情報処理学会論文誌　データベース　Vol.6　No.3　61–72 (June 2013)

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

山本 岳洋[1,a)]　中村 聡史[2,3,b)]

受付日 2012年12月20日, 採録日 2013年4月5日

概要：本稿では，印象に基づく楽曲検索実現のために，動画共有サイト上に投稿された楽曲動画を，可愛らしい，切ない，元気がでるといった印象に分類する手法を提案する．楽曲動画の印象分類のため，ユーザの投稿した時刻同期コメントに着目し，単語の品詞，文字の繰返し構造，楽曲のサビ区間の 3 つを利用する．実験では 1,314 本の楽曲動画を 7 印象クラスに分類し，提案手法が F 値のマクロ平均で 0.659 を達成しベースライン手法よりも高い精度を得た．また，楽曲の歌詞や音響特徴量を用いた分類手法とも比較し，提案手法の有効性を示した．

キーワード：印象推定，ユーザ生成メディア，音楽情報検索

Using Viewers' Time-synchronized Comments for Mood Classification of Music Video Clips

Takehiro Yamamoto[1,a)]　Satoshi Nakamura[2,3,b)]

Received: December 20, 2012, Accepted: April 5, 2013

Abstract: This paper proposes a method to classify music video clips, which are uploaded to the video sharing service, into the mood categories such as "cute," "sorrow" and "cheerful." The method leverages viewers' time-synchronized comments posted to video clips to classify the video clips into moods. It extracts features from the comments in the terms of (1) parts-of-speech, (2) lengthened words and (3) chorus parts of the music. Our experimental results showed that out method achieved the best classification performance (Macro F-measure of 0.659) compared with some baselines. In addition, our method outperformed the conventional approaches that utilize lyrics and audio features of musics.

Keywords: mood classification, consumer generated media, music information retrieval

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すごしている．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ソフトウェア [1] 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 [2]．たとえば，我々の調査によると，動画共有サービス「ニコニコ動画」[*1] では VOCALOID に関する楽曲の動画が 2012 年 8 月時点で 10 万本以上投稿されている．さらに，1,000 本を超える VOCALOID に関する楽曲動画が毎月投稿されており，新たな楽曲の数は増加し続けている．

人々にとってアクセス可能な楽曲の量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い

[1] 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University, Kyoto 606–8501, Japan
[2] 明治大学総合数理学部
School of Interdisciplinary Mathematical Sciences, Meiji University, Nakano, Tokyo 164–8525, Japan
[3] 科学技術振興機構 CREST
JST CREST, Chiyoda, Tokyo 102–0076, Japan
[a)] tyamamot@dl.kuis.kyoto-u.ac.jp
[b)] satoshi@snakamura.org

[*1] http://nicovideo.jp

えない．現状では楽曲名やアーティスト名といった書誌情報，再生数や投稿日，ユーザが付与したタグなどに基づく方法でしか検索を行うことができない．この問題を解決するため，近年，音楽情報検索の分野では，楽曲から受ける印象に基づく検索が注目を集めている [3], [4]．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，楽曲から受ける，視聴者の主観的な感情のことである．たとえば，

- VOCALOID 楽曲についてはあまり詳しくない．けれども自分は “切ない音楽” が好みなので，VOCALOID の有名な楽曲の中でも切ない楽曲を聴きたい．
- 今は友だちと喧嘩をして気分が落ち込んでいる．そこで “元気がでるような楽曲” を聴いて気分を癒したい．

このような楽曲の探し方が可能となれば，気に入っているアーティストやジャンルがまだな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．

音楽情報検索の分野では，楽曲の印象を推定するためのアプローチとして，楽曲の音響特徴量や，楽曲の歌詞情報と音響特徴量を組み合わせる手法 [5], [6] などが提案されている．しかし，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．一方，ユーザがタグを付与できるサービス上では，楽曲の印象に関するタグを付与することが可能だが，しかし，その割合はニコニコ動画では 5%，音楽に関するソーシャルメディアである Last.fm [*2] では 14%と少なく，現状では検索に利用するには不十分である（詳しくは 3.3 節で述べる）．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．

そのための第一歩として，本稿ではニコニコ動画に投稿された VOCALOID に関する楽曲動画を対象とし，楽曲動画を視聴中のユーザがその動画に付与した反応を利用して，その楽曲動画の楽曲に対する印象を推定する手法を提案する．本研究で対象とするニコニコ動画では，視聴者は動画の視聴中に動画に対してコメントを付与できる．このようにして付与されるコメントは，動画視聴中の視聴者の反応を文字列として表現したものと考えることができ，楽曲に対する印象に関するものも多く含んでいると考えられる．

本稿では，楽曲の印象推定の情報源として，視聴者が付与したそうしたコメントに着目し，以下の 2 つの課題に取り組む．

- **視聴者の付与したコメントからの素性抽出手法の検証**
  本稿では，視聴者のコメントから印象推定に有用な素性を抽出する手段として，(1) コメント中の形容詞，形容動詞，(2) コメント中の繰返し文字の正規化，(3) 楽曲のサビ区間中のコメントの利用，という 3 つの方法を組み合わせた手法を提案する．実験では，7 つの印象クラス計 1,314 本の楽曲動画を用いて精度評価を行った結果，コメント中の全単語を用いる手法やコメントそのままを素性として用いる手法に比べて，提案手法が最も高い分類精度を達成することが分かった．
- **歌詞や音響特徴量を用いた印象推定手法との比較**
  また，楽曲の印象推定におけるコメントの有用性を明らかにするため，提案手法と歌詞や音響特徴量に基づく印象推定手法とを比較した．6 つの印象クラス計 719 本の楽曲動画を用いて精度評価を行った結果，提案手法が既存手法に比べて印象推定に有用であることが分かった．

本稿の構成は以下のとおりである．まず，2 章で関連研究を紹介する．3 章では，ニコニコ動画の概要について述べるとともに，本研究で使用した楽曲の印象データセット作成方法について述べる．4 章で提案手法について述べ，5 章で視聴者コメントからの素性抽出手法に関する評価実験について述べる．6 章では提案手法と歌詞や楽曲特徴量を用いた分類手法との精度比較について述べ，7 章で提案手法について考察した後，最後に本稿をまとめる．

2. 関連研究

2.1 楽曲の印象推定

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 [3] 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．

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 [7] 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，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 [8]．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．さらに近年では，ソーシャルメディアの発展を背景に，楽曲に対して一般のユーザが付与したタグを基に印象を表現するアプローチもとられている [6], [9]．本稿でも，このアプローチを用いて楽曲動画に対する印象のデータセットを構築する．

楽曲の印象推定に関する研究の多くは楽曲の音響信号に基づく特徴量を利用した手法であるが，近年ではそうした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

