

ユーザーインタラクションを用いた  
楽曲の構造理解システム  
—ボーカロイドとニコニコ動画を題材に—

矢倉 大夢



# 目次

目次	3
第 I 章 はじめに	5
第 II 章 楽曲構造理解技術とニコニコ動画	7
1 楽曲構造理解技術 . . . . .	7
2 ニコニコ動画における楽曲構造理解技術 . . . . .	8
3 ボーカロイドとニコニコ動画 . . . . .	9
第 III 章 提案手法	11
1 前提 . . . . .	11
2 概略 . . . . .	12
3 検出手法 . . . . .	13
4 実装 . . . . .	15
5 過去の手法との比較 . . . . .	16
第 IV 章 評価実験	17
1 データセット . . . . .	17
2 評価手法 . . . . .	17
3 評価結果 . . . . .	17
第 V 章 おわりに	19
参考文献	21



# 第1章

## はじめに

インターネットの普及は、全世界に新たな自己表現の手段をもたらした。ブログの流行とともに、インターネット上で自分の考えを表現することが一般的になり、さらに、Twitter や Facebook などのソーシャルメディアの台頭とともに、そのハードルが大幅に低くなった。

インターネットがもたらしたものはそれだけではない。創作活動のハードルも大幅に低くなったのである。プロでなくとも、自らの作品を Web に投稿すれば、物理的距離など関係なしに、多くの人からフィードバックを得ることができるようになった。そして、こうした創作活動を支えているのが、数々のユーザー投稿型 Web サイトである。

その中でも、ニコニコ動画は、日本発のサービスとして最も有名であろう。ニコニコ動画は、Web サイト上で再生される動画に対してリアルタイムでコメントを付けられるサービスであり、2006 年から運営されている。コメントによって視聴体験を共有するという新たな楽しみが得られることや他の人からのフィードバックが得やすいということもあり、当初から人気を誇り、たくさんの動画が投稿されてきた [16]。2013 年 6 月末時点で、ユーザー数が 3415 万人、課金ユーザーが 200 万人を超えており [9]、日本最大の動画投稿プラットフォームであることは間違いないだろう。

また、楽曲の構造理解は近年研究されているテーマの 1 つである。その背景の 1 つとしては、音楽の販売プラットフォームのクラウド化に伴って、ユーザーの嗜好分析やリコメンデーションシステムを目的として構造理解システムが用いられるようになったことが挙げられよう。そのため、リズム分析やコード進行の分析などは様々なアプローチがなされてきた。また、そういった研究を応用した例として、マイクから取得した音声から流れている音楽を検索するモバイルアプリケーションや自動で譜面を生成するリズムゲームなどが挙げられる。

しかしながら、未だサビの検出については、高い精度が得られているとはいえない。その理由として、人間がサビを見つけ出す際はリズムやコード進行だけでなく歌詞のメッセージ性の強さなども考慮しているのに対し、プログラムでの歌詞の意義理解は非常に困難であるということが挙げられる。

## 第I章 はじめに

そこで、この研究ではニコニコ動画を題材として、その最も特徴的な機能であるコメントを利用することにより、歌詞を直接解析することなしに、メッセージ性の強い部分や盛り上がっている部分、いわゆる「サビ」を高精度で検知する手法を開発した。

## 第 II 章

# 楽曲構造理解技術とニコニコ動画

### 1 楽曲構造理解技術

楽曲構造理解技術とは、音響信号を解析することによって楽曲の繰り返し区間やビート構造、メロディラインやコード進行などを解析する技術のことである。日本では 1990 年代後半から様々な研究がなされ、HMM(隠れマルコフモデル)と呼ばれる手法や、音響信号から特徴ベクトルを抽出して非負値行列因子分解・K 近傍法などの機械学習アルゴリズムを用いる手法が提案されてきた。[18][19][7][8]

