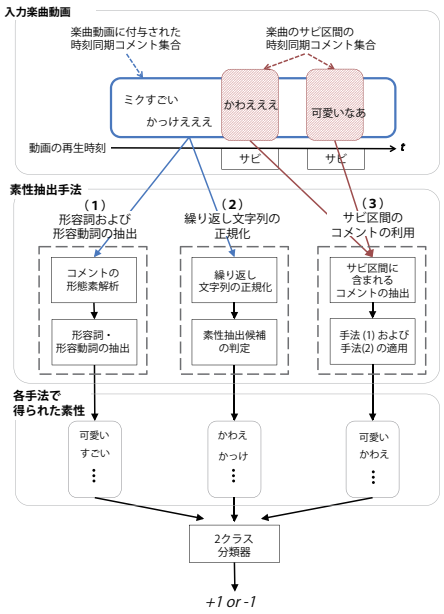

図 5 視聴者コメントを用いた印象推定手法の流れ.

素性として抽出する．本研究では，形態素解析器として MeCab[1]を使用した．

（2）正規化されたコメント: Brody らは，マイクロブログの 1 つである Twitter[2]に投稿される文章において，本来 “cool” と記述されるべき単語が，“cooooollllll” のように，“o” や “l” が繰り返された単語として記述されるなど，感情に関連した単語はこのように単語中の文字が繰り返されて Twitter 上に投稿されていていることを指摘している [1] 本研究で扱うニコニコ動画も，彼らの指摘した文字の繰り返し構造が起こったコメントが投稿され，また，そうしたコメントは印象と関連が深いと考えられる．そこで，Brody らの手法を利用し，“かっけえええ” のような，特定の文字列が連続して出現するようなコメントを正規化し，素性として抽出する．

（3）楽曲のサビ区間に出現する形容詞および正規化されたコメント: 楽曲動画に付与されたあるコメントがその楽曲の印象と関連するものかどうかは，コメントが投稿された動画の再生時刻とも関わっていると考えられる．特に，“サビ” は，コーラス（chorus）あるいはリフレイン（refrain）とも呼ばれ，楽曲全体の構造の中で一番代表的な，盛り上がる主題を表す部分 [2] であり，視聴者が楽曲に対して受ける印象を決定づける重要な区間ではないかと考えられる．そこで，まず，Goto により提案された手法 [2] を用い楽曲動画のサビ区間を求め，楽曲動画中のコメントに対して，そのコメントが投稿された時刻がとしてサビ区間に入っている場合に，上述した，形容詞および正規化されたコメントを素性として，分類として用いる．このとき，サビ区間を用いた手法と，上記（1），（2）で述べた手法とで文字列的に同じ素性が抽出されるが，サビ区間を用いた手法で得られた素性はそれらの素性とは異なる素性として扱い学習データを構築する．

5.2 実験設定

本稿では，各印象クラスに対してあらかじめ決められた閾値を超える楽曲動画を正例，そうでない動画を負例として扱い，2 クラス分類器を構築することで，印象推定の有効性を検証した．本稿では，評価者の評価値の平均が，C1-C6 については 3.5，Valence-Arousal については 0.5 より大きな値となっている楽曲動画を正例，そうでない動画を負例とした．本来，データセットに付与された評価値は連続値であるため，回帰モデルなどの適用も考えられるが，今回はどのような印象クラスに対してコメントを用いた印象推定が有効に働くかを検証することを目的としたため，簡易な 2 クラス分類器を用いた．

2 クラス分類器の構築には，分類器の構築手法として広く利用されているサポートベクターマシン（SVM）を用いた．実際の分類器の構築には，SVM のライブラリである LIBSVM[3]を使用し，カーネルとして線形カーネルを，その他のパラメータは LIBSVM の初期設定値を用いた．分類性能の評価尺度には F 値を用い，全体としての評価尺度には F 値のマクロ平均とマイクロ平均を用いた．そして，5 分割交差検定を行い評価値を求め，各値の平均値を求めた．また，分類器を構築する際，訓練データの正例負例の数が同数となるように，負例をダウンサンプリングすることで訓練データにおける正例と負例の不均衡を解消した．

5.3 実験結果

表 4 は 5.1 節で述べた手法（**comment**），楽曲分析において広く用いられている MARSYAS [8] により得られる音響特徴量を素性とした手法（**audio**），両者の素性を組み合わせた手法（**comment+audio**）での分類結果である．

表 4 の **comment** 手法を見てみると，C2（元気が出る），C6（可愛い），Valence-Arousal で高い F 値となっている一方で，C5（面白げな）で低い値となっていることが分かる．これは，4.3 節で述べたように，C5（面白げな）クラスについては評価者間の相関係数が低く，正解とする基準が他のクラスよりも評価者に依存しているため，結果として正解集合に特定のパターンが出現しづらく，精度が低くなったのではないかと考えられる．

我々が行った，タグベースで作成したデータセットでは，音響特徴量に基づく分類手法は，コメントに基づく分類手法よりも精度が著しく低かった．しかし，表

[1]MeCab，https://code.google.com/p/mecab/
[2]http://twitter.com
[3]http://www.csie.ntu.edu.tw/˜cjlin/libsvm/

