

動画共有サービス上の時刻同期コメントを用いた 音楽動画探索システムに関する検討

木下 裕一朗^{1,a)} 佃 洸撰² 渡邊 研斗² 中塚 貴之² 中野 倫靖² 後藤 真孝² 中村 聡史¹

受付日 2025年5月13日, 採録日 2025年11月11日

概要: 未知の音楽に出会い、興味の幅を広げることは、人々の音楽体験を豊かにする。しかし、動画共有サービスには日々膨大な音楽動画が投稿されており、ユーザがキーワードやタグに基づく従来の検索を用いて、普段は聴かないクリエイターや音楽ジャンルの中から、新たに聴いてみたいと思える音楽動画に出会うことは困難である。本研究では、こうした問題を解決するために、音楽動画内の時刻に同期して投稿されたコメント（時刻同期コメント）を利用する。具体的には、時刻同期コメントの中には、その時刻における音楽動画に対する個性的な表現を含むコメントなど、そのコメントが投稿された音楽動画の視聴を動機づけるコメントが存在すると仮定し、そうしたコメントをきっかけとしてユーザが音楽動画を視聴するというアプローチを提案する。本稿では、提案するアプローチの有用性を検証するために、提示された複数の時刻同期コメントのうち、ユーザが興味を持って選択したコメントについて、そのコメントが投稿された時刻の手前から音楽動画を試聴できる実験用システムを実装し、ユーザ実験を実施した。実験の結果、提案するアプローチによって、ユーザが普段聴かないジャンルでかつ嗜好に合う音楽動画に出会えることや、ユーザが興味を持ったコメントはユーザ間で異なり、多様であることが明らかになった。

キーワード: 時刻同期コメント, 音楽動画, 音楽探索

A Study of a Music Video Exploring System Using Time-synchronized Comments on a Video Sharing Service

YUICHIRO KINOSHITA^{1,a)} KOSETSU TSUKUDA² KENTO WATANABE² TAKAYUKI NAKATSUKA²
TOMOYASU NAKANO² MASATAKA GOTO² SATOSHI NAKAMURA¹

Received: May 13, 2025, Accepted: November 11, 2025

Abstract: Encountering unfamiliar music and broadening one's interests can enrich the overall music experience. However, since an overwhelming number of music videos are uploaded daily to video sharing services, it is difficult for users to discover music videos from creators or genres they do not usually listen to by relying solely on conventional keyword- or tag-based searches. To address this issue, this study uses time-synchronized comments, which are comments posted in synchronization with specific timestamps in music videos. Specifically, we assume that some of these comments, such as those containing unique expressions about particular scenes, can serve as motivational triggers for users to watch the corresponding music videos. Based on this assumption, we propose an approach that encourages users to explore music videos by using these comments as entry points. To evaluate the effectiveness of the proposed approach, we implemented an experimental system. The system presents multiple time-synchronized comments to the user; when the user selects a comment of interest, they can start watching the music video from a point slightly before the timestamp at which the comment was posted. A user study using this system revealed that the proposed approach enabled users to discover music videos that matched their preferences even in genres they do not usually listen to, and that the comments users found interesting varied across individuals, indicating a diversity of interests.

Keywords: time-synchronized comments, music videos, music exploration

1. はじめに

動画共有サービスの普及によって、膨大な数の様々な音楽動画が視聴可能になり、ユーザは視聴する楽曲を決める際に検索や推薦を利用することが一般的になった。これまで、ユーザの音楽検索を支援するため、ハミング [1], [2], [3] や音楽コンテンツ [4], [5], [6], 画像 [7], [8] をクエリとした検索や、ソーシャルタグ [9], [10] を利用した検索手法が提案されてきた [11]。これらの検索は、ユーザの探したい楽曲が比較的明確に決まっている場合には有用であるが、ユーザが普段聴かないクリエイターやジャンルの膨大な音楽動画の中から、そのユーザの嗜好に合った楽曲を見つけたいときに適した検索とはいえない。新たなクリエイターを知ったり、好みの音楽ジャンルの幅を広げたりすることは、人々の音楽体験をより豊かにすることにつながる。そのため我々は、ユーザが普段聴かないクリエイターやジャンルの楽曲から、新たに自身の嗜好に合う楽曲を発見することの支援を目指している。

ユーザが普段聴かないクリエイターやジャンルの楽曲を積極的に聴かせるためには、そうした楽曲を聴きたいと思うきっかけを作ることが重要である。本研究では、そうしたきっかけとして、音楽動画内の時刻に同期して投稿されたコメント（時刻同期コメント）に着目する。時刻同期コメントは動画を視聴しながら投稿されるため、動画のシーンに応じたコメントが投稿される傾向にある [12]。そのため、音楽動画であれば、「ここのギターかっこいい」「走り出したくなるようなイントロだ」「この演出は泣ける」のように、楽曲の特定のパートや映像の特定のシーンに関するコメントも多く投稿される。そうしたコメントの中には、それが投稿されたシーンを視聴したいと思うような、視聴のきっかけとなるコメントが含まれるのではないかと我々は考えた。さらに、上記の例のように、時刻同期コメントには、そのコメントが投稿された音楽動画のクリエイターやジャンルを想起させないものも多いため、自身が普段聴く音楽には影響されずに、動画視聴のきっかけを作れるという利点もある。

以上の点をふまえて本研究では、ユーザの興味を引く時刻同期コメントをきっかけとして音楽動画を視聴するというアプローチを提案する。時刻同期コメントは気軽にたくさん投稿される特性があるため、時刻に同期していないコメントやレビュー評価が投稿されないような音楽動画に対しても、時刻同期コメントに基づく提案アプローチならば幅広く適用できるという利点がある。また、時刻同期コメ

