

ニコニコ動画における映像要約とサビ検出の試み

青木 秀憲† 宮下 芳明*

† 明治大学大学院理工学研究科基礎理工学専攻情報科学系

* 明治大学理工学部情報科学科

本稿では、ウェブサービス「ニコニコ動画」でのコメント頻度によってその映像にともなう音楽のサビ部分の検出や映像要約に応用可能なのかを検証すべく、様々なジャンルの映像をとりあげ評価を行った。

A Trial for Video Summarization and Chorus-Section Detecting on Nicovideo

Hidenori AOKI† Homei MIYASHITA*

† Computer Science Course, Graduate School of Science and Technology, Meiji University.

* Department of Computer Science, Meiji University

In this paper, we investigated various movies on Nicovideo, and checked the relation between the frequency of comments and chorus-section in the music used in it. We also proposed the method for video summarization using the frequency of comments on Nicovideo.

1 はじめに

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[1]、500 万人以上のユーザー数を抱えている。このサービス上で入力されたコメントは投稿順に記録され、右から左に向かって流れるように動画上に表示される。携帯電話に対応した「ニコニコ動画モバイル」サービス、外部 SNS に動画を貼り付けられる「ニコニコ外部プレーヤー」など、日々改良と拡張が行われており、今後ますますの発展が予想されている。また、淘汰されていくタグに関する伊藤らの研究[2]や、ユーザ間のコミュニケーションに焦点をあてた川井らの研究[3]なども始まっており、ニコニコ動画を対象とした研究も増えていくと思われる。

論文執筆時点での動画数は約 100 万本、総コメント数は約 10 億件となっているため、単純に平均すればひとつの動画コンテンツに対して 1000 件のコメントがついている計算になる。このコメントづけをアノテーション付与と捉えれば、アップロードされた動画にはかなり充実した量のメタデータが付いているということもできる。本稿では、このコメントを量的に評価したときに、それによって映像コンテンツにおける最も重要な箇所（いわば動画におけるサビ）の判別や映像要約にどれくらい応用可能なのかを検証すべく、様々なジャンルの映像に対して評価を行ったものである。

2 検証方法

予備調査として、ニコニコ動画のユーザーである 20 代学生 5 名に対して聞き取り調査を行い、コメン

トをするのはどういときかを尋ねた。「弹幕作成」

（映像が見られなくなるほどのコメントを大量に投稿すること。アスキーアートのように二次元的な絵に見せる試みも多い）、「あいさつ」（おつかれさまを意味する「乙」など動画の最初や最後で他の閲覧者に対して行うコメント）をのぞけば、「疑問があったとき」「相づち」「驚いたとき」「感動したとき」「ツッコミをいれたいとき」「一緒に歌いたいとき」など、閲覧者の感情や思想が動かされたときであることを示唆する意見が多かった。本稿の第二筆者らは、三次元的に表示される地形に残される多人数のコメントを利用した小説創造支援システムを研究してきたが[4]、そのシステムにおいても特徴的な地形にコメントが集中する傾向があった。そこで、ニコニコ動画においても大量にコメントが書き込まれる箇所になんらかの強い意味があるのではないかと考えた。

ニコニコ動画では、ひとつの動画あたり一定数のコメントが表示されるようになっており（再生時間 1 分以内の動画では 100 件、1 分以上 5 分未満では 250 件、5 分以上 10 分未満では 500 件、10 分以上では 1000 件）、これを越える場合には古いコメントから削除される（有料会員はこれまでに記録された全てのコメントを閲覧することができるが、この応用については 4.6 で述べる）。

コメントは、その内容とともにその動画におけるコメント入力時刻が記録されているので、これを取集計すればどのタイミングでどれくらい数のコメントがなされたかを知ることができる。筆者らは様々なジャンルの動画に対してこのコメント数を抽出し、コメント数が顕著に多い箇所が実際の動画でどのような意味をもつのかを調査した。

