

Nicolizer: ユーザインタラクションに応じた 楽曲動画のコメント自動変換システム

山本 岳洋^{1,2,a)} 中村 聡史^{1,2,b)}

概要: 本稿では、動画共有サービス上に投稿されている楽曲動画に付与されたコメント内容をユーザがインタラクティブに変更可能なシステム「Nicolizer」を提案する。本研究の目的は、ユーザが動画に対してインタラクションすることで、「もっと大量にコメントを流したい」、「より泣けるコメントにしたい」、「コメントを感情的にしたい」、「真面目な内容にしたい」といったさまざまな要求に応じて動画のコメントを変更可能な仕組みを実現し、ユーザが能動的に動画を多様な観点から楽しむための方法を提供することである。本稿では、楽曲動画を対象として、視聴中の楽曲動画に合った、「笑えるコメント」や「泣けるコメント」をユーザの操作に応じて自動的に生成し、視聴中の動画に提示する仕組みを提案する。Nicolizerを用いることで、これまで受動的な閲覧しかしなかったコメントに対して視聴者は能動的なインタラクションが可能となり、同じ楽曲動画を聴覚的にも視覚的にも繰り返し楽しむことができると考えられる。

Nicolizer: An Automatic Comments Generation for Music Video Clips based on User Interactions

TAKEHIRO YAMAMOTO^{1,2,a)} SATOSHI NAKAMURA^{1,2,b)}

Abstract: This paper proposes a system called “Nicolizer”, which enables users to interactively modify comments posted on music video clips uploaded to a video sharing service. The purpose of this work is to enable users to actively interact with the comments on the music video clips by enhancing the comments from the various aspects. Nicolizer automatically generates comments related to audio moods such as “sadness” or “happy” by learning comment models from the existing music video clips. Nicolizer enables users to actively interact with comments posted on the music video clips. It also enables users to enjoy a music video clip from both visual and audio aspects.

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。

ニコニコ動画では、視聴者は動画の任意の再生時間に対して自由にコメントを投稿することができ、その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、投稿された再生時間と同じタイミングで動画中に流れて表示されるという、他の動画共有サービスにはない特徴を持つ。ニコニコ動画に投稿されている VOCALOID に関する楽曲の

¹ 京都大学 情報学研究所
Graduate School of Informatics, Kyoto University

² JST CREST
JST CREST

a) tyamamot@dl.kuis.kyoto-u.ac.jp

b) nakamura@dl.kuis.kyoto-u.ac.jp

*1 <http://nicovideo.jp>

動画にも、多くのコメントが投稿されており、視聴者は楽曲だけではなく、コメントを閲覧することで視覚的にも楽曲動画を楽しむことができる。

我々は、気に入った楽曲動画は、ブックマークをしたりプレイリストを作成したりすることで、何度も繰り返し聴くことでその楽曲を楽しむことをよく行う。しかし、楽曲動画の「楽曲」を繰り返し聴く一方で、そうした動画に付与された「コメント」をその度に繰り返し見ることはあまりないのではないかと考える。楽曲はそのメディアの性質上繰り返しの視聴を受け入れることができる。一方で、楽曲動画に付与されたコメントは他のユーザとの共感を感じる上で非常に有用なメディアであるものの、楽曲を繰り返し楽しむという観点からコメントを考えた場合、視聴者には以下のような欲求があるのではと考えられる。

- 同じ楽曲動画を1日や1週間に何回も再生したりするような場合、楽曲動画に付与されるコメントはあまり変化しない。また、投稿されてしばらく経過した動画に追加コメントが投稿されることも少ない。そのため、毎回同じコメントが提示されることになり、新鮮さが失われていくことがある。そこで、動画の再生の度に提示されるコメントをいろいろと変化させたい。
- 楽曲動画に対して、楽曲の印象とは関連しない説明的なコメントや歌詞、会話のためのチャットコメントが投稿されていることがある。他者と感情を共有しながら楽曲動画を視聴する場合には、楽曲の印象に関連したコメントのみを動画に表示したい。逆に、冷静に楽曲動画を視聴する場合は、説明や歌詞などのコメントを積極的に表示し、印象に関係あるコメントは消したい。
- 投稿されたばかりの楽曲動画や、あまり視聴者数を集めることができなかった楽曲動画には、動画自体に付与されるコメントが少ない。コメントが沢山投稿されていると、楽曲動画の楽しさが向上することがある。そこで、もっと多くのコメントを流しながら楽曲動画を視聴したい。