ントが投稿されたシーンから音楽動画を視聴可能にすることで、時刻に同期していないコメントやレビュー文をきっかけとして動画を最初から視聴する場合よりも、ユーザの嗜好に合ったシーン・動画に出会える可能性が高まることを期待できる。さらに、嗜好に合うかどうかの判断が早まればユーザは次々と音楽探索を行えるため、普段聴かないジャンルの音楽動画に多く出会えることも期待できる。

提案アプローチの有用性を検証するために我々は、時刻同期コメントを提示し、ユーザが興味を持ったコメントを選ぶとそれが投稿された動画内の再生時刻の少し前（5秒前）から音楽動画を再生して探索できる実験用システムを実装した。システムを用いた実験を通して、興味を引くコメントにはどのような特徴があるか、また提案アプローチによって普段は聴かないジャンルでかつ嗜好に合う音楽動画をどの程度発見できるかを評価した。さらに、興味を引くコメントがユーザ間でどの程度一致するか検証した。

本研究の貢献を以下にまとめる。

- 興味を引く時刻同期コメントをきっかけとして音楽動画を視聴するというアプローチを提案し、実験により、自身が普段聴かないジャンルで嗜好に合う音楽動画を見つけることに対して、提案アプローチが有用であることを示した。
- 楽曲に対して高い評価を述べているコメントは人の興味を引く傾向があり、またコメント投稿者独自の視点が含まれるコメントや特定の音楽動画にのみあてはまるコメントも人の興味を引く可能性が高いことを明らかにした。
- 人が興味を引かれるコメントの一致度合いを検証し、評価の高さを表す語や音楽動画の特定のシーンに関連する語を含むコメントは、多くの人の興味を引く傾向にあり、かつ興味を引かれると感じる基準は人によって大きく異なり多様性があることも明らかになった。

2. 関連研究

2.1 楽曲のメタ情報に基づく楽曲の検索・探索

ユーザと楽曲の出会いの支援を目的として、楽曲のメタ情報であるタグ [10], [13], 歌詞 [14], [15], ジャケット画像 [16], [17] などに基づいて楽曲の検索や探索が可能なシステムが提案されてきた。たとえば TagFlip [10] では、ユーザが選択したタグの組合せに応じて楽曲の再生リストが生成される。Lyric Jumper [15] では、歌詞のトピックに基づいて楽曲を探索したり、特定のトピックに関連する歌詞のフレーズを推薦することでユーザが興味を持ったフレーズを含む楽曲に出会えたりする機能が実現されている。また MusicGalaxy [16] では楽曲間の類似度に応じて、宇宙空間に見立てた 2 次元空間中に楽曲のジャケット画像を表示し、興味を持ったジャケット画像の楽曲を聴ける。

タグ、歌詞、ジャケット画像のいずれのメタ情報も、楽曲

¹ 明治大学
Meiji University, Nakano, Tokyo 164-8525, Japan

² 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8568, Japan

a) kinoshita@nkmr-lab.org

のジャンル推定に使われることがあるように、ジャンルとの関連が強いことが知られている [18], [19], [20]. したがって、上記のシステムでは、表示されたメタ情報からユーザはジャンルの想像がついてしまうため、自身が普段聴くジャンルの楽曲を探索するきっかけとしては有用であっても、普段聴かないジャンルの楽曲を探索するきっかけとしては不十分であると考えられる。また、特定の音楽ジャンルを表す3つのパーツ（頭部、体、背景）を選択して自身のアバターを作成すると、選択されたパーツに基づいて音楽のプレイリストを生成するシステム [21] も存在するが、このシステムでは、ユーザは音楽ジャンルを想起可能なパーツを選択するため、普段聴かないジャンルの楽曲と出会うためには、意識的にそのジャンルを表現するパーツを選択しなければならない。楽曲に対するレビュー文については、レビュー文を用いて音楽ジャンルの推定を行った研究 [22] が存在するため、レビュー文には音楽ジャンルを想起できる情報が含まれている傾向があると考えられる。本研究では、ユーザに特定のジャンルを想起させることが少ない時刻同期コメントを用いるため、普段聴かないジャンルの音楽動画を視聴するきっかけとなる効果が期待できる。

2.2 動画に投稿されたコメントの活用

動画の再生時刻とは非同期のコメントを用いて、コメントをインデックスに追加して検索精度を改善する研究 [23], [24], タグに相当するような動画のメタ情報をコメントから抽出する研究 [25], [26], コメントから抽出した感情とシチュエーションに基づき音楽動画を推薦する研究 [27] などに取り組まれてきた。これに対して、時刻同期コメントを用いた研究では、動画のシーンに応じたコメントが投稿されやすいという特性を活用することで、動画の重要シーンの抽出 [28], [29] や、登場人物が注目されるシーンの抽出 [12] が可能であることが示されている。また、時刻同期コメントには、その瞬間ごとの感情をより多く表現する傾向があるという特性もある [30]。この特性を活用し、コメント内の語に着目することで、視聴者が動画にいだいた印象に基づく動画ランキングの生成 [31], 視聴者の反応に基づく音楽動画ランキングの生成 [32], 音楽動画の印象の分類 [33] といった研究が行われてきた。

以上の研究はいずれも、コメントを何らかの方法で集約することで動画の特徴を表し、動画の検索や分類などを実現している。それに対して本研究では、個々の時刻同期コメントが持つ、音楽動画に対する多様な表現が、音楽動画を視聴するきっかけとしてユーザにとって有用であると考え、コメントを集約せずに個々のコメントをそのまま表示する点に独自性がある。