これらの研究においては、すでにかかなりの精度で検知できる手法が確立されている。例えば、齊藤らによる調の推定手法 [19] では、RWC 研究用音楽データベース [17] の楽曲に対して、93.8% の検出精度を誇っている。他にも、メロディラインの検出についても後藤の研究 [2] では、ポピュラー音楽に対しては 89.1% の精度を誇っている。

しかしながら、サビの検出については、未だ高い精度を得られる手法が確立されているとは言い難い。現在、唯一のサビ検出手法となっている、後藤らによる RefrainD[3] においては、80.0% の精度にとどまっている。この手法では、以下の 3 つの点を仮定することによって、音響信号の解析によって得られた繰り返し区間の情報からサビを検出している。

1. サビの区間は、毎回ほぼ類似したテンポで、一定の長さの区間として繰り返し演奏される。
2.  $((\Rightarrow A \text{ メロ} [\Rightarrow B \text{ メロ}]) \times n_1 \Rightarrow \text{サビ}) \times n_2$  に相当するような、長い区間での繰り返しがある場合、その末尾の部分がサビである可能性が高い。
3. ある繰り返し区間内にさらにその区間の半分程度の短い繰り返し区間がある場合には、それがサビである可能性が高い。

しかし、ポピュラー音楽の中にもこうした仮定から外れるものもあり、そのために精度を上げることが難しいとされている。

別のアプローチとしては、歌詞情報を解析してサビを検出するという手法も挙げられるが、これはさらに困難であると考えられる。理由としてはまず、楽曲からの歌声の抽出が

80% 程度の精度にとどまっている [11] こと、歌声からの歌詞認識の精度が現在の研究では 58% 程度と非常に難しいこと [12], さらに、歌詞の内容を解析して盛り上がりを見つけるには、歌詞の意義解析が必要となることが挙げられる。

## 2 ニコニコ動画における楽曲構造理解技術

ニコニコ動画は前述の通り日本最大の動画投稿プラットフォームであるが、その成長の鍵となっているのがその特殊性、特に以下の 2 点にあると考えられている [4].

1. コメントによって同時に動画を見ているかのように感じる「擬似同期」
2. 1 つの動画からいくつもの派生動画が生まれ、さらにそこから派生動画が生まれる「N 次創作」

このうち、「N 次創作」を促進する手段として、ニコニコ動画ではコンテンツツリーという仕組みが用意されている。コンテンツツリーは、ユーザーが動画投稿の際に創作の元となった親作品がある場合はそれを手動で登録することによって形作られている。このシステムは、ユーザーの創作意欲を刺激することを目的としたもので、以下の様なクリエイター助成プログラムも用意されている。[10]

クリエイター奨励プログラムは、作品の人気度に応じて奨励金が受け取れる仕組みです。子作品の人気に応じて、親作品のクリエイターは「子ども手当」が受け取れるため、通常よりも多くの奨励金を受け取れる可能性が出てきます。

しかしながら、コンテンツツリーの登録は投稿者に任せられているために、親作品の作者が正当なインセンティブを得られない場合もある。そこで、楽曲構造理解技術を用いることによって、自動的にコンテンツツリーを作成することが可能になる。

また、楽曲構造理解技術はユーザーのエクスペリエンス向上にも繋がると考えられる。例えば、コード進行やメロディラインの類似度を判定することによってユーザーが好むであろう楽曲をリコメンデーションすることができ、膨大な数の動画の中から好みの動画を簡単に見つけ出すことができるようになる。

一方、ニコニコ動画にアップロードされている楽曲は一般のポピュラー音楽より構造理解が難しいと考えられている。その理由として、以下の 2 点が挙げられる。

1. ニコニコ動画は、プロの作曲家でなくとも自由に楽曲をアップロードできる場所であり、一般のポピュラー音楽の形式(サビの配置など)に従っていないものが多い。
2. ボーカロイドと呼ばれる歌声合成技術を用いて作成された楽曲が多く、人間の声を対象にしたアプローチだけでは楽曲からの歌声抽出が難しい。