我々は、視聴者のこのような欲求を満たすためには、視聴者の興味に応じてコメントをインタラクティブに変更、生成可能な仕組みが重要なのでは無いかと考えている。

音楽情報処理に関する研究分野では、楽曲に合った映像の自動生成がいくつか行われている。たとえば後藤らは楽曲のビートに同期して踊るCGキャラクタを生成するための手法を提案している [3]。また、藤澤らは楽曲の音響特徴量に応じた色彩や形を描画する手法を提案している [4]。最近では、室伏らは既存のダンス動画を利用することで、楽曲に合ったダンス動画を半自動で生成する仕組みを提案している [5]。また、多くの音楽再生プレーヤーは楽曲の周波数成分や強弱に反応した視覚効果を描画する機能がある

(図1(b))。こうした研究は、楽曲を視覚的にも楽しむための映像の自動生成手法と考えることができる。

本研究の目的は、このような研究と同様に、「楽曲に合ったコメント」を自動的に生成することで、視聴者が楽曲動画を視覚的に何度も楽しむことができるような仕組みを実現することである。そのために、本研究では、既存の楽曲動画から楽曲に合ったコメントを学習し、ユーザのインタラクションによって視聴中の楽曲に付与されたコメントを自由に変更可能なシステム「Nicolizer^{*2}」を提案する。Nicolizerは楽曲を含む音響の調整に広く用いられているグラフィックイコライザをメタファとしたシステムであり、動画の視聴者はイコライザの各スライダを調整することで、動画に付与されたコメントを自由に変更することができる。

Nicolizerはユーザの操作に応じて、既存の楽曲動画から学習したコメントを自動で生成する本稿では、「泣ける」、「笑える」といった楽曲から受ける主観的な印象に着目し、そうした印象と関連するコメントを自動的に生成し視聴中の動画に提示することができる。この仕組みにより、動画の視聴者は「泣ける」楽曲を聴いている際に、より「泣ける」コメントを動画に多く提示することで、視覚的にも楽曲を楽しむことができる。また、こうしたコメント生成は確率的に行われるため、動画を再生する度に異なったコメントが提示される。このようなコメントを視覚的に楽しむことで、楽曲動画を視覚的にも聴覚的にも何度も繰り返し楽しむことができるのではないかと考えている。

2. Nicolizer の設計

本章では、提案システム「Nicolizer」の設計方針について述べる。まず、インタフェースとしての設計方針について述べた後、楽曲動画に合ったコメントの自動生成に必要な要件について整理する。

2.1 イコライザインタフェースを介したコメントへのインタラクション

Nicolizerは、楽曲動画に付与されるコメントを変更するための仕組みとして、音響の出力調整に良く用いられるイコライザをメタファとしたインタフェースを用いる。図1(a)は、AppleのiTunes^{*3}が提供するグラフィックイコライザの例である。通常のグラフィックイコライザでは、各周波数成分を強弱を各スライダで調整することで、出力となる音響に影響を与える。本システムも対象となるコンテンツは楽曲動画であり、イコライザのようなインタフェースはユーザにも受け入れやすいのではないかと考えられる。

^{*2} NicoNico + Equalizer

^{*3} <http://www.apple.com/jp/itunes/>

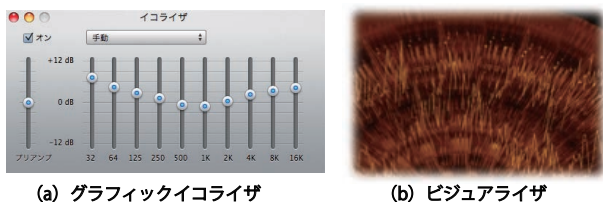


図 1 グラフィックイコライザと楽曲のビジュアライザの例.

図 2 に Nicolizer のインターフェースを示す. グラフィックイコライザが音響の出力を変化させるのに対して, Nicolizer では各スライダーの変更は動画に提示されるコメントを変化させる.

本稿では, 図 2 に示すように 3 種類のスライダーによりコメントを変化することができる.

(1) 元動画のコメントと自動生成コメントの混合量

ユーザは視聴中の動画に付与されているコメントのみを閲覧したり, システムが自動生成したコメントのみを閲覧したりすることができる.

(2) 動画上に提示するコメント量 本スライダーの値を調整することで, 楽曲動画上に表示されるコメントの分量を調整できる.

(3) コメントが関連する印象の調整 それぞれの印象に関するスライダーの値を調整することで, どの印象からコメントが生成されやすいかを調整できる. これにより, 視聴者は泣ける楽曲を聴いている際に, 「泣ける」コメントを元動画よりも作為的に大量に流したり, 「笑い」に関するコメントを流すことで普段の視聴とは異なる視聴の仕方が可能となる.