3. 提案アプローチ

本研究では、ユーザが普段聴かないクリエイターやジャン

ルの音楽動画の中から、新たに自身の嗜好に合う楽曲を発見できるようにするため、音楽動画に投稿された時刻同期コメントに着目し、コメントをきっかけとして音楽動画を視聴するというアプローチを提案する。具体的には、複数の時刻同期コメントをユーザに提示し、その1つが選択された際に、それが投稿された動画内の時刻の少し前から再生可能な動画リンクを作成して音楽動画を視聴できるようにする。時刻同期コメントが投稿されたシーンの少し前から再生するのは、音楽動画を視聴しながら投稿される時刻同期コメントが、その言及箇所よりもわずかに遅れて投稿される傾向にあることをふまえて、コメントで言及されているシーンをユーザがより確実に視聴できるようにするためである。さらに、そのようにすることで、ユーザは音楽動画を最初から最後まで視聴しなくても、その音楽動画が自身の嗜好に合ったものか判断しやすくなる。楽曲のサビだけを次々に聴くことで自身の嗜好に合う楽曲であるかの効率的な判断を支援するシステム [34] はこれまでも提案されているが、我々のアプローチでは、ユーザは興味を持ったコメントが投稿されたシーンを視聴しながら次々と音楽動画を探索できる。また、時刻同期コメントにはクリエイター名や音楽ジャンル名といった音楽情報が含まれることは少ないため、提案するアプローチによってそうした情報に依存しない音楽動画の探索が可能になる。その結果、ユーザは自身の日頃の音楽の好みにとらわれずに、普段視聴しないタイプの音楽動画の中から新たに自身の嗜好に合ったものを発見できると我々は考える。

4. 提案アプローチの有用性検証

本章では、以下の問いに答えるため、提案するアプローチに基づいた音楽動画探索が可能なシステムを実装し、評価実験を行う。

RQ1 音楽動画を視聴したいと思う、興味を引くコメントにはどのような特徴があるか。

RQ2 提案するアプローチにより、普段は視聴しないタイプでありかつ嗜好に合う音楽動画をどの程度発見可能になるか。

本実験では、RQ2において普段は視聴しないタイプの音楽動画を判定するために、音楽のジャンル情報を用いる。

4.1 実験設計

4.1.1 使用データ

動画共有サービスは様々な存在するが、本稿では、時刻同期コメントの投稿が可能な「ニコニコ動画*1」に着目し、国立情報学研究所が提供している「ニコニコデータセット*2」に含まれるデータを利用する。ニコニコデータセットには、2007年3月6日から2021年9月30日の間にニ

*1 <https://www.nicovideo.jp>

*2 <https://www.nii.ac.jp/dsc/idr/nico>



図 1 表示された 3 つのコメントから興味を引かれたコメントを選択すると、そのコメントが投稿されたシーンの手前から音楽動画を視聴できる実験用システム

Fig. 1 An experimental system that allows users to select one of three displayed comments and watch the music video starting slightly before the scene where the selected comment was posted.

ニコニコ動画に投稿された約 2,000 万件の動画のデータが含まれる。具体的には、各動画のメタデータ（タイトル、動画 ID、タグなど）と時刻同期コメントデータ（コメント本文、コメントの投稿日時、コメントが投稿された動画内の再生時刻など）が含まれており、コメントの総数は約 41 億である。

本稿では、VOCALOID [35] などの歌声合成ソフトウェアを用いて創作されたオリジナル楽曲（以下、VOCALOID 楽曲）の音楽動画を対象とする。ニコニコデータセットに含まれる動画の中から対象となる音楽動画を取得するために、Hamasaki ら [36] による音楽視聴支援サービス Songrium^{*3}の分析データ（VOCALOID 楽曲かどうかの判定データ）を用いた。音楽動画に投稿された時刻同期コメントの中で、極端に短いコメントは多くの動画に書き込まれる定型的な表現であるものが多く、記号のみのコメントや極端に長いコメントは動画の内容とは無関係であるものが多い。そうしたコメントは興味を引くコメントである可能性は低いと考えられるため、連続して記述される記号・文字を 1 つにした際に 5 文字以下のコメント、記号のみのコメント、50 文字以上のコメントを除外した。また、過度に暴力的な表現などの不適切な表現を含むコメントを除外するために、単語のブロックリストを人手で作成し、ブロックリスト中の単語を 1 つでも含むコメントも除外した。

投稿されたコメントの数は音楽動画によって大きく異なるが、本実験では様々な動画に投稿されたコメントについて分析するために、上記のコメントの除外処理を行ったうえで各動画に投稿された直近 50 件のコメントのみを使用し、コメント数が 50 件に満たない動画は実験の対象外とした。以上の処理により、実験に使用するデータセットに含まれる動画数は 27,420、コメント数は 1,371,000 となった。

4.1.2 実験用システム

前項で作成したデータを用いて、時刻同期コメントを選

択するとそのコメントが投稿されたシーンの手前から音楽動画を探索できる実験用システムを実装した（図 1）。システムはコメントフィールドと動画フィールドからなり、システムにアクセスすると、コメントフィールドに時刻同期コメントが 3 件表示される。具体的には、システムはデータセットから 3 件の音楽動画をランダムに選択し、さらにそれぞれの動画に投稿された 50 件のコメントから 1 件ずつランダムに選択したコメントを表示する。実験参加者が興味を引かれたコメントを選択すると、動画フィールドにそのコメントが投稿された動画を視聴するためのプレーヤ（ニコニコ動画が提供する外部プレーヤ）と選択したコメントが表示される。プレーヤの再生ボタンを押すと、ユーザが選択したコメントが投稿されたシーンの少し前から動画が再生され、時刻同期コメントも動画上に表示される。本実験システムでは、ニコニコ動画の視聴者が何らかの感想を持ったとき、反応してから 2 秒程度遅れてコメントを投稿する傾向 [37] や、あるコメントに関連したコメントがその前後 5 秒以内に投稿されている可能性が高い [38] ことをふまえ、動画は選択したコメントが投稿された 5 秒前から再生されるようにした。