[*2] http://www.last.fm/

されている [5], [6]．たとえば，Hu らは歌詞特徴量と音響特徴量の両者による分類精度の比較を行っている [6]．彼らは，歌詞から特徴量を抽出する方法として，歌詞を単語集合に分割して得られる素性を TF-IDF 法により重み付けする手法を提案している．彼らは楽曲に付与されたタグを基に 18 個の印象クラスを作成し，2,829 曲の楽曲を用いて評価実験を行っている．その結果，歌詞特徴量のみを利用した手法が音響特徴量による分類手法の精度をわずかに上回ること，また，歌詞と音響特徴量を組み合わせることで，分類精度がさらにわずかではあるが向上することを示している．

2.2　ニコニコ動画におけるコメント利用

ニコニコ動画では，動画の視聴中に，任意の再生時刻に対してコメントを投稿できるという，これまでの動画共有サイトにはない新しい仕組みを提供している．そのため，ニコニコ動画に投稿されるコメントの分析や，コメントを利用したあらたな応用例は，近年研究者の注目を集めている．たとえば，Nakamura らは “泣ける”，“笑える” といった感情を含むコメントに着目し，人手で作成した辞書を用いてそうした感情を含むコメントの時間的な分布の分析や，感情に基づく動画検索を提案している [10]．佃らは動画に登場する人物に関連するコメントの時系列パターンに着目し，対象とする人物の登場パターンに基づく動画検索手法を提案している [11]．ほかにも，ニコニコ動画に投稿された楽曲動画に着目したものとして Yoshii らの研究がある [12]．彼らは，楽曲動画のある時刻のコメントとその時刻における音響特徴量を素性として用いることで，楽曲動画に付与されるコメントを推定し自動生成する手法を提案している．また，我々もこれまでに，視聴者が付与したコメントと楽曲構造理解技術を組み合わせることで，楽曲動画から 15 秒のサムネイル動画を自動生成する手法を提案している [13]．

3.　楽曲動画印象データセットの作成

本章では，まず，本研究で対象とするニコニコ動画の概要を述べる．その後，本研究で用いた，印象推定のためのデータセットの作成方法について説明し，作成したデータセットについての簡単な分析結果を述べる．

3.1　ニコニコ動画の概要

「ニコニコ動画」は，2012 年第 2 四半期の時点で 2,946 万人の利用者がおり，日本で最も利用されている動画共有サービスの 1 つである．本研究で対象とする音楽コンテンツをはじめとして，さまざまなコンテンツに関連した動画がニコニコ動画に日々 5,500 本程度投稿されている [14]．ニコニコ動画における動画視聴時のインタフェースの概要を図 1 に示す．ニコニコ動画における動画視聴インタフェースは，本研究と関連する重要な特徴として下記の 2 点を持つ．

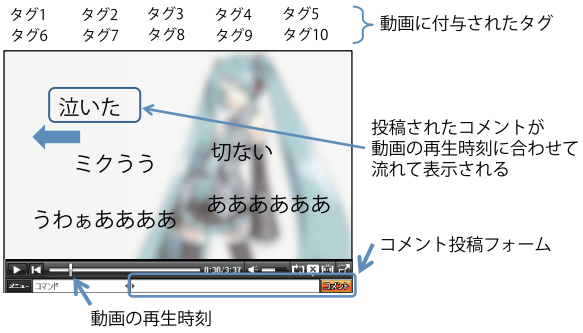

図 1　ニコニコ動画における動画視聴インタフェースの概要
Fig. 1　Overview of viewer interface in NicoVideo.

- 視聴者は動画の視聴中に，任意の再生時刻に合わせて自由にコメントを投稿することができる．投稿されたコメントは，他の視聴者がその動画の視聴中に，投稿された再生時刻と同じタイミングで動画上に流れながら表示される．
- 視聴者は動画に対してタグを付与することができる．このタグはユーザごとではなく動画ごとに管理され，付与できるタグの数は各動画 10 個までである．

本稿では，上記の特徴である，コメントが投稿された際の動画の再生時刻付きのコメント（以降，“時刻同期コメント” と呼ぶ）を用いることで，楽曲動画の印象推定を行う．また，動画に付与されたタグ情報を基に楽曲動画の印象データセットを作成し評価に用いる．

3.2　データセット作成

本研究では，楽曲動画の印象推定問題を，あらかじめ用意した印象のクラスとそれに属する楽曲動画を訓練データとして用意し学習を行うことで，入力楽曲動画をいずれかの印象クラスに分類するという多クラス分類問題として扱う．そのためには，印象としてどのようなクラスが存在するのかを決定し，その印象クラスに属する楽曲動画を用意する必要がある．本稿では，Hu ら [6] と同様に，楽曲動画に付与されたタグから印象クラスを作成し，データセットを作成するというアプローチをとった．Hu らは，分類対象とする印象クラスを決定するため，Last.fm 上で楽曲に付与されたタグから楽曲の印象に関連すると思われるものを抽出し，それを人手でいくつかのグループにまとめあげることで印象クラスを用意している（Mood Tag Dataset として公開されている[*3]）．さらに，得られた印象クラス中のタグが付与された楽曲をそのクラスに属する楽曲と見なしデータセットを作成している．彼らの手法の利点は，ボトムアップ的に楽曲の印象クラスを決められる点と，印象に属する動画を半自動的に決定することでデータセッ

[*3] http://music-ir.org/mirex/wiki/2010:Audio_Tag_Classification

ト作成のコストを削減できるという点があげられる．彼らがデータセットの構築に用いた Last.fm は，音楽を扱ったソーシャルネットワーキングサイトであり，ユーザは楽曲に対して自由にタグを付与することができる．我々はニコニコ動画中の楽曲動画を対象としており，Hu らが対象とした Last.fm とは異なるサービスを対象としているものの，両サービスとも楽曲（ニコニコ動画においては動画）に対してユーザがタグを付与することができるという点で，本研究においても彼らと同様のアプローチでデータセットが作成できると考えられる．

実際のデータセット作成は下記のとおりに行った．

（1）印象に関するタグの抽出　「ニコニコ大百科」[*4,*5]ページ上の「VOCALOID 関連タグ一覧」記事中の「雰囲気による分類」節に記載されているタグ 144 個を収集した．その後，得られた 144 個のタグから，楽曲に対する印象とはいえないもの（楽曲の品質に関するタグ，楽曲のジャンルに関するタグなど）を除去した．

（2）タグのグループ化　残ったタグ集合を，Mood Tag Dataset を参考にしながら，著者らの手により類似する印象を表すタグどうしをまとめ，クラス名を付与し，13 種類の印象クラスにまとめた．