表 II.1 固有名詞を含むタグに対する合計再生数 (降順)

タグ	合計再生数 (回)
東方	1,901,348,088
VOCALOID	1,488,655,423
アイドルマスター	973,098,753
初音ミク	907,016,417
MikuMikuDance	493,647,610
東方手書き劇場	446,687,190
VOCALOID 殿堂入り	442,306,035
東方ヴォーカル	353,426,183
ポケモン	327,294,723
ミクオリジナル曲	323,278,475

### 3 ボーカロイドとニコニコ動画

ボーカロイド (Vocaloid) は、ヤマハが開発した歌声合成技術、及びそれを用いた製品の総称であり、これを用いることでパソコン上で人間らしいボイスを作ることが可能になる。ニコニコ動画においては、ボーカロイドを用いた楽曲がたくさんアップロードされている。

II.1 は、国立情報学研究所から提供されているニコニコデータセット [13] から算出したものである。これを見ると、ニコニコ動画のコンテンツのうちでボーカロイドが占める割合がいかに多いかが一目瞭然であろう。初音ミクは、ボーカロイドを用いた製品のひとつであり、MikuMikuDance も初音ミクから派生したソフトウェアであるため、上位 10 タグのうち、実に半分を占めていることが分かる。

ボーカロイドに関するコンテンツで大変特徴的なのは、前述の「N 次創作」が非常に顕著だという点である。ニコニコ動画では、ある楽曲が流行するとその楽曲を別のボーカロイド製品に歌わせた動画やその楽曲に独自の映像を組み合わせた動画、そしてその楽曲を人間が歌うという、いわゆる「歌ってみた動画」など、派生作品がたくさん投稿される。

そのため、ボーカロイド楽曲にも有効な楽曲構造理解技術が確立されれば、こうした派生作品も自動で判別することができ、特にユーザーエクスペリエンスの改善に繋がることは間違いないだろう。



## 第 III 章

# 提案手法

### 1 前提

提案する手法では、ニコニコ動画の特徴のひとつである「擬似同期」を実現しているコメント機能を使用する。ニコニコ動画においては、楽曲鑑賞の際に盛り上がる部分で多くのユーザーがコメントを投稿する傾向があり、それを抽出することでサビを検知することを目的としたものである。

III.1 は、ニコニコ動画に投稿されているボーカロイド楽曲で最も再生数の多い「みくみくにしてあげる」のコメント数の時間に対する推移を示しており、さらに、背景が濃くなっている部分がサビの区間を表している。この図より、盛り上がり部分はコメント数が顕著であり、さらに、繰り返しごとにコメント数の推移が類似しているというのが分かるであろう。