本実験では、表示したコメントすべてがクリックされてしまうと、ユーザが実際に興味を引かれるコメントの特徴を分析しにくくなるため、3 件のコメントのうち 1 件だけしか選択できないように制限した。コメントを選択した後にコメントフィールドの「他のコメントを見る」のボタンを押すことで、新たな 3 件のコメントが表示される。表示された 3 件のいずれにも興味を引かれなかった場合は、コメントを選択せずに「他のコメントを見る」のボタンを押すこともできる。本実験では、より多くのコメントを分析するために、ある実験参加者に表示されたコメントは、その後すべての実験参加者に対して表示されないようにした。つまり、各コメントが表示されるのは実験を通して最大でも 1 回のみとした。さらに、コメントをきっかけとす

^{*3} <https://songrium.jp>

ることで嗜好に合う音楽動画を発見できたかどうかを検証するため、動画フィールドに高評価（嗜好に合う）と低評価（嗜好に合わない）のボタンを設けた。ただし、動画を視聴した際に評価のボタンを押すことは必須ではない。

4.1.3 実験手順

実験参加者に実験内容と注意を記載した Web ページを閲覧してもらった後、我々が実装した実験用システムを 10 日間使用してもらった。いずれの実験参加者も自身の PC を用いて、任意の Web ブラウザを通して実験用システムにアクセスし、ユーザ登録をしたうえでシステムを使用した。各実験参加者は 3 件 1 組で表示されるコメントセットを評価し、3 件のコメントから 1 件を選んだ場合には、必ず音楽動画を視聴するよう指示をした。なお、音楽動画を最後まで視聴することは強制しなかったため、実験参加者は任意のタイミングで動画の視聴を止めることができた。表示された 3 件の中に興味を引くコメントが 1 つもなかった場合は、実験参加者は「他のコメントを見る」のボタンを押し、システムが表示する新しいコメントセットを評価した。

システム利用実験の終了後、実験参加者は Web 上でアンケートに回答した。アンケートの項目は、性別と年齢、日常生活におけるニコニコ動画や VOCALOID 楽曲の視聴頻度、10 種の各音楽ジャンル（ポップス・ロック・ダンス・ジャズ・ラテン・クラシック・行進曲・ワールド・声楽・邦楽^{*4}）の音楽を日常的に聴く頻度、実験用システムで視聴した楽曲の中に普段聴かない音楽ジャンルのものがどの程度あったか、どのようなコメントのときに実際にその動画を視聴したいと思ったか、実験用システムを今後も継続して利用したいか、から構成される。

4.2 実験結果

実験参加者は、筆頭著者が所属する研究室の 20 歳から 24 歳の学生 20 名（女性 5 名、男性 15 名）で、平均年齢は 21.7 歳であった。実験参加に対する報酬は 3,000 円であった。実験参加者のうちニコニコ動画を月に 1 回以上利用する者は 4 名、VOCALOID 楽曲を月に 1 回以上聴く者は 12 名であった。20 名の実験参加者に 10 日間実験用システムを使用してもらった結果、4,148 件の音楽動画が視聴された。選択されたコメント（「選択コメント」と呼ぶ）の数は 4,480 であり、そのうち動画視聴後に高評価を得た数は 2,811、低評価を得た数は 973、いずれの評価も得られなかった数は 696 であった。システムのコメントフィールドに表示された 3 つのコメントのうちの 1 つが選択されたために選択されなかったコメント（「非選択コメント」と呼ぶ）の数は 8,960、3 つのいずれも選択されずに「他のコメントを見る」ボタンが押されたコメント（「スキップコメント」と呼ぶ）の数は 4,098 であった。なお、4,480 件の選択コメントのうち、85 件のコメントは削除などの理由で音楽動画の再生時にプレーヤ上に表示されなかった。プレーヤ上でコメントの存在が確認できなかったことによる評価への影響を考慮し、本節ではこの 85 件を除いた 4,395 件の選択コメントを分析に使用する。

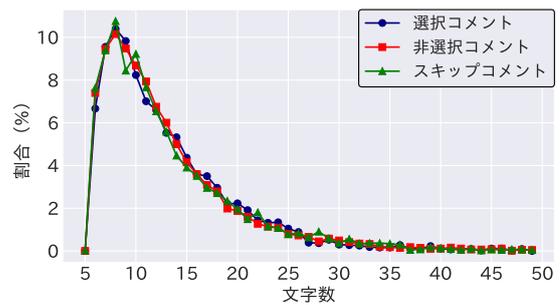


図 2 選択・非選択・スキップコメントの文字数の分布

Fig. 2 Distribution of the number of characters in selected, non-selected, and skipped comments.