（3）楽曲動画の取得　ニコニコ動画から “VOCALOID” タグの付与された動画 186,987 本を収集し，これを VOCALOID を利用した楽曲を扱った楽曲動画であると見なし，得られた動画集合中で，（2）で得られた印象クラスに属するタグを含む動画をその印象クラスに属する動画として採用した．このとき，複数の印象クラスに属するタグが付与された動画は除去した．

（4）動画数の少ない印象クラスの除去　得られた 13 種類の印象クラスから，そのクラスに属する動画が 30 本以上存在するクラスのみを採用し，最終的に 11 種類の印象クラスを得た．

表 1 に得られたデータセットに関する情報を示す．表中の “印象クラス名” は，著者らが付与した印象クラス名を，“タグ数” はそのクラスにまとめられたタグの数，“動画数” はその印象クラスに属する動画の数，“タグ例” は実際のタグの例を表している．表に示すように，最終的に 50 の印象に関するタグから得られる 8,749 本の楽曲動画をデータセットとして得た．

3.3　印象に関するタグを含む動画の割合

最後に，3.2 節で得られたデータセットを基に，表 1 に示した楽曲の印象に関する 50 のタグ（以降，本節では “印象タグ” と呼ぶ）を含む動画がどの程度存在するのかを分析した．**図 2** は “VOCALOID” タグを含む動画のうち，印象タグを含む動画の割合を，すべての動画およびタグが 10 個付与されている動画それぞれについて再生数ごとに表したグラフである．図 2 より，印象タグの付与されている動画の割合が平均すると 5.0%であることが分かる．このことは，現状では印象に基づく楽曲動画の検索は，たとえ印象タグを手がかりとして検索を行ったとしても，再現率の観点から不十分であることを示している．

このように，印象タグが付与される割合がきわめて低い理由の 1 つには，1 つの動画に付与できるタグ数が 10 個に制限されているというニコニコ動画特有の制約が考えられる．この制約のため，楽曲動画に付与されるタグは，楽曲のメタデータに関するタグ（アーティスト名，使用している VOCALOID 名，楽曲名，“カラオケ配信中” など）が多く，印象に関するタグは付与されにくいのではないかと考えられる．しかし，そのような制約のない Last.fm でも，Hu らが用意した印象に関するタグが付与されている楽曲の割合は 14%程度である．また，ソーシャルブックマークの大規模な分析を行った Heymann らの研究 [15] でも，主観的なタグは全体の 5%であると報告されている．このように，メタデータに比べて，印象に関するタグはやはり付与されにくいのではないかと考えられる．

表 1　作成した楽曲印象データセット
Table 1　Our music mood dataset.

印象クラス名	タグ数	動画数	タグ例
sorrow	4	1,865	切ないルカうた
cute	9	1,851	かわいいボカうた
fresh	7	1,645	爽やかな姉さんリンク
cool	9	928	クールなボカうた
homely	3	383	素朴なミクうた
cheerful	5	335	元気が出るリンうた
darkness	4	273	黒ミク
aggressive	4	253	イケレン
relaxed	3	117	なごミク
depressed	1	116	元気が減るミクうた
dreamy	1	79	巡音幻想曲
合計	**50**	**8,749**	

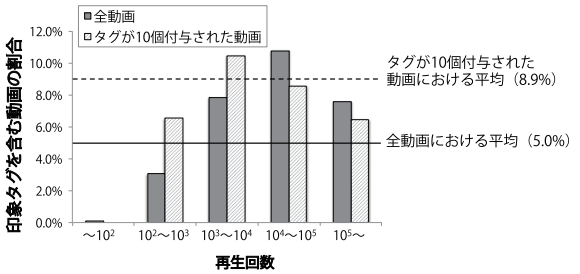

図 2　印象タグを含む動画の割合．図中の直線は全動画の全再生数における平均を，破線はタグが 10 個付与された動画の全再生数における平均値を表す
Fig. 2　Ratio of video clips that contain mood tags.

*4　http://dic.nicovideo.jp/
*5　ニコニコ動画に関連する記事をユーザ同士が自由に作成，編集可能なサービス．

また，印象タグを含む動画の割合を再生数という観点から見てみると，再生数が 1,000 回以下，特に 100 回以下の動画には印象タグがほとんど付与されていないことが分かる．これは，動画が投稿されてからまだ日が浅かったり，動画の人気がなかったりするために動画の視聴者数が十分に集まっておらず，印象タグに限らずタグそのものが付与される機会が少ないためであると考えられる．また，再生数が 10 万回を超えるような動画について見てみると，全動画，タグが 10 個付与された動画ともに，印象タグが付与される割合が再生数が 1 万回～10 万回の動画と比較して低いことが分かる．これは，動画が多くの再生数を獲得して人気動画になるにつれて，動画の人気を表すタグ（“VOCALOID 殿堂入り” など）の付与される割合が高くなるため，印象タグが付与される割合が低下するのではないかと考えられる．

このような結果から，印象の自動推定技術が，印象に基づく検索を実現するにあたり重要であると考えることができる．

4. 時刻同期コメントに基づく印象推定手法

4.1 概要

本稿では，楽曲動画に付与された時刻同期コメントから素性を抽出し，多クラス分類器にかけることで楽曲動画の印象を推定する．提案手法の概要を図 3 に示す．提案手法では，時刻同期コメントから楽曲の印象推定に有用な素性を抽出するため，(1) コメント中の形容詞および形容動詞，(2) コメント中の文字の繰返し構造の正規化，(3) 楽曲のサビ区間中のコメントの利用という 3 つの手法を用いる．この 3 手法はそれぞれ，(1) コメント中の形容詞や形容動詞は楽曲の印象を表している，(2) “かっけええ” や “かっけえええ” のように，コメント内で同じ文字が繰り返し出現するようなコメントは印象と関連が深い，(3) 楽曲のサビはその楽曲の代表的な区間であり，その区間に投稿されたコメントは楽曲の印象を表す，という仮説に基づいている．以降，それぞれの手法の具体的なアイデアと実際の手法の流れについて説明し，最後に素性の重み付け方法を述べる．

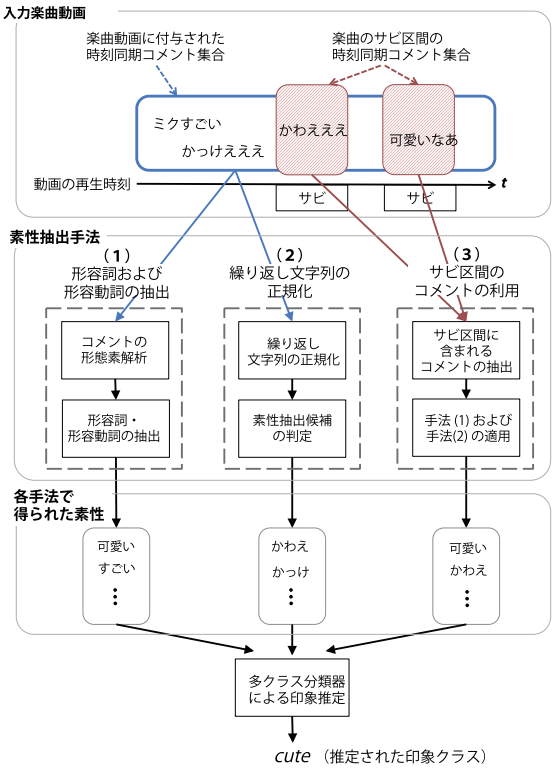

図 3　視聴者の時刻同期コメントからの素性抽出手法の概要
Fig. 3　Overview of our feature extraction method.