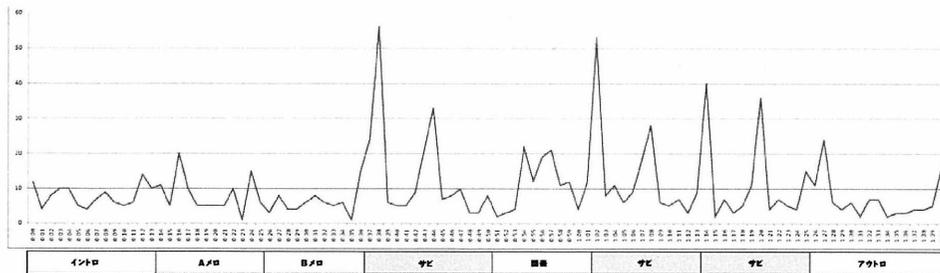


図1 「3D みくみく PV ♪」におけるコメント頻度

3 音楽系コンテンツにおけるサビ検出の試み

前章の予備調査で、コメントする動機に「一緒に歌いたいとき」という感想があった。三省堂の「デイリー 新語辞典」によれば、サビとは「ジャズやポピュラー音楽で曲想の変化した部分。また曲の途中の最も印象的な部分・メロディー」であるが、歌い出したいと思うタイミングはサビの冒頭部分である可能性が高いため、結果としてコメント数が多いところはサビの冒頭部と一致するのではないかという仮説を筆者らはもった。そこで、PV 的な音楽系コンテンツにおいてコメント数調査を行い、実際の楽曲構造と照らし合わせた。

3.1 3D みくみく PV ♪

最初に使用した音楽系コンテンツは、「3D みくみく PV ♪[5]」(長さ:1分40秒 投稿日:2007年10月25日)である。歌声ソフトウェアシンセサイザー「初音ミク」[6]を用いた楽曲「みくみくにしてあげる♪」の音楽に3DCGの映像を合わせたものであり、論文執筆時点で再生数150万回以上、コメント数10万件以上を誇る人気コンテンツである。横軸を再生時間、縦軸をコメント数としたグラフ、および楽曲の構造を示したのが図1である(コメントデータ取得日:2007年12月30日)。

この楽曲の構造は、イントロ(0~13秒)、Aメロ(14秒~25秒)、Bメロ(26秒~36秒)、サビ(37秒~50秒)、間奏(51秒~1分1秒)、サビ(1分2秒~1分13秒)、サビ(1分14秒~1分25秒)、エンディング(1分26秒~1分40秒)となっており、サビは3回繰り返されている。

グラフのピークとなっている部分を3点あげると、順に38秒、1分2秒、1分14秒となっており、コメント内容を見るとサビの歌詞にあたる「みくみくにしてあげる」というコメントを打ち込んでいるケースがほとんどであった。最初のサビだけピークの発現が2秒ほど遅れているが、この1回目のサビについてだけ、歌詞が「(君のこと) みくみくにしてあげる」と2拍前から開始されるアウトタクトとな

っており、サビの頭はこのアウトタクト部からなるものの、ユーザーのコメントはこのアウトタクト部分が終わったところからスタートしていることが原因だということがわかった。キャッチーな歌詞がサビの頭から始まっていない場合やアウトタクトが付与されている場合に、サビ検出がうまくいかない例ということができる。

3.2 私の時間 3DPV ♪

次に調査した音楽系コンテンツは「私の時間 3DPV ♪[7]」(長さ:1分52秒 投稿日:2008年1月2日)である。このコンテンツも初音ミクを用いた楽曲であり、執筆時点で再生数44万回、コメント数2万回の人気コンテンツである。コメント頻度と楽曲構造を図示したものが図2である(コメントデータ取得日:2008年1月9日)。

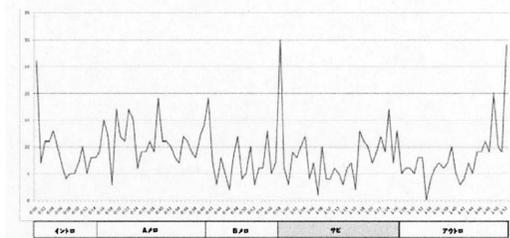


図2 「私の時間 3DPV ♪」におけるコメント頻度

この楽曲の構造でとても特徴的なのは、「サビが一度しか歌われない」(58秒時点)ということである。後藤の SmartMusicKIOSK におけるサビ検出アルゴリズム[8]では、楽曲の繰り返し構造を解析し、それに基づいて8割近い正答率でサビ検出を行うことに成功しているが、このように1回のみしかサビが用いられない楽曲についてはその検出を苦手とする傾向がある。図2のグラフをみてみると、最もコメント頻度が高いのは58秒の時点であり、ちょうど本楽曲のサビの開始時点に一致している。ちなみにコメン