表 4　手法ごとの F 値の比較．太字で表された数値は手法間での最大値を示している．

		comment	audio	comment+audio
印象クラス	**C1（堂々）**	.471	.482	**.530**
	C2（元気が出る）	.597	.473	**.616**
	C3（切ない）	**.487**	.393	.469
	C4（激しい）	.469	.479	**.495**
	C5（面白げな）	**.273**	.186	.221
	C6（可愛い）	.585	.489	**.597**
	Valence	.688	.526	**.699**
	Arousal	.614	**.652**	.602
	マクロ平均	.523	.460	**.529**
	マイクロ平均	.547	.480	**.553**

4 中の **audio** 手法を見てみると，C5 や Arousal では **comment** 手法よりも高い精度となっていることが分かる．これは，トップダウン的に，人手の評価に基づいて作成した印象データセットでは，タグベースで作成したデータセットよりも，音響特徴量の有効性が高くなっているのではと考えられる．

コメントと音響特徴量を組み合わせた **comment+audio** 手法が全体的に最も高い精度を得ているが，その精度は最大でも F 値で 0.699 と，高い精度とは言えない．今後，印象に基づく楽曲検索を実現するためには，より精度の高い印象推定手法を考えていく必要がある．

6　まとめと今後の展望

本稿では，印象に基づく楽曲検索に関する研究を促進するため，動画共有サイト上に投稿された楽曲動画 500 件を対象とし，各動画に対して 8 つの印象クラスに基づいて，評価者から評価データを収集した．また，作成された評価データを分析し，基本的な特性について明らかにするとともに，既存の印象推定手法を適用することで，印象推定における課題を明らかにした．

我々は，本研究で得られたデータセットを楽曲動画の印象推定手法の評価に利用するだけでなく，下記の様な研究に利用することを考えている．

印象に基づく楽曲動画検索のためのインタラクション: 印象を含んだ情報要求においては，クエリの作成方法から検索結果の提示方法にいたるまで，既存のインタフェースでは不十分な点が多いと考えられる．そこで，本研究で得られたデータセットを正解データとして，印象に基づいた情報要求下における，検索結果の提示方法やインタラクション方式について研究していく予定である．

楽曲動画に対する知覚された類似度の検証: 人々は，ある楽曲と別の楽曲を試聴したときに，類似していると感じることもあれば，類似していないと感じることもある．しかし，楽曲のどのような特徴が類似していれば，人々が楽曲を類似していると感じるのかについては，ほとんど明らかにされていない．今後，人々が知覚する楽曲の類似性と，本研究で得られた印象の評価をもとに，印象という観点から，楽曲間がどのような場合に類似するのかを検証することで，人間が知覚する楽曲の類似性と印象の関係を明らかにしていく予定である．

最後に，本研究で作成したデータセットは，自由に利用可能な形で我々のホームページ上で公開する予定である．本研究で作成したデータセットが，楽曲の印象に基づく検索技術の発展に寄与すれば幸いである．

謝辞

本研究の一部は，科学技術振興機構 OngaCREST プロジェクトによる支援を受けました．ここに記して謝意を表します．

参考文献

[1] S. Brody and N. Diakopoulos. Cooooooooooooooollllllllllllll!!!!!!!!!!!!!!: using word lengthening to detect sentiment in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 562–570, 2011.

[2] M. Goto. A chorus section detection method for musical audio signals and its application to a music listening station. *IEEE Transactions on Audio, Speech and Language Processing*, 14(5):1783–1794, 2006.

[3] K. Hevner. Experimental studies of the elements of expression in music. *The American Journal of Psychology*, 48(2):246–268, 1936.

[4] X. Hu, J. Downie, and A. Ehmann. Lyric text mining in music mood classification. In *Proceedings of the 10th International Society for Music Information Retrieval Conference*, pages 411–416, 2009.

[5] X. Hu, J. Downie, C. Laurier, M. Bay, and A. Ehmann. The 2007 MIREX audio mood classification task: Lessons learned. In *Proceedings of the 9th International Conference on Music Information Retrieval*, pages 462–467, 2008.

[6] Y. Kim, E. Schmidt, R. Migneco, B. Morton, P. Richardson, J. Scott, J. Speck, and D. Turnbull. Music emotion recognition: A state of the art review. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*, pages 255–266, 2010.

[7] J. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[8] G. Tzanetakis and P. Cook. MARSYAS: A framework for audio analysis. *Organised sound*, 4(3):169–175, 1999.

[9] 剣持秀紀. 歌声合成の過去・現在・未来: 「使える」歌声合成のためには. **情報処理** *(***情報処理学会誌***)*, 53(5):472–476, 2012.

[10] 後藤真孝. 初音ミク，ニコニコ動画，ピアプロが切り拓いた CGM 現象. **情報処理** *(***情報処理学会誌***)*, 53(5):466–471, 2012.

[11] 山本岳洋，中村聡史. 視聴者の時刻同期コメントを用いた楽曲動画の印象推定. *Web* **とデータベースに関するフォーラム** *(WebDB Forum 2012), A3-1*, 2012.

[12] 山本岳洋，中村聡史. 視聴者の時刻同期コメントを用いた楽曲動画の印象分類. **情報処理学会論文誌（トランザクション）データベース**, 6(3):61–72, 2013.