2.2 楽曲動画に合ったコメントの生成

提案システムは, 楽曲動画をコメントという観点から視覚的に楽しむために, 既存の楽曲動画からコメントを学習することで, 楽曲に合ったコメントの自動生成を行う. 本稿では, 特に下記 2 つの対応関係に注目して, コメントの自動生成を行う.

1. 楽曲の印象

楽曲を聴いている最中, 我々は「泣ける」, 「笑える」, 「かわいい」, 「かっこいい」といったさまざまな印象を受ける. 楽曲動画に付与されるコメントには, そうした印象を反映したコメントが多く含まれる. たとえば, 「泣ける」印象を与える楽曲の動画には, 「切ない…」, 「泣けてきた」といったコメントが付与されることが多い. こうした, 楽曲を聴いた印象に合ったコメントを自動的に生成することができれば, 楽曲の雰囲気合ったコメントを視聴者に提示することができる.

2. 楽曲の時間的構造

一般的に, 楽曲は「イントロ」, 「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



図 2 Nicolizer のインターフェース画面. ユーザはスライダーを操作することで, 動画に提示されるコメントを調整することができる.



図 3 楽曲の時間的構造と, 動画に付与される典型的なコメントの例.

解することができ, その繰り返しの構成される. 楽曲動画に付与されるコメントは, こうした楽曲の時間的な構造を密接に関連していると考えられる. 例えば, 楽曲が始まる動画の前では「期待」, 「wtkk」といった, これから始まる楽曲に期待するコメント, サビ区間では「サビすげえ」といったコメントが, さらに動画の終了時には「GJ」, 「8888」といった動画を讃えるコメントが付与されると考えられる (図 3). このように楽曲の構造に合ったコメントを生成することで, 違和感の無い, 楽曲動画に合ったコメントが提示できると考えられる.

3. 実装

本章では, Nicolizer の実現方法について説明する. Nicolizer は大きく分けて, (1) コメント学習フェーズ, (2) 動画視聴フェーズの 2 つのフェーズから構成される. それぞれの概要を図 4 と図 5 に示す. コメント学習フェーズでは, 楽曲動画に合ったコメントを生成するために, 既存楽曲動画集合からコメントを収集し, 楽曲構造と楽曲の印

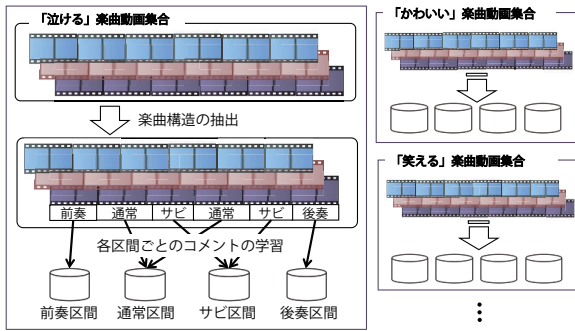


図 4 コメント学習フェーズの概要. 印象と楽曲の区間に応じてコメントを学習する.

象に合ったコメント生成モデルを学習する. そして, 動画視聴フェーズでは, 学習したコメント生成モデルを基に, ユーザのインタラクションに応じてコメントを自動的に生成し視聴中の動画に提示する.

以降, 各フェーズの詳細について説明する.

3.1 コメント学習フェーズ

3.1.1 印象ごとの楽曲動画の収集

図 4 に示すように, 楽曲の印象に関するコメントを自動的に生成するために, まず, 「笑い」や「泣ける」といった印象ごとに関連する楽曲動画を用意する必要がある. 本研究では, 我々がこれまでに取り組んできた, 楽曲動画の印象推定に関する研究 [6] で用いたデータセットを使用した. これは, ニコニコ動画上で「VOCALOID」タグの付与された動画 186,987 本の中から, 「かわいいミクうた」や「泣けるミクうた」といった, 楽曲の印象に関連するタグが付与されている楽曲動画のみを抽出し, 類似する印象ごとに楽曲動画を 7 つの印象クラスにまとめたデータセットである. 本研究では, 「笑い」, 「泣き」, 「かつこいい」, 「かわいい」という 4 つの印象に関する動画集合 (計 1,806 本) を用意し, コメントの学習に用いた.