ト頻度が 2 番目, 3 番目に高いところはそれぞれ動画開始時と終了時に対応していた。

3.3 ロックマン2 おっくせんまん! (Version コム)

この動画コンテンツ「ロックマン2 おっくせんまん! (Version コム) [9]」(長さ: 2分 51 秒 投稿日: 2007 年 3 月 6 日)はニコニコ動画発足初期から人気のある曲で, 再生回数 270 万以上, コメント数 82 万件を越えている。ゲームタイトル「ロックマン2」の BGM にオリジナルの歌詞をのせた楽曲である。コメント頻度と楽曲構造は図 3 で示される(コメントデータ取得日: 2007 年 12 月 30 日)。

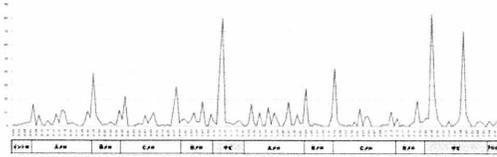


図 3 「ロックマン2 おっくせんまん! (Version コム)」におけるコメント頻度

楽曲中でサビは 3 回繰り返されており, それぞれ 1 分 11 秒, 2 分 26 秒, 2 分 36 秒の箇所である。コメント頻度をみると顕著なピークが 3 箇所あり, 1 分 14 秒, 2 分 28 秒, 2 分 39 秒となっている。本楽曲においてサビの部分は「君がくれた勇気はおっくせんまん! おっくせんまん!」「見過ごしてた景色はおっくせんまん! おっくせんまん!」という歌詞になっており, 「おっくせんまん!」の箇所ですべてを連呼する「弹幕」が入っている。このため, これらのピークはサビに対する反応ではあるものの, 「君がくれた勇気は」「見過ごしてた景色は」というサビの冒頭部分を示しているわけではなかった。

3.4 コンビニ【オリジナル曲】に勝手に絵を(ry FULL.ver

このコンテンツ「コンビニ【オリジナル曲】に勝手に絵を(ry FULL.ver[10]」(長さ: 4分 26 秒 投稿日: 2008 年 2 月 12 日)も, 初音ミクによる楽曲であるが, 映像はこれまでとは異なりストーリー仕立てのアニメーションとなっている。コメント頻度と楽曲構造は図 4 で示される(コメントデータ取得日: 2008 年 2 月 28 日)。

この楽曲では, 1 分 5 秒, 2 分 23 秒, 3 分 14 秒, 3 分 43 秒の時点でサビが開始されるが, コメント頻度が高いのは 1 分 53 秒, 3 分 22 秒, そして 4 分 25 秒と, 音楽の構造とは無関係な部分である。1 分 53 秒の位置については, コンビニの店員が愛らしくほほえむシーンであり, ここではハート型の「弹幕」が投稿されている。3 分 22 秒の位置では, 主人公の描

画が特定の漫画のそれと似通っているため, それに呼応したコメントが投稿されている。また, 最後の 4 分 24 秒時点では, このコンテンツの感想が大量に書き込まれていた。これらのことから, このコンテンツに関してはコメント頻度が音楽の構造よりも映像のストーリーやシーンに強くひきずられた例であるということが出来る。

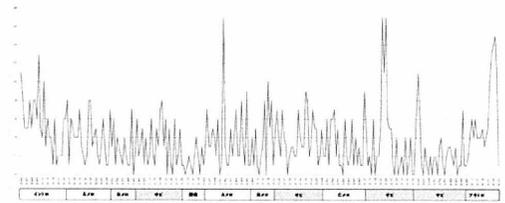


図 4 「コンビニ【オリジナル曲】に勝手に絵を(ry FULL.ver)におけるコメント頻度

3.5 考察

他にも同様な手順で様々な PV 系の動画について調査を行ったところ, 3.4 のように楽曲構造と完全に異なるピークを持つ例は少なく, 大半はサビの冒頭部かサビ中のキャッチーな部分の開始点に主要なピークが存在することがみてとれた。