また、開始の部分と終了の部分に対しては、特徴的なコメント数の増加が見られるが、こ

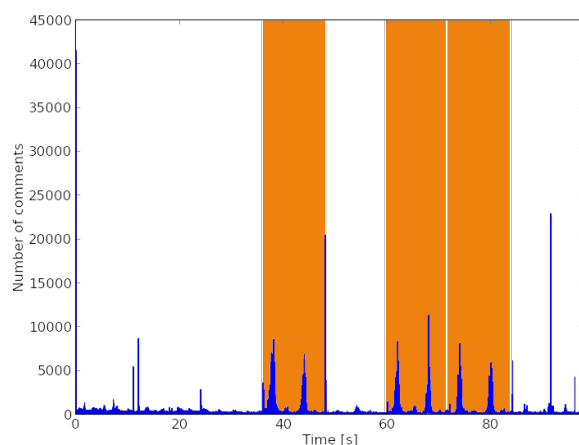


図 III.1 「みくみくにしてあげる」における時間に対するコメント数の推移

### 第 III 章 提案手法

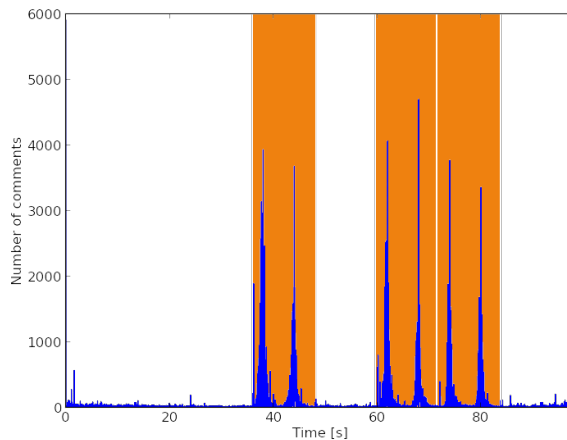


図 III.2 「みくみくにしてあげる」における時間に対する「みっくみく」を含むコメント数の推移

これは、ニコニコ動画において開始と終了の時点で作者に対する賞賛のコメントを残すという習慣があることが影響していると思われる。

そして、ニコニコ動画に特有の現象として「弹幕」と呼ばれるものがある。「弹幕」は、動画の盛り上がる部分や特徴的な部分で、たくさんの類似したコメントが一度に投稿されるという現象を指す。

III.2 は、先程と同じ「みくみくにしてあげる」のコメントのうち、「弹幕」に用いられているこの動画特有のキーワードである「みっくみく」を含むコメント数の推移を示している。この図より、動画特有のキーワードを抽出することがサビ区間の検出精度を高める上で重要であると考えられる。

## 2 概略

提案手法では、まず楽曲内の繰り返し区間を検出する必要があるため、後藤らによる RefraiD[3] を基にした手法を利用した。RefraiD では、III.3 のような流れで繰り返し区間の検出を行なっている。具体的には、単位時間ごとに楽曲を区切り、クロマベクトルの類似度を判定することで類似区間を検出し、さらにそうして検出した類似区間を統合していくことによって、繰り返し区間の検出を行なっている。

そうして検出された繰り返し区間の中から、独自の手法によりスコアを算出することによってサビ区間を選ぶ。

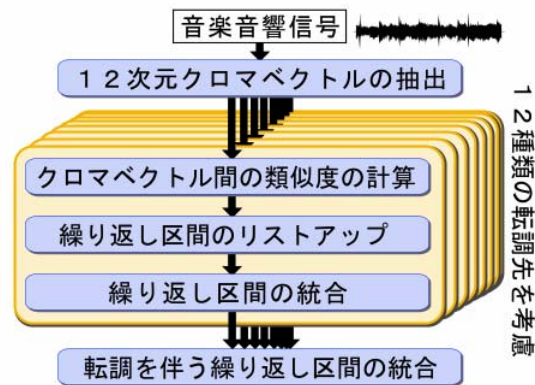


図 III.3 RefraiD[3]における繰り返し区間の検出手法 ([15] より引用)

### 3 検出手法

#### 定義

具体的な検出手法について述べるにあたり，以下の通りに定義する．

1.  $Length$  は，動画の長さ (秒) を表す．
2.  $Comment(x, y)$  は，動画に投稿されたコメントのうち開始  $x$  秒から  $y$  秒の間に投稿されたコメントの集合を表す．
3.  $Noun(c)$  は，コメント  $c$  に含まれる名詞の集合を表す．

#### 手順 1: コメントについて重み付けをする

まず，「弹幕」の抽出を目的にその動画において中心となっているキーワードを抽出する．そのため，すべてのコメントについて形態素解析を行い，出現する名詞を数え上げる．

はじめに，名詞  $v$  についての出現率を以下の様に定義する．

$$C = Comment(5, Length - 5)$$

$$P(v) = \frac{\sum\{\{w \mid w \in Noun(c), v = w\} \mid c \in C\}}{\sum\{|Noun(c)| \mid c \in C\}}$$