3.1.2 楽曲構造の抽出

各印象ごとに動画集合を用意した後, 各動画の楽曲構造ごとにコメントを抽出し, コメントの生成モデルを構築する. 楽曲の構造毎にコメントを学習することで, 2.2 節で述べたように, 楽曲の構造に即したコメントが生成可能になると考えられる.

本稿では, 楽曲の構造検出手法として Goto により提案された手法「RefraiD」[7]を用いた. RefraiD は, 楽曲の断片的な繰り返し区間の相互関係を調べながら楽曲の構造を求める手法であり, 80%の楽曲についてサビ区間を正しく検出可能な, 非常に高精度な楽曲構造手法である. 本稿では RefraiD で得られるサビ区間をもとに, 動画を下記の 4 種類の区間に分割した.

サビ区間 RefraiD でサビ区間として検出された区間

前奏区間 楽曲動画の開始 15 秒間

1. 楽曲動画にアクセス



2. イコライザの値に基づきコメントを生成・提示

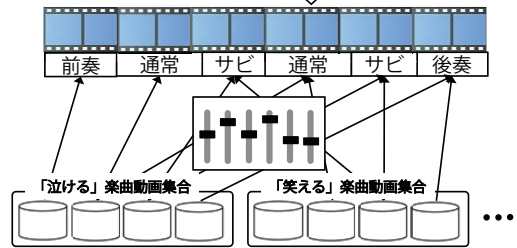


図 5 楽曲動画視聴フェーズの概要. コメント学習フェーズで学習したコメントモデルと, イコライザの値に基づいて, 各楽曲区間ごとにコメントを自動生成し, 視聴中の動画に提示する.

後奏区間 楽曲動画の終了 15 秒間

通常区間 上記 3 つのどれにも当てはまらない区間

そして, それぞれの区間について, 楽曲動画の印象毎にコメントの生成モデルを構築する. 本稿では, 簡単のため, コメント生成モデルとして多項分布を仮定し, 最尤推定により多項分布を推定することでモデルを学習した. すなわち, それぞれの印象の楽曲構造における, 各コメントの出現確率に応じてコメントが生成される.

3.2 楽曲動画視聴フェーズ

3.3 コメントの生成

視聴者が Nicolizer を通して楽曲動画にアクセスすると, まず, システムは RefraiD の結果を用いて楽曲動画を 3.1.2 項で述べた 4 つの区間に分割する. その後, イコライザの初期設定値を基に, 各区間ごとにコメントを生成する. 各区間のコメント生成の際には, 2.1 節で述べた 3 種類のスライダーの値を元にコメントを生成する. また, 生成されたコメントが付与される再生時刻については, 一様分布に従ってランダムに決定した. さらに, サビ区間に関しては, 楽曲動画の盛り上がる部分であると考えられるため, 他の区間よりも多くのコメントを生成することで, 楽曲の盛り上がりを表した.

4. 実行例

実際に Nicolizer を利用して, 楽曲動画のコメントを変更する例を示す. 図 6 と図 7 は同一の動画に対して, それぞれ異なるイコライザの値でコメントを生成した様子である. 図 6 と図 7 で再生している楽曲動画には, 元々「泣ける」や「切ない」といったコメントが付与されており, 視聴者にとって「泣ける」印象を与える楽曲であると考えられる. このような動画を視聴をしている際に, Nicolizer の「泣き」に関するスライダーの値を強調することで, より多くの「泣ける」印象をコメントを生成することができた.

このようにすることで、元々の動画に付与されたコメントから受ける「泣ける」という印象を、より強く感じることができるのではないかと考えられる。

一方、図7は「笑い」に関するスライダの値を強調した際の、自動生成されたコメントの例である。図から分かるように、「笑える」印象に関連したコメントが動画に提示されていることが分かる。このようなインタラクションを行うことで、普段の楽曲から受ける印象とは異なる印象をコメントから受ける事ができる。このように、イコライザを介してコメントにインタラクションを行うことで、これまでにない新しい楽曲の楽しみ方を体験できるのではと考えられる。また、楽曲の構造毎にコメントを学習しているため、動画のはじめの方では動画に期待するコメントが、動画の終了時にはクリエイタに感謝するようなコメントが生成されており、全体的に違和感なくコメントを生成することができていた。

しかし、今回のコメント生成手法は、コメント文字列をそのまま学習し生成するというシンプルな手法なため、違和感のあるコメントも生成されることも多い。たとえば、初音ミクに関する楽曲であるのにも関わらず他の VOCALOID に関するキャラクタ名を含むコメントが生成されたり、他の楽曲の歌詞が生成されたりということもあった。今後は、Yoshii らが提案しているような、より洗練されたコメント生成手法 [8] を用いることで、より楽曲に合った違和感のないコメントの生成を行う必要がある。