「サビ中のキャッチーな部分の開始点」はいわば「サビの中のサビ」のようなものであるから, 辞書的意味に照らしたときに「曲想の変化した部分」でこそないものの, 「曲の途中の最も印象的な部分・メロデー」であることは間違いない。こうしたことから, コメント頻度とサビはかなり強い関係をもっている結論できるのではないだろうか。後藤の SmartMusicKIOSK[8]のようなインタフェースでコメント頻度の高いポイントに移動するような頭出しボタンを設けるなどすれば, 有用性の高いアプリケーションにつながることもできるのではないだろうか。

4 映像要約への応用

映像要約やサムネイル表示に関する研究は盛んに行われている。三浦らは料理映像の特徴に着目し, 動き検出による料理映像解析を行い, 映像要約を行った[11]。伊藤らは音声と映像の一貫性を考慮し, ニュース報道番組への要約システムを作成した[12]。出口らは, 映画の文法に基づいた映像要約手法を提案している[13]。また, 林らも概念グラフを用いたニュース映像要約システムを作成している[14]。本稿では, コメント数が多い箇所を映像中で重要な箇所であると仮定し, ニコニコ動画におけるコンテンツを 30 秒の映像に要約する手法を考えた。

まずは, 映像開始と終了の 3 秒ずつを取得し, これを要約映像の開始・終了と一致させる。これは,

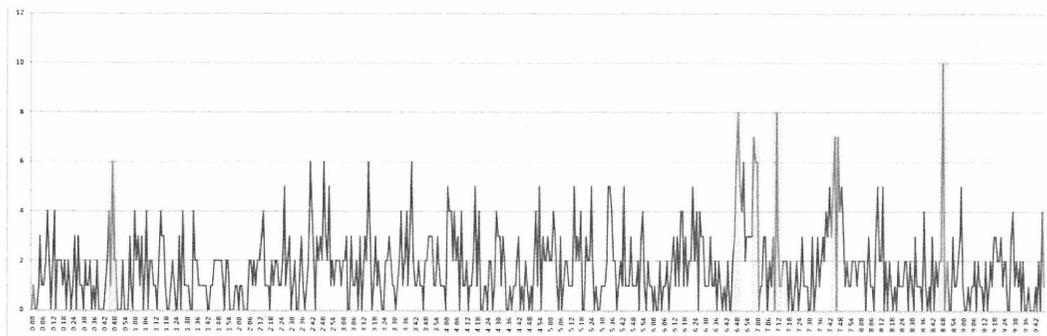


図5 「ハイパーション作ってみた」におけるコメント頻度と要約採用箇所

映像開始・終了時にタイトルやクレジットが記述されることが多いからである。次に、コメント頻度が高いポイントを抜き出し、その前2秒、後3秒をひとつのユニットとみなして取り出す。ニコニコ動画では、意図したタイミングでちょうど文字が流れるように調整して投稿するケース（3章での音楽系コンテンツではそうした投稿が多い）と、動画に何らかの感想をもったときにその場で投稿するケースがある。後者の場合は、反応してから投稿するまでに2秒程度遅れることがある。そこでコメントのピークポイントの前2秒から取得するルールとしたわけである。後ろを3秒としたのは、コメントが表示される時間が約3秒であることによる。なお、この領域に他のユニットが重なる場合はひとつの大きなユニットとして結合させるようにする。

これらのユニットの準備ができれば開始・終了の3秒ずつのユニットを30秒の区間の最初と最後に配置し、コメント頻度の高いユニットから順にこの間に入れていく。ユニットどうしの順番はその時刻に準じ、各ユニットは1秒ずつクロスフェードさせて連結する。これでぴったりと30秒収まる場合はそれで完成とする。最後に1ユニットの幅未満の隙間が残った場合は、ここにもうひとつユニットを削って詰め込む。ユニットの削り方は、①コメント直後の1秒間、②次の1秒間、③その次の1秒間、④コメント直前の1秒間、⑤その前の1秒間の順に優先して保持しながら尺をつめることとする。以上のルールに従えば、一意に要約映像ができあがる。