集合  $C$  は，開始 5 秒後から終了 5 秒前までの区間に投稿されたコメントの集合である．よって，出現率は  $C$  に含まれるコメントの中で， $v$  の出現数をすべての名詞の出現数で割ったものである．ここで開始からの 5 秒間と終了までの 5 秒間を含めなかったのは，前述の通り開始と終了の部分では盛り上がりと関係なくコメントが投稿される傾向があると考えられるからである．

### 第 III 章 提案手法

そして、コメント  $c$  について、以下のように重み付けをする。

$$Weight(c) = \frac{\sum\{P(v) \mid v \in Noun(c)\}}{|Noun(c)|}$$

これは、コメント  $c$  に含まれる名詞の出現率を平均したものである。ここで、総和ではなく平均としたのはニコニコ動画においては単語数の少ないコメントが多数を占めており、単語数の多いコメントに重点を置くと逆にノイズになると考えられるからである。

#### 手順 2: 各区間についてスコアを求める

検出された繰り返しに属するそれぞれの区間において、どれだけ盛り上がっているかを示すスコアを求める。開始  $a$  秒から  $b$  秒までの区間  $s$  について、そのスコアを以下のように求める。

$$Score(s) = \sum\{Weight(c) \mid c \in Comment(a, b)\}$$

#### 手順 3: 各区間について特徴ベクトルを求める

検出された繰り返しに属するそれぞれの区間において、コメント状況の類似度を求めるための特徴ベクトルを求める。区間を 3 秒ごとに分割し、各分割区間ごとに特徴度を求めてそれを要素とするベクトルを求める。つまり、区間が  $n$  秒である場合、 $\frac{n}{3}$  の分割区間に分けられ、ベクトルの次元は  $\frac{n}{3}$  次元となる。

開始  $a$  秒から  $b$  秒までの区間  $s$  について、その特徴ベクトルを  $\vec{v}(s)$  とする。この時、 $i$  番目の要素を以下のように求める。

$$C_i = Comment(a + (i - 1) \times 3, a + i \times 3)$$

$$\vec{v}_i(s) = \frac{\sum\{Weight(c) \mid c \in C_i\}}{\sum\{Weight(c) \mid c \in Comment(a, b)\}}$$

ここで、集合  $C_i$  は  $i$  番目の分割区間に含まれるコメントの集合を意味している。つまり、特徴ベクトル  $\vec{v}(s)$  の各要素は、3 秒ごとに分割されたそれぞれの分割区間に含まれるコメントの重みが、区間  $s$  に含まれるコメントの重みに占める割合を示す。

#### 手順 4: 各繰り返しについて属する区間の類似度を求める

同じ繰り返しに属する区間  $s, t$  について、その 2 区間の類似度を以下のように求める。

$$Distance(s, t) = \sqrt{\frac{\min\{\dim \vec{v}(s), \dim \vec{v}(t)\}}{\sum_{k=0}^{\min\{\dim \vec{v}(s), \dim \vec{v}(t)\}} |\vec{v}_k(s)^2 - \vec{v}_k(t)^2|}}$$
$$Similarity(s, t) = \frac{1}{Distance(s, t)}$$

これは、ベクトルの距離の逆数であり、距離が近いほど類似度が高くなる。

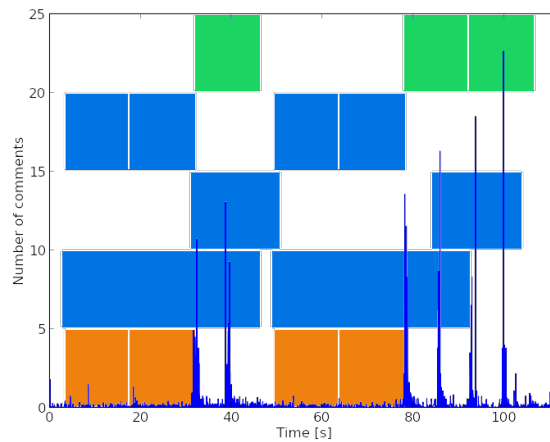


図 III.4 「リンリンリンってしてくりん」における時間に対するコメントの重みの総和の推移

#### 手順 5: 繰り返しに対してスコアを求める

区間  $\{s_1, s_2, \dots, s_n\}$  が同じ繰り返しに属しているとする場合に、その区間の集合を  $S$  とする。そして、検出されたすべての繰り返しについて以下の様にスコアを求める。