ト」と呼ぶ)の数は 4,098 であった。なお、4,480 件の選択コメントのうち、85 件のコメントは削除などの理由で音楽動画の再生時にプレーヤ上に表示されなかった。プレーヤ上でコメントの存在が確認できなかったことによる評価への影響を考慮し、本節ではこの 85 件を除いた 4,395 件の選択コメントを分析に使用する。

4.2.1 選択された時刻同期コメントの分析

選択コメント・非選択コメント・スキップコメントそれぞれの文字数の分布を図 2 に示す。3 種類（選択・非選択・スキップ）のコメントすべてにおいて 7~10 文字のものが多く、16 文字以上のコメントは 4% 未満であった。全体的な分布も 3 種類の間で大きな差はなく、平均文字数は選択コメントが 13.1、非選択コメントが 13.0、スキップコメントが 13.1 であった。また、各動画を長さ（秒数）に応じて 3 つの区間（序盤・中盤・終盤）に等分割し、3 種類のコメントが各区間に出現する割合も比較したが、大きな違いはみられなかった（たとえば、選択コメントの割合は序盤が 43.4%、中盤が 29.8%、終盤が 26.8% であった）。

次に、3 種類の各コメントに含まれる具体的な語に基づいて分析を行うために、MeCab [40] を用いて各コメントを形態素に分割した。表現の特徴が強く表れると思われる名詞と形容詞のみを対象として、まず 3 種類のコメントすべてを合わせた全体のコメント集合（全コメント）において各形態素が含まれるコメントの割合を求めた。そのうえで、選択・非選択・スキップそれぞれのコメント集合において、各形態素が含まれる割合から、全コメントにおいて同じ形態素が含まれる割合を引いたときの値を求めた。たとえば、選択コメントにおいて「曲」という形態素が含まれる割合は 10.5% であり、全コメントにおいて「曲」が含まれる割合は 6.9% であるので、選択コメントにおける「曲」の値は 3.6% となる。この値が大きい形態素は、そのコメント集合に特有の形態素であるといえる。また、上記で求めた、選択・非選択・スキップの 3 つのコメント集合における、全コメント集合と比較したときに出現割合の差が大きかった形態素に対して、出現割合に有意な差があるか求めため、 z 検定を行った。多重比較を行うため、 p 値には

*4 10 種類の音楽ジャンルは「RWC 研究用音楽データベース：音楽ジャンル [39]」の「大分類」を採用した。

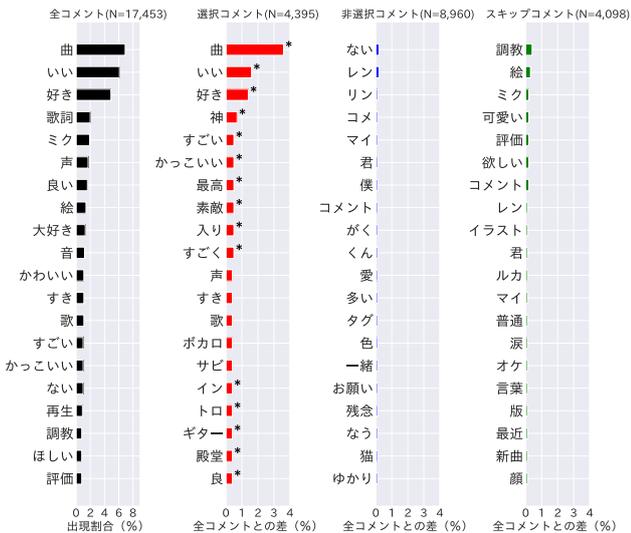


図 3 全コメントにおいて出現割合上位 20 の形態素と、選択・非選択・スキップのそれぞれのコメントにおいて各形態素が含まれる割合から、全コメントにおいて同じ形態素が含まれる割合を引いた値が上位 20 の形態素。全コメントにおける出現割合との差が有意である形態素については、棒グラフに*記号を付与している

Fig. 3 The top 20 morphemes based on their occurrence ratio in all comments, and the top 20 morphemes obtained by subtracting the occurrence ratio of each morpheme's appearance in all comments from the ratio of its appearance in selected, non-selected, and skipped comments. Bars marked with an asterisk (*) indicate morphemes whose occurrence ratios differ significantly from those in all comments ($p < .05$).

ボンフェローニ補正を適用し、有意水準は 0.05 を用いた。

選択・非選択・スキップの各コメントにおける、上記の値が上位 20 の形態素を図 3 に示す。参考のため、図 3 には全コメントで出現割合が高かった上位 20 の形態素も掲載している。また、全コメント集合と比較したときの出現割合に有意差が認められた形態素については、棒グラフに*記号を付与している。全コメントでは「曲」「いい」「好き」の出現割合が高いが、選択コメントでもそれらの値が特に高く、全コメントと比較して有意に高いことが分かった ($p < .05$)。したがって、「曲」について言及しているコメントや、音楽動画の何かしらの側面を「いい」あるいは「好き」と述べているコメントは、人の興味をより引きやすく、選ばれる回数が多かったといえる。また、選択コメントでは「神」「すごい」「最高」の値も有意に高く ($p < .05$)、特に高く評価する際に用いられる語が含まれるコメントも選ばれやすいことが分かる。実験後に実施したアンケートでは、「肯定的なコメント」や「評価がかなり高いと受け取れるもの」が表示されたときに実際に動画を視聴したいと思ったという回答が得られており、これらの結果とも一致していた。一方で、非選択コメントやスキップコメントでは、バーチャルシンガーの名前である「レン」「リン」「ミ