さらに、コメントの自動生成が動画の視聴者に与える影響についても考える必要がある。今回はコメントを自動的に生成し、元動画と組み合わせてユーザに提示していた。しかし、現在表示されているコメントが、その動画に対してコメントされたものではなく、自動的に生成されたことが分かると、視聴者にとって、そのコメントが機械的なものと感じられ、コメントに対して違和感を感じてしまうと考えられる。コメントの背後に存在する、そのコメントを投稿したユーザの存在感を残したまま、視聴者の要求に応じてコメントを変換するために、元のコメントを生かしたままコメント内容を変換するための手法について考えていく必要がある。

5. まとめと今後の課題

本稿では、既存の楽曲動画を再利用してコメントを学習することにより、楽曲動画に付与されるコメントをユーザ側で自由に変更可能なシステム「Nicolizer」を提案した。Nicolizer を使用することで、視聴者は楽曲から受ける印象をより強く反映したコメントを動画に提示したり、一方で楽曲から受ける印象と全く異なる印象に関するコメントを動画に提示したりすることができる。視聴者はこのようなコメントを視覚的に楽しむことで、楽曲動画を視覚的にも聴覚的にも何度も繰り返し楽しめることができるのでは

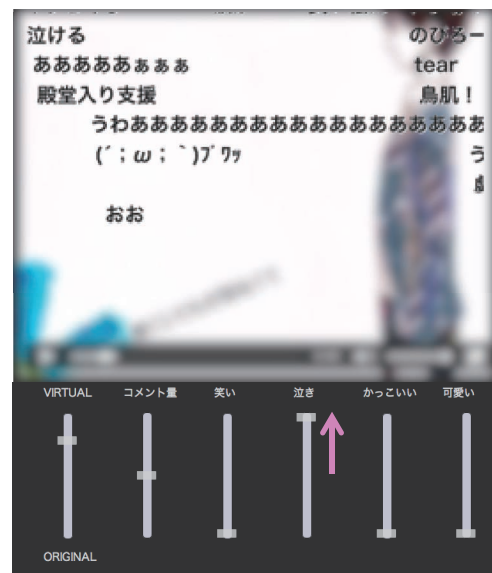


図6 「泣き」印象の度合いを強調した際のコメントの自動生成例。



図7 「笑い」印象の度合いを強調した際のコメントの自動生成例。

ないかと考えている。また、Nicolizer はコメントを自動生成することが可能なため、コメントがあまり付与されていない楽曲動画などでもより楽曲を視覚的に楽しむことができる。

今回は、楽曲とコメントの対応関係として、楽曲から受ける印象に着目した。しかし、楽曲とコメントの対応関係には他にもさまざまな関係が考えられる。たとえば、楽曲のテンポとコメントを提示する速度を合わせたり、楽曲の音色とコメントの色や大きさを合わせるといった提示手法が考えられる。さらに、今後の方向性として、コメントだけではなく、映像についても既存の楽曲動画を再利用することで楽曲に合った映像の自動生成が可能ではないかと考えている。たとえば、「泣ける」楽曲動画の映像を集約することで、「泣ける」楽曲に合った映像を自動的に生成するこ

とができるのではと考えている。今後は、このような映像の自動生成にも取り組み、楽曲動画を視覚的にインタラクティブに楽しむことができる仕組みを実現していく予定である。

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 プロジェクトによるものです。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] Goto, M.: An Audio-based Real-time Beat Tracking System for Music With or Without Drumsounds, *Journal of New Music Research*, pp. 159–171 (2001).
- [4] 藤澤隆史, 谷光彬, 長田典子, 片寄晴弘：和音性の定量的評価モデルに基づいた楽曲ムードの色彩表現インタフェース, 情報処理学会論文誌, pp. 1133–1138 (2009).
- [5] 室伏空, 中野倫靖, 後藤真孝, 森島繁生：DanceReProducer: 既存のダンス動画の再利用により音楽に合った動画を作成できるシステム, 第17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 (WISS2009), pp. 63–68 (2009).
- [6] 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).
- [7] Goto, M.: 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*, Vol. 14, No. 5, pp. 1783–1794 (2006).
- [8] Yoshii, K. and Goto, M.: MusicCommentator: Generating comments synchronized with musical audio signals by a joint probabilistic model of acoustic and textual features, *Proceedings of the 8th International Conference on Entertainment Computing*, pp. 85–97 (2009).