もし、映像にも「起・承・転・結」のような流れがあるならば、映像要約のためには「起」と「結」がまず大事であるが、これは映像の最初と最後の部分であるから取り出すのは比較的容易である。「承」は内容的に「起」の延長上に存在するものであるから、省略してかまわない。「転」は、全体の流れを面白くするという意味ではぜひ要約映像の中に入れていたが、通常はそれがどこにあたるかわからない。し

かし、ニコニコ動画においてはこの「転」の部分にこそたくさんコメントが集まる可能性がある。筆者らは、このような仮定のもと本ルールを設定した。次からは、本ルールに基づいて様々な動画ジャンルにおいて映像要約を試み、その結果を検証した。

4.1 ハイパーション作ってみた.

コンテンツ「ハイパーション作ってみた[15]」（長さ：9分45秒 投稿日：2007年12月29日）は、馬の仮面をかぶった投稿者がドリンク「ポーション」に次々と栄養ドリンクを混ぜ、煮込んだ末に飲もうとするが嘔吐してしまい（ポートの映像が流れる）、再び飲むも嘔吐、カレーと混ぜて食しようとするがさらに嘔吐、最後に乾燥剤が入っていたことに気づくという内容である。いさか品がない内容ではあるが再生数94万回、コメント数64万件となる人気コンテンツであり、ニコニコ動画に投稿されるいわゆる「ネタ映像」の典型でもあるといえる。この映像のコメント頻度のグラフを図5に示す。このうち網掛け部分が、映像要約に使用された部分である。

要約映像は次のような内容となっている。投稿者挨拶→ドリンクを鍋に入れる映像→ドリンクを飲み、むせる→ポートの映像→再び飲もうとする→ポートの映像→カレー皿→再び嘔吐→乾燥剤が写って終了。このように、もとのコンテンツの1/20の時間であるにもかかわらず、エッセンスをほぼすべて取り込んで要約に成功しているといえる。この要約映像だけを見ても、十分内容把握が可能である。コメントをみてみると、笑いを誘う箇所（嘔吐するところなど）に対して一斉にコメントがなされているため、「面白いところ」をかいつまむことができたのではないかと考えている。

4.2 ウサビッチ 第22話「戦車注意」

ストーリー性を持つ動画の例としてとりあげたこのコンテンツは、人気シリーズとして投稿されてい

る 3DCG アニメーション[16]である(長さ:1分29秒 投稿日:2007年12月12日)。二匹のウサギが運転する自動車が戦車と出会い、攻撃を受けながら追いかけるが、運転しながらのタイヤ交換、砲撃されて車が空中分解してもそれを瞬時に組み立て直して逃げ切るといったストーリーである。映像のコメント頻度のグラフが図6である。先ほどと同様、網掛け部分が映像要約に使用されたシーンである。

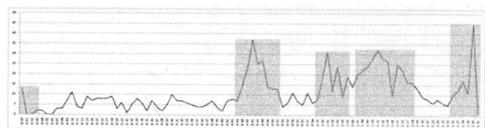


図6 ウサビッチ 第22話「戦車注意」におけるコメント頻度と要約採用箇所

要約映像を見てみると、タイヤ交換、空中分解と修理という大事なシーンがちゃんと収録されており、要約としては成功しているといえる。ただし、もとのコンテンツが短いので、30秒の要約映像だと1/3であるから、これは当然の結果なのかもしれない。

4.3 自作の改造マリオ(スーパーマリオワールド)を友人にプレイさせる

ニコニコ動画にはゲームを操作しているところをキャプチャした動画が多くアップロードされている。本コンテンツ[17]もそのひとつで、難易度が異常に高くなるように改造を施したゲームをプレイしている映像であり、130万回の再生数、200万回のコメント件数を誇る(長さ:9分22秒 投稿日:2007年6月9日)。難易度が高いだけあって、マリオが死亡してはリトライすることの繰り返しとなっている。映像のコメント頻度のグラフおよび映像要約に使用された部分を表したのが図7である。