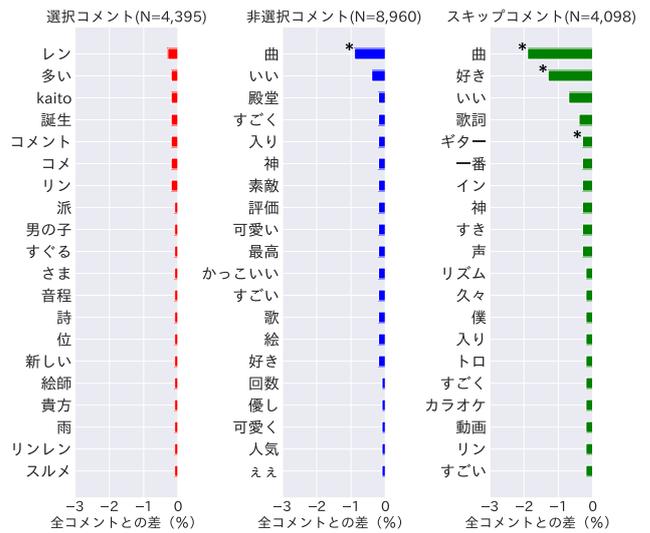


図 4 選択・非選択・スキップのそれぞれのコメントにおいて各形態素が含まれる割合から、全コメントにおいて同じ形態素が含まれる割合を引いたときの値が下位 20 の形態素。全コメントにおける出現割合との差が有意である形態素については、棒グラフに*記号を付与している

Fig. 4 The bottom 20 morphemes obtained by subtracting the occurrence ratio of each morpheme's appearance in all comments from the ratio of its appearance in selected, non-selected, and skipped comments. Bars marked with an asterisk (*) indicate morphemes whose occurrence ratios differ significantly from those in all comments ($p < .05$).

ク」「ルカ」や、歌声合成パラメータ調整に関連した「調教」の値が、全コメントと比較してわずかに高かったが、有意差は認められなかった ($p > .05$)。これは、実験参加者のうち 4 割が VOCALOID 楽曲をまったく聴かない人であったことが影響している可能性がある。実験後のアンケートでは、「知らない専門用語が入っているコメント」は表示されても選択しなかったという回答が得られており、たとえばスキップコメントで値の高かった「調教」はこれに該当すると考えられる。

次に、先ほどとは逆に、3 種類のコメントにおける各形態素の出現割合から、全コメントにおけるそれぞれの形態素の出現割合を引いたときに、値が下位 20 の形態素を図 4 に示す。この値が小さい形態素は、そのコメント集合に固有の出現しにくい形態素であるといえる。全コメント集合と比較したときの出現割合に有意差が認められた形態素については、棒グラフに*記号を付与している。選択コメントでは、バーチャルシンガーの名前である「レン」「kaito」「リン」の値が全コメントと比較してやや低かったが、有意差は認められなかった ($p > .05$)。また、非選択コメントやスキップコメントの両方で値が低かった語として、「曲」「いい」「好き」が含まれており、両方のコメント集合において「曲」の値は有意に低く ($p < .05$)、スキップコメントにおいては「好き」の値が全コメントと比較して有意に

低かった ($p < .05$). さらに、非選択コメントでは「殿堂入り」「殿堂」および「入り」「神」「素敵」などの高く評価する語が、スキップコメントでは「歌詞」「リズム」「イントロ」「イン」および「トロ」といった楽曲の構成要素に関連する語がみられた。しかし、これらの語の選ばれやすさについては有意差が認められなかった ($p > .05$)。

以上のように、本稿では選択コメントの一般的な傾向を調べたが、選択コメントには「このレトロなビートが強烈」「なんという疾走感のあるヤキイモ」「夏の切なさ、繊細さ、美しさが全部揃ってる」「これ聞きながらドライブすると自然とアクセルと踏んでしまう・・・」のように、コメントを投稿したユーザの独自の視点が含まれるコメントや、特定の音楽動画にのみあてはまるコメントも含まれていることが分かった。今後はそうしたコメントも対象として分析を行い、人の興味を引くコメントの性質に対する理解を深めたい。また、本稿では形態素に着目し、選択・非選択・スキップコメントにおいて出現割合が高い名詞や形容詞を求めることで、3種類のコメントの特徴を分析したが、今後はそれぞれのコメントに対して自然言語処理モデルを用いて分析を行い、文脈理解やクラスタリングなどを実施することで、人の興味を引く表現についてさらに明らかにしていきたい。

4.2.2 視聴された音楽動画の分析

実験参加者によって視聴された音楽動画のジャンルを調べるため、VOCALOID 楽曲などを対象とした音楽発掘サービス「Kiite」*5において「ジャンルフィルター」として採用されている音楽動画のタグのうち、4.1.3 項で述べた10種類の各ジャンルに対応するタグを手手で抽出し、「ボカロバラード」「ボカロレゲエ」など、計43のタグを得た。ニコニコデータセットの動画メタデータに含まれるタグを調べた結果、実験で使った27,420件の音楽動画で10種類中1つ以上のジャンルに対応するタグを含む動画は9,238件、実験で視聴された4,148件の音楽動画で10種類中1つ以上のジャンルに対応するタグを含む動画は1,516件であった。

図5に、上記の4,148件と1,516件それぞれの音楽動画集合における各ジャンルの割合を示す。データセットにおける各ジャンルの割合と、視聴された音楽動画の各ジャンルの割合はおおむね一致していた。この結果は、時刻同期コメントから特定のジャンルが想起されることが少ないことを示唆しており、ユーザが興味を引かれるコメントを選択すると、母集団の動画集合におけるジャンルの割合に応じて各ジャンルの動画を視聴することになる。実験参加者ごとの、視聴した音楽動画におけるジャンルの内訳を図6に示す。多くの実験参加者が様々なジャンルの音楽動画を視聴しており、視聴した動画に含まれる平均音楽ジャンル

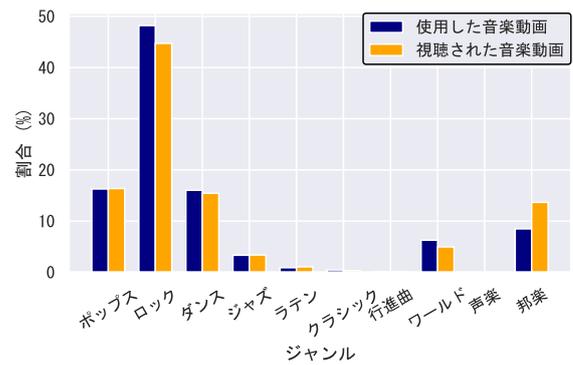


図5 実験に使用した音楽動画全体の音楽ジャンルの分布と、実験で視聴された音楽動画の音楽ジャンルの分布

Fig. 5 Distribution of music genres across all music videos used in the experiment and those viewed by participants during the experiment.