4.2 形容詞，形容動詞の抽出

コメントから印象を分類するための方法として，コメント中に存在する単語に注目することが考えられる．たとえば，切ない印象を与える楽曲動画に対しては，視聴者は “泣ける” や “悲しい” といった単語を含むコメントを付与すると考えられる．

このための方法として，コメントを形態素解析器を用いて単語集合に分割し，各単語を素性として用いる手法が考えられる．しかし，単語の中には楽曲の印象と関連しないものも多く存在し，そのような単語を素性として抽出してしまうと，過学習の原因となってしまい分類性精度の低下を引き起こすと考えられる．一方，3.2 節で作成した印象クラスや Hu らの作成した Mood Tag Dataset もそうであるように，印象を表現する際には形容詞（および形容動詞）を用いる場合が多い．つまり，単語集合の中でも品詞情報に着目し，形容詞や形容動詞の単語のみを用いることで印象と関連深い素性を抽出することができると考えられる．

実際の素性の抽出には，対象とする動画に投稿された視聴者コメント集合を形態素解析し，得られた形態素集合の中から品詞が形容詞または形容動詞であるもののみを抽出し，各形態素の原型を素性として抽出した．

4.3 繰返し文字の正規化

4.2 節では，形態素解析を用いて時刻同期コメントを単語集合に分割し，形容詞，形容動詞を抽出する手法について説明した．しかし，ニコニコ動画のように，ユーザが自由な形式で即応的にコメントを投稿可能な仕組み上では，形態素解析器がうまく働かないことも多い．たとえば，“かっこいい” という意味を表すコメントを投稿する場合を考えてみても，“かっけえええ”，“かっけえええええええええ”，“かっっっけえええ” といった，多様なコメントが投稿される．こうした，形態素解析器がうまく働かないようなコメントを扱う最も単純な方法は，コメント文字列そのままを 1 つの素性として分類に用いることである．しかし，コメント文字列そのままを 1 つの素性として扱ってしまうと，

学習データが疎になり，また，印象と関連しないようなコメントも素性としてしまうため，やはり分類精度低下を引き起こすと考えられる．一方で，“かっけえええええええ” や “かっっっけえええ” といったコメントを同一の素性として扱うことができれば，コメントの形態素解析からは得られない有用な素性が抽出でき，分類精度向上が期待できる．

上記の問題を解決するために，本稿では Brody らの手法 [16] を利用する．Brody らは，マイクロブログの 1 つである Twitter [*6] 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，“cooooolllll” 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，ある種の単語はこのように単語中の文字が繰り返されて Twitter 上に投稿されていることを指摘している．さらに，そうした特定の文字が繰り返されるような単語は，ユーザの感情と強く関連していることに着目し，文字の繰返し構造を持つ単語のみを利用した Twitter の感情（ポジティブおよびネガティブ）分析手法を提案している．彼らの研究では，Twitter に投稿された英語文を感情分析の対象としていたが，本研究で扱うニコニコ動画も，彼らの指摘した文字の繰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．

実際に Brody らの手法を利用した，コメントからの素性抽出手法の流れを以下に示す．まず，以下の流れに従って，素性の候補集合を得る．

(1) 訓練データ内の楽曲動画中のすべてのコメントに対して，日本語以外の文字を除去し，拗音を大文字に変換する前処理を行っておく（例 “すっげぇぇぇ www → “すっげえええ””）．
(2) (1) で得られたコメント集合に対して，連続する同文字を 1 文字として正規化したコメント（例 “すっげえええ” → “すっげえ”，“かわいいいいい” → “かわい” など）の集合を得る（以降，“すっげえ” や “かわい” などを正規化コメントと呼ぶ）．
(3) (2) で得られた正規化コメント集合のうち，正規化コメントの元となったコメント内の文字の繰返し数に着目し，繰返し数が 3 文字に満たない正規化コメントを集合から除去する．たとえば，“このまま” というコメントのみから “このま” という正規化コメントが得られているとすると，この正規化コメントは “ま” が 2 回繰り返されたコメントのみから生成されているため除去する．
(4) (3) で得られた正規化コメント集合を，素性の候補集合とする．

そして，実際の素性抽出段階では，対象となっているコメントすべてについて，その正規化コメントを求め，それが上記で得られた正規化コメント集合に含まれていれば，その正規化コメントを素性として採用する．

4.4　楽曲のサビ区間の利用

4.2 節および 4.3 節で述べた手法は，時刻同期コメント集合を，コメントが付与された動画の再生時刻によらずに等しく扱った手法ととらえることができる．しかし，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．たとえば，最も分かりやすい例では，実際の楽曲開始前に付与されたコメントは楽曲の印象と関連した内容である可能性は低い．一般的に，楽曲は “A メロ”，“B メロ”，“サビ” といったような，メロディのいくつかのまとまりに分解することができ，その繰返しにより構成される．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 [17] 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．つまり，楽曲のサビ区間中に投稿されたコメントは，楽曲の印象と関連深いコメントが多く，そうしたコメントから素性を抽出することで印象推定の精度が向上するのではないかと考えられる．

本稿では，楽曲のサビ区間検出手法として Goto により提案された手法 “RefraiD” [17] を用いた．RefraiD は，楽曲の断片的な繰返し区間の相互関係を調べながらサビ区間を求める手法（複数区間がサビとして出力）であり，80%の楽曲についてサビ区間を正しく検出可能な，非常に高精度なサビ区間検出手法である．

実際の手法は，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 RefraiD によりサビとして抽出された区間に入っている場合に，4.2 節および 4.3 節で述べた抽出方法により素性を抽出し，分類として用いる．このとき，サビ区間を用いた手法と，4.2 節および 4.3 節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．

4.5　素性の重み付け

各素性の重みは Hu らの研究の歌詞特徴量の重み付け [6] を参考に，TF-IDF に基づく重み付けを行った．訓練対象とする動画集合を V とするとき，分類対象の動画 v における素性 w の重みを $\mathrm{tf}_{w,v} \log \frac{|V|}{\mathrm{df}_{w,V}}$ として求める．ここで，$\mathrm{tf}_{w,v}$ は動画 v 中のコメント集合において素性 w が出現する回数，$\mathrm{df}_{w,V}$ は動画集合 V 中で素性 w が現れるコメントを含む動画数を表す．

5.　評価実験

4 章で述べた素性抽出手法の有用性を評価するため，分

[*6] http://twitter.com

表 2　5 章での印象分類実験に使用したデータセット
Table 2　Dataset used in experiments in Section 5.