$$\begin{aligned} \text{Score}(S) = & \max\{\text{Similarity}(s, t) \mid s, t \in S, s \neq t\} \\ & \times \sum\{\text{Score}(s) \mid s \in S\} \end{aligned}$$

そして、このスコアが最も高い繰り返しをサビとして検出する。

III.4 は、「リンリンリンってしてくりん」において、コメントの重み付けがどのようになるかを示したものだ。背景の 5 段の区間は RefraiD が検出した 5 つの繰り返し区間を示しており、そのうち RefraiD の仮定に基づいた推定ではオレンジの繰り返し区間をサビ区間として検出した。しかしながら、提案手法では緑の繰り返し区間のスコアが最も高くなり、正しくサビ区間の検知ができていることがわかる。

## 4 実装

実装には、Python を用いた。また、コメントの形態素解析には MeCab[5] を使用した。しかし、MeCab にデフォルトで使用されている IPA 辞書は一般的な単語しか収録しておらず、ネットスラングや顔文字などを解析することができない。そこで、新たにユーザー辞書を作成することにより解析精度の向上を図った。

今回は、はてなキーワード及びニコニコ大百科から単語を抽出した。はてなキーワードは Wiki システムの 1 つで、Wikipedia よりユーザーが気軽に項目を作成できるのが特徴である。そのため、インターネット用語もたくさん登録されておりコメントの解析にも有用

### 第 III 章 提案手法

だと判断した。

ニコニコ大百科はニコニコ動画と連動しているサービスのひとつで、これも Wiki システムである。ニコニコ動画では、動画に付与されているタグと連動してニコニコ大百科の記事を参照できるシステムがあり、そのため、ニコニコ動画のコンテンツと密接に関係した単語がたくさん投稿されている。

これらに掲載されている単語を形態素解析の辞書に登録する際には単語のスコアを設定する必要があるため、以下の式を用いて算出した。

$$score = \max\{-36000, 6000 - 200 \times length^{1.3}\}$$

この式は、MeCab のマニュアル [6] に記載されていた以下の式を、インターネット用語固有の長い単語に対応しつつ、分布が広がるように改良したものである。

$$score = \max\{-36000, -400 \times length^{1.3}\}$$

## 5 過去の手法との比較

コメントを使用して動画を解析するという手法は青木ら [14] によって試行されている。しかしながら、青木らの研究では、コメント数の推移とサビ区間の関連性について数例にわたり調査を行ったのみであり、コメントを使用して自動で楽曲構造の解析を行うというのは本研究が初めてである。さらに、提案手法はコメントの重み付けや音響信号を基にした繰り返し区間の検出を行っており、青木らの手法に対して優位性があると考えられる。

また、他のサビ検出手法としては、RefraiD[3] が挙げられるが、前述のとおり、RefraiD は楽曲のサビ区間に対する仮定を基にしてサビ区間を選んでおり、仮定から外れている楽曲での検出精度が低いという問題がある。しかし、提案手法は RefraiD を基にして、さらにコメントを使用することによって盛り上がりの検出を行っているため、さらに高精度な検出が可能になるのではないかと考えられる。



## 第 IV 章

# 評価実験

### 1 データセット

評価実験を行うにあたって、動画情報やコメントデータの取得に国立情報学研究所から提供されているニコニコデータセット [13] を利用した。また、サビ区間データの教師データとして、Songle[1] を利用した。Songle は、楽曲をサビやメロディ、コード、ビートなどの情報と共に動画を鑑賞できるシステムであり、その楽曲構造の解析には RefraiD[3] が用いられているが、ユーザーがサビ情報などを訂正できるシステムもあるため、ユーザーが訂正済みのサビ情報を正解データとした。つまり、RefraiD の解析によるものではなく、ユーザーが Songle 上で登録したサビ情報を正解データとしている。