数は6.4であった。本実験で使った音楽動画の多くはジャンルがロックであったため(図5)、どの実験参加者においてもロックの占める割合が高い結果となった。

次に、時刻同期コメントをきっかけとすることで、普段聴かないジャンルで嗜好に合う音楽動画をどの程度見つけられるかを評価した。各ジャンルの楽曲を日常的に聴く頻度は5段階リッカート尺度(1:まったく聴かない, 5:ほぼ毎日聴く)を用いて質問した。本実験では、2以下の回答をした場合に、そのジャンルを実験参加者が普段聴かないジャンルであると見なした。ジャンルタグを持つ音楽動画の中で、実験参加者 $u \in \{A, B, \dots, T\}$ がコメントを選択して視聴した音楽動画集合を V^u 、そのうち u が普段聴かないジャンルでかつ高評価をした動画集合を $V^u_{unfam+fav}$ とする。 V^u に対する $V^u_{unfam+fav}$ の割合は、興味を引かれたコメントをきっかけにして、普段聴かないジャンルで嗜好に合う音楽動画を見つえられる確率を表す。表1に、実験参加者ごとの数値を示す。割合のマクロ平均は0.384であった。標準偏差は0.256であったため、実験参加者によって割合にばらつきはあるものの、興味を引かれたコメントを選択して音楽動画を視聴すると、平均38.4%の確率で普段聴かないジャンルでかつ嗜好に合う音楽動画に出会えることが示された。竹川ら[41]が実施した実験では、ユーザがランダムに音楽を探索したときに、自身の嗜好に合った音楽に出会えた割合は、31.3%や33.9%*6であった。ユーザがランダムに音楽探索を行ったとき、普段聴かないジャンルでかつ自身の嗜好に合った音楽に出会える確率は、これらの値よりも低くなると予想されるが、本研究では、提案アプローチにより平均38.4%の確率で遭遇できていたことから、提案アプローチは普段聴かないジャンルでかつ嗜好に合う音楽の発見に有用であると考えられる。

さらに、実験参加者 u がコメントを選択して視聴した、

*6 文献[41]では、参加者を2つのグループに分けて実験を実施していたため、各グループにおいて嗜好にあった音楽に出会えた割合が示されていた。

*5 <https://kiite.jp>

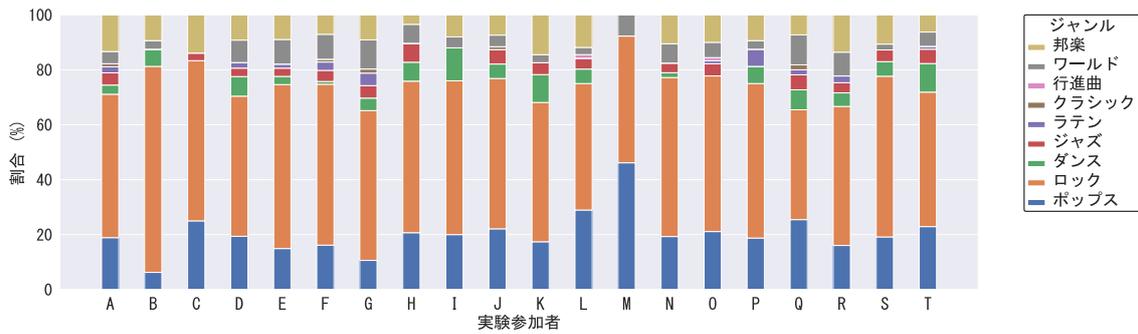


図 6 各実験参加者が視聴した音楽動画の音楽ジャンルの内訳

Fig. 6 Breakdown of music genres in music videos viewed by each participant.

表 1 実験参加者 u がコメントを選択して視聴した音楽動画集合 V^u のうち、普段聴かないジャンルでかつ高評価をした動画集合 $V_{unfam+fav}^u$ の割合

Table 1 The proportion of $V_{unfam+fav}^u$ within the set of music videos V^u that participant u viewed by selecting comments, where $V_{unfam+fav}^u$ represents the set of videos from genres that u does not usually listen to but rated highly.

実験参加者	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V^u	91	32	37	99	69	106	66	29	26	99	70	76	13	59	91	32	56	82	97	97
$V_{unfam+fav}^u$	64	2	23	46	49	47	3	19	0	18	13	34	10	11	49	1	22	49	48	14
割合	0.703	0.063	0.622	0.465	0.710	0.443	0.046	0.655	0	0.182	0.186	0.447	0.769	0.186	0.539	0.031	0.393	0.598	0.495	0.144

表 2 実験参加者 u がコメントを選択して視聴した、普段聴かないジャンルの音楽動画集合 V_{unfam}^u のうち、高評価をした動画集合 $V_{unfam+fav}^u$ の割合

Table 2 The proportion of $V_{unfam+fav}^u$ within the set of music videos V_{unfam}^u that participant u viewed by selecting comments and that belong to genres u does not usually listen to.