印象クラス名	動画数	印象クラス名	動画数
cute	365	cool	133
sorrow	239	aggressive	78
cheerful	224	darkness	71
fresh	204	–	–

類精度の評価実験を行った．本章での実験の目的は，時刻同期コメントからの素性抽出方法として，どのような手法が良いのかを明らかにすることである．

5.1　データセット

実験にあたり，3.2 節で作成した楽曲印象データセットのうち，コメントが 200 件以上付与されている楽曲動画のみを対象とし，対象となる動画数が 30 本以上存在する印象クラスのみを扱った．**表 2** に対象とした印象クラスおよびそのクラスに属する動画数を示す．表から分かるように，本章の実験では 7 印象クラス，計 1,314 本の楽曲動画を対象とした．

5.2　手法

視聴者の時刻同期コメントからの素性抽出手法として，4.2 節で述べた，形容詞，形容動詞を抽出する手法（**adj** 手法），4.3 節で述べた，繰返し文字を正規化する手法（**rep** 手法），4.4 節で述べた楽曲のサビ区間中のコメントを利用した手法（**sabi** 手法）およびそれらの手法の組合せで得られる素性による分類精度の評価を行う．また，ベースライン手法として，(1) コメントを形態素解析して得られる名詞，形容詞，形容動詞，動詞すべてを用いる手法（**allpos** 手法），(2) コメント文字列をそのまま素性として用いる手法（**allcomment** 手法）の 2 つを用意した．

5.3　実験設定

多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 [*7] 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．また，形態素解析には日本語形態素解析器である MeCab [*8] を用いた．

分類性能の評価尺度には F 値を用いた．あるクラスにおける F 値は，そのクラスに関する分類の適合率を P，再現率を R とするとき，それらの調和平均 $\frac{2PR}{P+R}$ として求められる．今回は，実際の印象推定手法を用いたアプリケーションとしては，適合率と再現率の両者が重要であると考え，この F 値を採用した．また，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 [18]（本稿ではそれぞれ，“マクロ F 値”，“マイクロ F 値” と呼ぶ）を用いた．マクロ F 値は各クラスで得られた F 値の平均をとった値であり，各クラスの重要性を等しく扱った尺度である．また，マイクロ F 値は各クラスに所属するメンバ数を考慮した F 値である．そして，表 2 に示した動画集合に対して，5 分割交差検定を行い評価値を求め，それを分割を変更しながら 10 回試行し，各値の平均値を求めた．

[*7] http://www.csie.ntu.edu.tw/~cjlin/libsvm/
[*8] http://mecab.googlecode.com/svn/trunk/mecab/doc/

5.4　結果

5.2 節で述べた各手法での分類結果を**表 3** に示す．表中の “+” で表された手法は，それぞれの手法で得られる素性を組み合わせた手法を表している．たとえば，**adj+rep** 手法は **adj** 手法から得られる素性と **rep** 手法から得られる素性を学習および分類に用いる手法を表している．

まず，**adj**，**rep**，**sabi** の 3 手法それぞれ単体のみを用いた手法の分類精度を比較してみると，表 3 より，コメントの形容詞を用いる **adj** 手法がマクロ F 値，マイクロ F 値ともに他の 2 手法と比べて高い分類精度となっていることが分かる．また，**adj** 手法は 2 つのベースライン手法と比較しても高いマクロ F 値，マイクロ F 値となっており，コメント中の形容詞が楽曲動画の印象推定を行ううえで重要な素性となっていることが分かる．

次に，**rep** 手法の効果について見てみる．**rep** 手法単体での分類精度は **adj** 手法よりも低いものの，両者を組み合わせた **adj+rep** 手法は **adj** 手法よりもマクロ F 値，マイクロ F 値ともに高い分類精度となっていることが分かる．**adj** 手法と **rep** 手法についてより詳しく見てみるため，いくつかの印象クラスにおいて，素性の重みが高かった素性を著者らがいくつか選択して表示したものを**表 4** に示す．たとえば，cute クラス中の素性を見てみると，**adj** 手法では “かわいい” という形容詞が，一方 **rep** 手法では “かわい”，“かわえ” や “かあい” などに正規化されるようなコメント素性として抽出されていることが分かる．この例から分かるように，繰返し文字を正規化したコメントを素性として用いることで，“かわえええぇぇぇぇ” や “かぁぁいいいいいい” のような，形態素解析による単語集合的なアプローチではうまく扱うことができないようなコメントも素性として抽出できることが分かる．このような理由のため，**adj** 手法と **rep** 手法の両者の素性を組み合わせることで，より分類精度が向上したと考えられる．また，**rep** 手法の各印象クラスごとの分類精度を見てみると，cute クラスや aggressive クラスで比較的高い分類精度となっており，cheerful クラスや fresh クラスでは低い分類精度となっていることが分かる．これは，**rep** 手法が落ち着いた雰囲気の印象クラスに対して有効に働かないためではないかと考えられる．**rep** 手法は “かっっけえええええええ” のよう

表 3　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表している
Table 3　Comparison of F-measure for methods. Highest value among methods are indicated in bold.

		ベースライン手法		提案手法				
		allpos	**allcomment**	**adj**	**rep**	**sabi**	**adj+rep**	**adj+rep+sabi**
印象クラス	cute	.667	.558	.729	.692	.678	.747	**.758**
	sorrow	.675	.452	.716	.600	.567	.710	**.721**
	cheerful	.523	.226	**.568**	.347	.408	.525	.535
	fresh	.527	.319	**.662**	.394	.438	.602	.630
	cool	.626	.463	.625	.576	.465	**.659**	.654
	aggressive	.742	.638	.529	**.760**	.653	.755	.749
	darkness	.521	.313	.560	.520	.470	**.597**	.573
マクロ F 値		.612	.424	.627	.555	.526	.656	**.659**
マイクロ F 値		.619	.437	.661	.556	.540	.667	**.676**

表 4　印象クラスに特徴的な素性の例
Table 4　Example features in three mood classes.