### 2 評価手法

評価には、「VOCALOID」タグが付けられた動画のうち、再生回数が 10 万回を超えているものを使用した。そのうち、Songle においてユーザーの手によってサビ区間が登録されているものから 100 個を選び、評価を行った。

また、「弹幕」や単語の出現率が盛り上がりに関連があることを確認するために、重み付けなしでの検出、つまりすべての単語の出現率を 1 として検出したものと検出精度の比較を行った。

### 3 評価結果

評価結果は、IV.1 の通りであった。全体での検知率が 89.0% と、80.0% にとどまっていた RefraiD[3] に対して、大きく改善されていることがわかる。また、重み付けの有無によって検知精度が大きく左右されていることから、「弹幕」や単語の出現率が盛り上がり部分と大きく関連しているということも示された。

しかし、5 分以上の動画については、5 分未満の動画と比べてかなり検知精度が劣ってい

#### 第 IV 章 評価実験

表 IV.1 評価結果

重み付けの有無	あり	なし
5分未満の動画 (88 個)	92.0%	78.4%
5分以上の動画 (12 個)	66.7%	50.0%
合計	89.0%	75.0%

ることもわかった。これは、動画が長いと後半でのコメント投稿が少なくなり、盛り上がり部分でもコメント数の増加があまり見られないということが原因だと考えられる。

## 第 V 章

# おわりに

本論文では、ニコニコ動画における楽曲構造理解支援技術として、コメントを用いて楽曲のサビを検出する手法について述べ、その有効性を示した。提案した手法では、ニコニコ動画に投稿されたコメントを特殊な辞書を用いて形態素解析を行い、その出現率を基に重み付けを行うことで、高い精度で楽曲のサビ区間を検出することに成功した。また、類似性の高いコメントが短時間に大量に投稿される「弹幕」と呼ばれる現象が、盛り上がり部分と大きく関連しているということも示した。今後の課題としては、長い楽曲に対しては精度がかなり落ちるという部分についての改善と、「きたあああ」といったニコニコ動画で顕著な語尾を伸ばす表現について、形態素解析ができずに出現率に誤差が生まれている場合があるので、コメントの正規化を行い形態素解析の精度を上げることを考えている。

また、今回は音響信号だけでは構造理解が難しいジャンルとしてボーカロイドを取り上げたが、別のジャンルでも十分に有用な手法だと考えられるので、これから検証を行なっていきたい。

この手法では動画の盛り上がっている部分を検知できるため、楽曲に対する構造理解以外にも有用であると考えられる。例えば、動画要約やユーザーが不快に思いにくい広告表示のタイミングの検出、それ以外にも、コメントの分析による動画間の類似度の算出やそれを用いたりコメンテーションなど、様々な応用が考えられるのでぜひ挑戦していきたい。