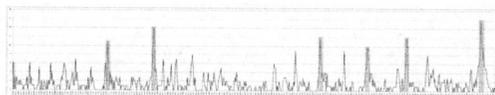


図7 「自作の改造マリオ(スーパーマリオワールド)を友人にプレイさせる」におけるコメント頻度と要約採用箇所

実際に映像を見てみると、マリオが死亡したり、うまく通過したりするシーンが断片的にめまぐるしく流れる映像となっており、前後の関係がわからず難解なものとなってしまっている。この映像をわかりやすくするには、詰め込むユニットの数を減らし、そのぶん前後の映像を長くすることが必要であると考えられる。

4.4 一から作るドラ焼きとお団子

このコンテンツは、ドラ焼きとみたらし団子の作り方を解説するもの[18]であり、三浦らの研究[11]で対象としている料理映像の典型としてとりあげた(長さ:8分11秒 投稿日:2007年11月19日)。図8が映像のコメント頻度のグラフおよび映像要約に使用された部分である。

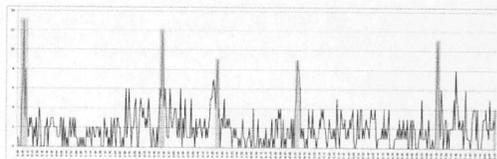


図8 「一から作るドラ焼きとお団子」におけるコメント頻度と要約採用箇所

実際の映像をみてみると、ドラ焼きの材料一覧→卵を割る→生地を焼く→みたらし団子の材料一覧→串に刺さった団子→たれを作っている映像→ドラ焼きを二つに切って確認している映像、という構成になっている。材料一覧は取得されているものの、生地をどうやって作るのか、団子はどうやって作るのかなど、料理を行うにあたって重要な部分がいくつか欠けており、この要約映像だけをみて料理を再現することはできない。

4.5 らきすたの報道ニュース

最後に、伊藤らの研究[12]で対象としているニュースコンテンツの一例として、「らきすたの報道ニュース」と題してアップロードされているニュース映像[19]をとりあげた。これは、アニメの舞台となった鷲宮神社に多くのファンが訪れており、コスプレで参拝したり、アニメキャラクターを描いた絵馬を残したりしている様子が紹介されているニュースである。図9は映像のコメント頻度のグラフおよび映像要約に使用された部分である。

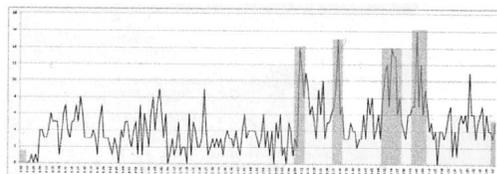


図9 「らきすたの報道ニュース」におけるコメント頻度と要約採用箇所

要約映像をみてみると、「神社が...オタクの聖地」というタイトル→記念撮影している様子→コスプレの様子→アニメキャラが描かれた絵馬→神主インタ

ビューの冒頭部→次のニュースを読み上げるアナウンサーという流れになっている。おおまかな流れは理解できるが、その神社がどこの神社であるのかという情報が抜けるなど、要約映像としては不十分な印象を与えている。

4.6 考察

以上の検証から、ニコニコ動画を利用した提案手法による映像要約では、うまくいくケースとそうでないケースがあることがわかった。ニコニコ動画でコメント数が増えるのは、広義での「面白いシーン」であることが多く、要約映像に取り入れる価値の高いことは間違いないが、そのおもしろさを理解するために必要なコンテキストが他にある場合は十分に面白さが伝わらないことがある。

また、料理番組の要約にはまったく不向きであるということが予想される。本稿のコンテンツでは料理の合間のトークが入っていなかったが、もし多くの料理番組のようにトークが入っていた場合は、むしろその楽しいトークが選択されてしまって料理プロセスが捨象されることも予想され、さらに要約としては役に立たないものになる可能性もある。