印象クラス = **cute**		印象クラス = **aggressive**		印象クラス = **fresh**	
adj 手法	**rep** 手法	**adj** 手法	**rep** 手法	**adj** 手法	**rep** 手法
かわいい	かわい	かっこいい	かっこい	爽やか	もっと評価されるべき
うまい	かわえ	かっこよい	かっけえ	さわやか	いなあ
仕方	ミクーン	やばい	もっと評価されるべき	かわいい	いねえ
楽しい	キュン	カッコイイ	れえん	きれい	あらかわい
あつい	めーちやあん	大好き	かっけー	元気	これはい
甘い	みくー	かわいい	りん	綺麗	きれい
大好き	うめえ	綺麗	すげえ	気持ちいい	なぜ伸びない
幸せ	めえちやあん	可愛い	きやあ	大好き	ありがとう

な，特定の文字が繰り返されるコメントを正規化する手法であったが，このように，特定の文字を繰り返すようなコメントは，視聴者の強い感情を表すために行われると考えられる．逆に，落ち着いた雰囲気の楽曲ではそのようなコメントはあまり投稿されず，**rep** 手法では印象に関連した良い素性が抽出できなかったために，結果としてそのような cheerful クラスや fresh クラスのような印象クラスでの分類精度が低くなったのではないかと考えられる．

sabi 手法について見てみると，**sabi** 手法単体での分類精度は **adj** 手法，**rep** 手法および **adj+rep** 手法よりも低くなっていることが分かる．その理由として，**sabi** 手法はサビ区間のみのコメントを素性として使うため，**adj+rep** 手法と同様の素性を抽出するものの，単体で用いただけでは十分な量の素性が得られないためではないかと考えられる．しかし，すべての素性を組み合わせた **adj+rep+sabi** を見てみると，**adj+rep+sabi** 手法はマクロ F 値で 0.659，マイクロ F 値で 0.676 と他の手法と比較して最も高い分類精度を達成していることが分かる．このことは，**sabi** 手法は単体では不十分ではあるが，サビに注目することで印象と関連した素性を抽出することができ，結果としてすべてを組み合わせた **adj+rep+sabi** での分類精度が向上したのではないかと考えられる．**sabi** 手法を組み合わせることで分類精度が向上したという結果は，時刻同期コメントを単なるコメントの集合として扱うだけではなく，音楽情報処理の分野で行われている楽曲構造理解に関する技術と組み合わせることで，より高い分類精度が得られる可能性を示していると考えられる．

表 5　**adj+rep+sabi** 手法によるコメント数と分類精度の関係
Table 5　Relationship between number of comments and classification accuracy of **adj+rep+sabi** method.

	動画に付与された時刻同期コメント数				
	200–271	272–389	390–607	608–1317	1318–
分類精度	.656	.670	.681	.689	.692

5.5　コメント数と分類精度の関係

本手法は，コメントから素性を抽出して印象を推定するため，その精度は動画に付与されるコメント数に依存すると考えられる．そこで，動画に付与されているコメント数と分類精度の関係を調べた．まず，5.1 節に示した楽曲動画を，動画に付与されているコメント数に応じて 5 つのグループに分割した．ここで，各グループ内の動画数が均等となるようにコメント数の範囲を設定した．次に，5.3 節の実験設定で **adj+rep+sabi** 手法により得られた分類精度を各グループごとに評価した．各グループごとの分類精度を**表 5** に示す．表中の分類精度は，各グループの動画集合の中で正しく分類された動画の割合として求めた．

表 5 より，コメント数が最も少ないグループ（コメン

表 6　ある印象クラスの楽曲動画がどの印象クラスに分類されたかを割合で表した表．表中の網掛で表されたセルは，誤った分類かつ値が 0.100 を超えていることを表す

Table 6　Classification results for mood classes. Wrong classification with 0.100 or higher value are indicated with shaded cell.

		推定された印象クラス						
		cute	sorrow	cheerful	fresh	cool	aggressive	darkenss
真の印象クラス	cute	.779	.038	.105	.048	.010	.005	.014
	sorrow	.057	.731	.037	.075	.077	.004	.019
	cheerful	.217	.051	.496	.133	.034	.003	.064
	fresh	.122	.117	.076	.617	.056	.005	.007
	cool	.026	.144	.010	.079	.685	.033	.022
	aggressive	.015	.056	.004	.028	.204	.680	.012
	darkness	.108	.107	.138	.003	.040	.002	.603

ト数が 200 件から 271 件のグループ）が最も分類精度が低く，コメント数が最も多いグループ（コメント数が 1,318 件以上）が最も分類精度が高くなっていることが分かる．3.3 節では，再生数が 10 万回を超えるような人気の楽曲動画には印象タグが付与されにくいことを示したが，その一方でそうした動画にはコメントが多く付与されているため，再生数が少ない動画よりも高い精度で印象を推定できると考えられる．

5.6　印象クラス間の関連

本稿では，楽曲動画の印象推定を多クラス分類問題として扱っていた．つまり，各印象クラスを排他的なものとしてとらえ，データセット作成および実験を行った．しかし，現実には印象クラスが排他的ではなかったり，どの印象クラスに属するかが曖昧であったりする．たとえば，3.2 節で述べたように，複数の印象クラスのタグが付与されているような楽曲動画もごくわずかではあるが存在（“VOCALOID” タグが付与された楽曲中の 0.3%）した．

印象クラス間の関係をより分析するために，提案手法で楽曲動画がどのように分類されたのかを分析した．**表 6** は，**adj+rep+sabi** 手法において，ある印象クラスの楽曲動画が，どの印象クラスに分類される割合が多かったのかを示した表である．表中の行要素は楽曲動画の真のクラスを表し，列要素は手法が推定したクラスを表す．たとえば，表中で (cute，sorrow) 要素の値が 0.038 というのは，本来 cute クラスに属する動画が sorrow クラスと分類された割合が 3.8%であった，ということを意味している．また，表中の網掛で表示された要素は分類が誤っており，かつその割合が 10%を超えていることを表している．

表 6 を見てみると，cheerful クラスは cute クラスに分類されることが多いことが分かる．この理由は，cheerful クラスと cute クラスが提案手法の精度や実験に用いたデータセットにおける動画数の偏りなども考えられるが，1 つの楽曲動画から，この 2 クラスの印象を受けやすいということも理由の 1 つに考えられる．cheerful クラスは，楽曲を聴いた際に “元気が出る” ような印象を与える楽曲動画に関するクラスであり，cute クラスのように “かわいい” と感じる楽曲動画を視聴した際に “元気がでる” と感じることも十分に考えられる．

今回は分類精度の評価を行いやすくするために，複数の印象クラスのタグが付与されているような動画を評価の対象とはしなかったが，ユーザの多様な情報要求に応えるためには，より柔軟な印象推定モデルが必要となると考えられる．たとえば，各クラスごとにバイナリ分類器を作成し，その出力を統合して最終的に印象を決定するようなモデルが考えられる．