## 参考文献

- [1] AIST-Songle-Project. Songle. <http://songle.jp> [Accessed: September 30, 2013].
- [2] Masataka GOTO. A real-time music scene description system : Predominant-f0 estimation for detecting melody and bass lines in real-world audio signals. *Speech Communication*, Vol. 43, No. 4, pp. 311–329, 2004.
- [3] Masataka Goto. A chorus section detection method for musical audio signals and its application to a music listening station. *Audio, Speech, and Language Processing, IEEE Transactions on*, Vol. 14, No. 5, pp. 1783–1794, 2006.
- [4] Masahiro Hamasaki, Hideaki Takeda, and Takuichi Nishimura. Network analysis of massively collaborative creation of multimedia contents: case study of hatsune miku videos on nico nico douga. In *Proceedings of the 1st international conference on Designing interactive user experiences for TV and video*, pp. 165–168, New York, NY, USA, 2008. ACM.
- [5] Taku Kudo. Mecab: Yet another part-of-speech and morphological analyzer. <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html> [Accessed: September 30, 2013].
- [6] Taku Kudo. Mecab の辞書構造と汎用テキスト変換ツールとしての利用. <http://mecab.googlecode.com/svn/trunk/mecab/doc/dic-detail.html> [Accessed: September 30, 2013].
- [7] 伊藤綾, 酒向慎司, 北村正. パラメータ共有 hmm に基づく音響信号からの自動和音認識の検討. 情報処理学会研究報告. [音楽情報科学], Vol. 2010, No. 4, pp. 1–5, nov 2010.
- [8] 角尾衣未留, 小野順貴, 嵯峨山茂樹. リズムマップ: 音楽音響信号からの単位リズムパターンの抽出と楽曲構造の解析 (音響分析一般 (2)). 情報処理学会研究報告. [音楽情報科学], Vol. 2008, No. 78, pp. 149–154, jul 2008.
- [9] 株式会社ドワンゴ, 株式会社ニワンゴ. niconico のプレミアム会員数が 200 万人を突破. <http://info.dwango.co.jp/pdf/news/service/2013/130624.pdf> [Accessed: September 26, 2013].
- [10] 株式会社ニワンゴ. コンテンツツリーについて: ニコニ・コモンズ ヘルプ. <http://help.nicovideo.jp/niconicommons/010/#027904> [Accessed: September 26,

## 参考文献

- 2013].
- [11] 藤原弘将, 後藤真孝, 緒方淳, 駒谷和範, 緒方哲也, 奥乃博. 音楽音響信号と歌詞の時間的対応付け手法: 歌声の分離と母音の viterbi アラインメント. 情報処理学会研究報告. [音楽情報科学], Vol. 2006, No. 90, pp. 37–44, aug 2006.
  - [12] 尾関弘尚, 鎌田貴幸, 後藤真孝, 速水悟. 歌声の歌詞認識における音高の影響について. 日本音響学会 2003 年秋季研究発表会 講演論文集, Vol. 1, No. 1, pp. 637–638, 2003.
  - [13] 国立情報学研究所. 情報学研究データリポジトリ ニコニコ動画コメント等データ. <http://www.nii.ac.jp/cscenter/idr/nico/nico.html> [Accessed: September 26, 2013].
  - [14] 青木秀憲, 宮下芳明. ニコニコ動画における映像要約とサビ検出の試み (セッション 2). 情報処理学会研究報告. [音楽情報科学], Vol. 2008, No. 50, pp. 37–42, may 2008.
  - [15] 後藤真孝. リアルタイム音楽情景記述システム: サビ区間検出手法 (セッション 3: 音楽音響信号処理). 情報処理学会研究報告. [音楽情報科学], Vol. 2002, No. 100, pp. 27–34, oct 2002.
  - [16] 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 (特集 cgm の現在と未来: 初音ミク, ニコニコ動画, ピアプロの切り拓いた世界). 情報処理, Vol. 53, No. 5, pp. 466–471, may 2012.
  - [17] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一. Rwc 研究用音楽データベース: 音楽ジャンルデータベースと楽器音データベース. 日本音響学会研究発表会講演論文集, Vol. 2003, No. 1, pp. 843–844, mar 2003.
  - [18] 武田晴登, 篠田浩一, 嵯峨山茂樹. リズムベクトルを用いたリズム認識. 情報処理学会研究報告. [音楽情報科学], Vol. 2002, No. 63, pp. 23–28, jul 2002.
  - [19] 斉藤翔一郎, 武田晴登, 西本卓也, 嵯峨山茂樹. Specmurt 分析と chroma vector を用いた hmm による音楽音響信号の調認識. 情報処理学会研究報告. [音楽情報科学], Vol. 2005, No. 82, pp. 85–90, aug 2005.