ユーザが面白いと思うところを取り出す要約は、良くも悪くもニコニコ動画的であるといえる。2008年4月に発表された第4回国際ニコニコ映画祭での大賞は「15秒でわかる日本のむかしばなし[20]」であった。これは昔話をおもしろおかしく要約したものであり、たとえば竹取物語は「むかしむかしあるところに竹取の翁がいましたが、ある日のこと光る竹の中からそれはそれは美しい女の子を発見しかぐや姫と名付け大事に育て周囲の男たちにモテモテでしたが、月に帰りました」というように不均等な要約を行うことで笑いを誘った映像作品である。本稿の提案手法による映像要約は一部失敗しているが、要約の正しさではなく面白さを評価するユーザになれば、受け入れられる可能性をまだ秘めているといえるのかもしれない。

また、ニコニコ動画の有料会員はこれまでに記録された全てのコメントを閲覧することができるが、これを利用すると様々な時期におけるコメント頻度を取得し、その変遷をみることができる。ニコニコ動画のコンテンツでは、アップロード当初は面白いと思われなかったシーンが、空耳として違う台詞に聞こえるなどといった新たな価値を発見され、注目を浴びることが往々にしてある。このように時代とともに動的に変化していく解釈内容についても追っていくのが、本稿における提案のひとつの強みなのかもしれない。

参考文献

- [1] <http://www.nicovideo.jp/>
- [2] 伊藤聖修, 鈴木育男, 山本雅人, 古川正志. ニコニコ動画におけるタグ共起ネットワークの特徴抽出, 第80回人工知能学会知識ベースシステム研究会 (SIG-KBS), 1月15-16日, NTT武蔵野研究開発センター, 2008.
- [3] 川井康寛, 志築文太郎, 高橋伸, 田中二郎. 動画共有に基づいた非同期コミュニケーションの連帯感を向上させるインタフェース, 第15回インタラクティブシステムとソフトウェアに関するワークショップ (WISS2007), pp.135-136, 2007.
- [4] 海沼賢, 宮下芳明, 西本一志: 他者からの触発を活用する小説創造プロセスの分析, 情処研報 2006-EC-3, Vol.2006, No.24, pp.113-120, 2006.
- [5] 3D みくみく PV ♪
<http://www.nicovideo.jp/watch/sm1359820>
- [6] 初音ミク
<http://www.crypton.co.jp/mp/pages/prod/vocaloid/cv01.jsp>
- [7] 私の時間 3DPV ♪
<http://www.nicovideo.jp/watch/sm1929913>
- [8] 後藤真孝."SmartMusicKIOSK: サビ出し機能付き音楽聴機", 情報処理学会論文誌, Vol.44, No.11, pp.2737-2747,2003.
- [9] ロックマン2 おっくせんまん!
<http://www.nicovideo.jp/watch/sm83>
- [10] コンビニ【オリジナル曲】に勝手に絵を(ry FULL.ver
<http://www.nicovideo.jp/watch/sm2302757>
- [11] 三浦宏一, 浜田玲子, 井手一郎, 坂井修一, 田中英彦: "料理映像の特徴を利用した要約手法の検討," 信学技報, PRMU2002-22, pp.15-20, Jun. 2002.
- [12] 伊藤一成, 酒井康旭, 斎藤博昭: 音声と映像の一貫性を考慮した要約動画の生成, 電子情報通信学会・データベース学会合同ワークショップ DEWS, (2004)
- [13] 出口嘉紀, 吉高淳夫. 映画の文法に基づく要約映像の生成. 情報処理学会研究報告. データベース・システム研究会報告, Vol.2004, No.3(20040115) pp. 33-40, 2004.
- [14] 林英俊, 李龍, 上林弥彦, "概念グラフを用いたニュース映像要約システムの構築", 第14回データベース工学ワークショップ DEWS2003, Mar 2003. 4.
- [15] ハイボーション作ってみた.
<http://www.nicovideo.jp/watch/sm1890440>
- [16] ウサビッチ 第22話「戦車注意」
<http://www.nicovideo.jp/watch/sm1756027>
- [17] 自作の改造マリオ(スーパーマリオワールド)を友人にプレイさせる
<http://www.nicovideo.jp/watch/sm423963>
- [18] 一から作るドラ焼きとお団子
<http://www.nicovideo.jp/watch/sm1569143>
- [19] らきすたの報道ニュース
<http://www.nicovideo.jp/watch/sm731876>
- [20] 15秒でわかる日本むかしばなし
<http://www.nicovideo.jp/watch/sm2849637>