6.　歌詞，音響特徴量との分類精度比較

5 章の実験より，視聴者の時刻同期コメント中の形容詞，繰返し文字，楽曲のサビ区間に着目した素性を組み合わせた手法が最も高い分類精度を達成したことが分かった．本章では，印象推定における視聴者の時刻同期コメントの有用性をさらに示すため，楽曲の歌詞や音響特徴量に基づく分類手法との分類精度比較を行う．また，コメントとそうした手法の組合せについても精度を分析する．

以降，歌詞および音楽特徴量の構築方法について説明した後，分類精度の評価について述べる．

6.1　歌詞特徴量

楽曲動画の歌詞については，我々が 2012 年 6 月 1 日から 6 月 30 日の期間に，「VOCALOID 楽曲データベース」[*9] から収集した，ニコニコ動画に投稿されている楽曲動画の歌詞情報 18,045 件中の情報を用いた．

歌詞からの素性抽出については，Hu らの研究 [6] と同様に，歌詞を形態素解析して得られる名詞，形容詞，動詞の集合を用いた．また，各素性の重みも 4.5 節および彼らと同様，TF-IDF による重み付けを用いた．

6.2　音響特徴量

音響特徴量の抽出については，楽曲分析において広く用いられている MARSYAS [19] を利用した．MARSYAS はスペクトル特徴量やメル周波数ケプストラム係数，テンポ情報などの特徴が抽出可能であり，そうした特徴量は楽曲のジャンル推定や印象推定など楽曲分析に広く利用されている．本稿では，楽曲動画の楽曲から，上記にあげた特徴量を含む，MARSYAS の初期設定で得られる 52 次元の特徴を抽出した．そして，抽出された特徴ベクトルに対して主成分分析を適用し，主成分で張られる空間における 52 次元のベクトルを分類のための素性とした．

6.3　データセットおよび実験設定

5 章で用いたデータセットのうち，6.1 節で述べた歌詞

[*9] http://www5.atwiki.jp/hmiku/

表 7　6 章での印象分類実験に使用したデータセット
Table 7　Dataset used in experiments in Section 6.

印象クラス名	動画数	印象クラス名	動画数
sorrow	166	cool	111
cute	159	cheerful	87
fresh	150	aggressive	46

表 8　各手法における F 値．太字で表された数値は手法間での最大値を表している
Table 8　Comparison of F-measure for methods. Highest value among methods are indicated in bold.

		comment	lyric	audio	l+a	c+l+a
印象クラス	cute	**.766**	.500	.506	.515	.762
	sorrow	.725	.498	.480	.563	**.739**
	cheerful	.420	.205	.070	.240	**.444**
	fresh	.630	.467	.208	.463	**.657**
	cool	**.673**	.437	.314	.468	.668
	aggressive	.691	.088	.126	.100	**.711**
マクロ F 値		.651	.365	.301	.392	**.664**
マイクロ F 値		.673	.441	.380	.467	**.687**

情報中に歌詞が存在する楽曲動画のみを実験に用いた．本章での実験に用いたデータセットを**表 7** に示す．表 7 に示すように，本章では 6 印象クラス，計 719 本の動画を実験に用いた．その他の実験設定は 5.3 節で述べた方法と同様である．

6.4　結果

時刻同期コメントから得られる特徴量，歌詞特徴量，音響特徴量に基づく分類結果を**表 8** に示す．表中の **comment** は 5 章で述べた **adj+rep+sabi** 手法，**lyric** は 6.1 節で得られた素性を用いた手法，**audio** は 6.2 節で得られた素性を用いた手法をそれぞれ表す．また，**l+a** は **lyric** と **audio** 手法の組合せ，**c+l+a** は **comment**，**lyric**，**audio** 手法の組合せをそれぞれ表している．

表 8 より，視聴者の時刻同期コメントを用いた **comment** 手法は，すべての印象クラスにおいて歌詞情報を用いた **lyric** 手法や音響特徴量を用いた **audio** 手法の分類精度を上回っていることが分かる．これは，時刻同期コメントは楽曲動画の視聴中に投稿されるため，その楽曲に対する印象を強く反映しており印象推定に有用であるということを示していると考えられる．また，**l+a** 手法は **lyric** 手法および **audio** 手法よりも高いマクロ F 値およびマイクロ F 値が向上しており，これは Hu らの結果と一致している．最後に，表よりすべての手法を組み合わせた **c+l+a** 手法が最も高いマクロ F 値とマイクロ F 値を達成していることが分かる．このことは，時刻同期コメント，歌詞，音響というそれぞれ異なる種類の情報源を組み合わせることで，分類精度が向上する可能性があることを示している．

7.　考察

最後に，提案手法の現在の課題について整理する．提案手法の分類精度を改善し，また，提案手法の有用性をさらに明らかにするために，3 つの取り組むべき課題があると考えている．

1 つめは視聴者の付与したコメントが，楽曲に対するものなのか映像に対するものなのかを判定する技術である．今回提案した手法は，印象に関するコメントはすべて楽曲に対するものであるという仮定をして素性を抽出している．しかし，ニコニコ動画上に投稿される楽曲動画にはいくつかの種類があり，静止画像が 1 枚表示されているだけの動画もあれば，ユーザが作成した PV（プロモーション映像）をともなう動画も存在する．そのため，動画に付与されるコメントが，楽曲の印象ではなく映像の印象に関するコメントであることもある．実際に，本稿での実験でも，aggressive クラスに属する楽曲動画が，その動画に付与されている映像に対して多くの視聴者が “かわいい” とコメントしていたため，cute クラスに判定されるという例が存在した．より精度良く楽曲の印象を推定するためには，コメントが何に対する反応なのかを判定する技術について取り組む必要がある．

2 つめはデータセットの作成方法である．本稿では，視聴者が付与したタグを基に正解セットを作成し，評価実験に用いた．しかし，タグとコメントは相互に独立したものではなく，場合によっては，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，因果関係も少なからず存在すると考えられる．こうしたタグとコメントの関係とは独立に提案手法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のではなく，手動でデータセットを構築する必要がある．

最後は，提案手法の適用範囲についてである．6.4 節の実験結果から，時刻同期コメントが歌詞や音響特徴量と比べて高い分類精度を達成することが分かった．しかし，時刻同期コメントを用いる手法は，歌詞や音響特徴量などを用いた手法にはない欠点も存在すると考えられる．時刻同期コメントを用いる手法の大きな問題点は，分類に十分な量のコメント数が必要なことである．本稿では，200 件以上のコメントが持つ動画を分類対象に用いたが，投稿されて日が浅い動画はそのような数のコメントが付与されていないことも多く，そのような動画に対しては提案手法がうまく働かないと考えられる．一方で，音響特徴量に基づく分類は，時刻同期コメントを用いる場合と比較して分類精度が低いものの，楽曲動画が投稿されてすぐに分類結果を得ることができる．今後，実際のアプリケーションとして，より多くの楽曲動画に対して印象推定を行うためには，こうしたさまざまな特徴量を組み合わせた分類手法が必要と

なってくると考えられる．たとえば，楽曲動画が投稿された初期の段階では楽曲特徴量に基づく分類をしておき，十分な時刻同期コメントが動画に投稿されると本稿での提案手法による分類結果を用いる，といった組合せが必要となってくるだろう．

8. まとめと今後の課題

本稿では，楽曲動画に付与された視聴者の時刻同期コメントに基づいて，楽曲動画を可愛らしい，切ない，元気がでる，爽やかといった印象に分類する手法を提案し，実験により有用性を示した．提案手法では，時刻同期コメントから（1）形容詞，形容動詞，（2）繰返し文字の正規化，（3）楽曲のサビ区間中のコメントという 3 つの手法を用いて素性を抽出した．実験では，7 印象クラス計 1,314 本の楽曲動画を対象に実験し，3 つの素性抽出手法を組み合わせた手法がマクロ F 値，マイクロ F 値ともに最も高い精度を達成することを示した．

今後は，7 章であげた課題を解決し，印象に基づく楽曲動画の検索を実現していく予定である．また，今回はデータセット作成の容易さから VOCALOID に関する楽曲動画を扱ったが，本手法はその他の楽曲動画に対しても適用できると考えられる．今後はより多くの楽曲動画を対象に実験する予定である．

最後に，より高レベルな印象に基づく楽曲検索の実現についても取り組んでいく予定である．楽曲から受ける印象としては，本研究で扱ったもののほかに，“昭和っぽい楽曲” や “お酒が飲みたくなる楽曲” なども考えられる．こうした印象は “元気が出る” や “爽やか” といった印象とは異なり，1 語の形容詞や形容動詞では表現することのできない，より高度な印象クラスではないかと考えられる．こうした印象に基づく検索を実現するためのアプローチとして，動画に付与されたタグに注目することを考えている．楽曲動画の中には，数は少ないものの “お酒が飲みたくなる曲” というタグが付与されていることもある．こうした情報を手がかりとすることで，より高度な印象に基づく楽曲検索が実現できるのではないかと考えられる．

謝辞　本研究を進めるにあたり，産業技術総合研究所の後藤真孝氏，濱崎雅弘氏，石田啓介氏に貴重なご助言をいただきました．また，本研究の一部は，科学技術振興機構 OngaCREST プロジェクトと，文部科学省科学研究費補助金若手研究（A）（研究代表者：中村聡史，課題番号 23680006）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] 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），Vol.53, No.5, pp.472–476 (2012).

[2] 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），Vol.53, No.5, pp.466–471 (2012).

[3] Hu, X., Downie, J., Laurier, C., Bay, M. and Ehmann, A.: The 2007 MIREX audio mood classification task: Lessons learned, *Proc. 9th International Conference on Music Information Retrieval*, pp.462–467 (2008).

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

[5] Laurier, C., Grivolla, J. and Herrera, P.: Multimodal music mood classification using audio and lyrics, *7th International Conference on Machine Learning and Applications*, pp.688–693 (2008).

[6] Hu, X., Downie, J. and Ehmann, A.: Lyric text mining in music mood classification, *Proc. 10th International Society for Music Information Retrieval Conference*, pp.411–416 (2009).

[7] Hevner, K.: Experimental studies of the elements of expression in music, *The American Journal of Psychology*, Vol.48, No.2, pp.246–268 (1936).

[8] Russell, J.: A circumplex model of affect., *Journal of Personality and Social Psychology*, Vol.39, No.6, pp.1161–1178 (1980).

[9] Hu, X., Bay, M. and Downie, J.: Creating a simplified music mood classification ground-truth set, *Proc. 8th International Conference on Music Information Retrieval*, pp.309–310 (2007).

[10] Nakamura, S. and Tanaka, K.: Video search by impression extracted from social annotation, *Proc. 10th International Conference on Web Information Systems Engineering*, pp.401–414 (2009).

[11] 佃　洸摂，中村聡史，山本岳洋，田中克己：映像に付与されたコメントを用いた登場人物が注目されるシーンの推定，情報処理学会論文誌「情報爆発」特集号，Vol.52, No.12, pp.3471–3482 (2012).

[12] Yoshii, K. and Goto, M.: MusicCommentator: Generating comments synchronized with musical audio signals by a joint probabilistic model of acoustic and textual features, *Proc. 8th International Conference on Entertainment Computing*, pp.85–97 (2009).

[13] 中村聡史，山本岳洋，後藤真孝，濱崎雅弘：視聴者反応と音楽的特徴量を用いたサムネイル動画の自動生成，Web とデータベースに関するフォーラム（WebDB Forum 2012），A2-3 (2012).

[14] 戀塚昭彦：ニコニコ動画の創造性：動画コミュニティサービス「ニコニコ動画」の 5 年間，情報処理（情報処理学会誌），Vol.53, No.5, pp.483–488 (2012).

[15] Heymann, P., Koutrika, G. and Garcia-Molina, H.: Can social bookmarking improve web search?, *Proc. 1st International Conference on Web Search and Web Data Mining*, pp.195–206 (2008).

[16] Brody, S. and Diakopoulos, N.: Cooooooooooooooollllllllllll!!!!!!!!!!!!!!: Using word lengthening to detect sentiment in microblogs, *Proc. Conference on Empirical Methods in Natural Language Processing*, pp.562–570 (2011).

[17] Goto, M.: A chorus section detection method for musical audio signals and its application to a music listening station, *IEEE Trans. Audio, Speech and Language Processing*, Vol.14, No.5, pp.1783–1794 (2006).

[18] Manning, C., Raghavan, P. and Schütze, H.: *Introduction to Information Retrieval*, Cambridge University

Press (2008).
[19] Tzanetakis, G. and Cook, P.: MARSYAS: A framework for audio analysis, *Organised sound*, Vol.4, No.3, pp.169–175 (1999).

山本 岳洋　（正会員）

1984 年生．2011 年京都大学大学院情報学研究科博士後期課程修了．同年日本学術振興会特別研究員（PD），2012 年京都大学大学院情報学研究科特定研究員，現在に至る．博士（情報学）．主に情報検索，特に情報検索におけるユーザインタラクションに関する研究に従事．日本データベース学会会員．

中村 聡史　（正会員）

1976 年生．2004 年大阪大学大学院工学研究科博士後期課程修了．同年独立行政法人情報通信研究機構専攻研究員．2006 年京都大学大学院情報学研究科特任助手，2009 年同特定准教授，2013 年明治大学総合数理学部准教授，現在に至る．博士（工学）．サーチとインタラクションや，情報曖昧化技術，ソーシャルアノテーション分析等の研究活動に従事．ヒューマンインタフェース学会等各会員．

（担当編集委員　喜田 拓也）