実験参加者	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V_{unfam}^u	74	6	28	72	58	80	23	23	6	23	16	54	13	11	91	6	47	69	79	27
$V_{unfam+fav}^u$	64	2	23	46	49	47	3	19	0	18	13	34	10	11	49	1	22	49	48	14
割合	0.865	0.333	0.821	0.639	0.845	0.588	0.130	0.826	0	0.783	0.813	0.630	0.769	1.0	0.538	0.167	0.468	0.710	0.608	0.519

普段聴かないジャンルの音楽動画集合を V_{unfam}^u としたときに、 V_{unfam}^u に対する $V_{unfam+fav}^u$ の割合を求めた。これは、普段聴かないジャンルの音楽動画集合を視聴したときに、その動画が嗜好に合う動画である確率を表す。実験参加者ごとの数値を表 2 に示す。割合のマクロ平均は 0.603 (標準偏差は 0.269) であった。普段聴かないジャンルの音楽動画であるにもかかわらず、嗜好に合っていた確率が平均 60.3% というのは高い値であるといえる。これは、自身が興味を引かれたコメントが音楽動画を視聴するきっかけになっているためであると考えられる。これがいかに高い値であるかは、普段聴かないジャンルの音楽動画をランダムに選択して視聴させたときに、高評価をする動画の割合を求めて比較することなどで検証できるため、今後より詳細な検証を行いたい。

実験後に実施したアンケートでは、本システムで視聴した音楽動画の中に普段聴かないジャンルの動画がどの程度あったか、5段階リッカート尺度 (1: まったくなかった, 5: とても多かった) で回答を得た。その結果、20 名の回答の平均値は 3.15 (1: 1 名, 2: 7 名, 3: 3 名, 4: 6 名, 5: 3 名) であり、12 名が 3 以上の回答をしていることか

ら、ユーザの主観的な観点からも、普段聴かないジャンルの音楽動画との出会いを実現できていることが示された。

4.2.3 実験用システムの定性評価

実験終了後のアンケートでは、実験用システムを今後も継続して利用したいかという質問に対して、20 名の実験参加者のうち 14 名がシステムを今後も使用したいと回答した。その理由として、「投稿日時が新しい順にソートすると、最近投稿された曲が一度に出てきてしまい、動画を取捨選択しなければならない負担があった。そのため、このシステムのようにコメントから選ぶとすべて視聴しなければいけない気持ちが無くなり、動画選択の負荷が軽くなった。」「コメントの箇所から再生すると人気のあるメロディーの箇所やサビなどから再生されることが多かったので、好みの曲を見つけやすかった」「ボカロ曲はよく聞けるが、自分の知っている曲を何度も聞いており、新しい曲を探すことが少ない。また探そうとしても、サムネイルとタイトルのみで自身の好みに合うかどうかを判別するのが難しい。そのため動画内のコメントから曲を探せる本システムは非常に便利だと感じた」などの回答が得られた。一方、システムの継続使用に対して肯定的でなかった 6 名の

実験参加者のうち、5名はシステムの改善があれば今後も使用したいと回答し、1名は自身の気に入っている曲のみを聴きたいためシステムを利用したいとは思わないと回答した。システムの改善点には、自身が高評価したコメントや動画の保存機能の追加や、VOCALOID 楽曲だけでなくより多くの音楽動画を視聴できるようにすることなどがあげられた。

本実験では、実験参加者がシステムの操作に戸惑ったり、実験参加者の認知負荷を高めたりしないように、意図的に簡素なインターフェースの実験用システムを実装して使用した。今後は上記の「高評価したコメントや動画の保存機能」をはじめとする、ユーザにとってより使いやすいシステムとなる機能を追加したり、インターフェースを洗練させたりすることで、実際にユーザが使用するためのシステムを実装する予定である。そのようにして実装したシステムを、最終的には誰もが利用できるように Web サービスとして公開したいと考えている。

本実験システムでは、コメントの言及箇所を確実に視聴できるようにするために、先行研究 [37], [38] をふまえてコメントの投稿時刻の 5 秒前から音楽動画が再生されるようにしたが、異なる秒数にした場合にユーザ体験がどう変わるかは今後の研究の余地がある。また、本実験の参加者は研究室内の学生に限定されていたため、システムの定性評価の結果が肯定的になるバイアスが生じた可能性がありうる。一般に同一研究室の学生を参加者とする実験では、参加を拒否することが心理的に難しく、匿名性が確保しにくいという問題が指摘されている。加えて、実験実施者は学生の筆頭著者だったが、参加者である学生の学業評価に影響を及ぼす立場だとも認識された場合には、公平性が損なわれる可能性が否定できない。システムの継続利用意図をより確かな形で検証するには、研究室外の環境で参加者を募集して実験を実施することが今後の研究として考えられる。

本稿はニコニコ動画に投稿された VOCALOID 楽曲を用いて実験を行ったが、今後は様々な音楽動画も対象にしたいと考えている。また、コメントの投稿シーンの手前から音楽動画を再生する本アプローチに対して、たとえばコメントを選択した後に音楽動画を最初から再生する従来の方法を用いて比較実験を行うなどして、提案するアプローチの有用性をさらに検証したい。

5. 興味を引くコメントの一致度の検証

4 章の実験では、より多くのコメントを分析することで人の興味を引くコメントの傾向を明らかにするために、各コメントが表示される回数は実験を通して最大でも 1 回のみとした。そうして多数の評価データを収集することで、4.2.1 項の分析により、興味を引くコメントや興味を引かないコメントには一定の傾向があることが示された。その

一方で、我々は分析を行うなかで、同じような内容のコメントであっても、選択・非選択・スキップのコメント集合のうち、複数のコメント集合に含まれるコメントが一定数存在することに気づいた。また、選択される傾向にあった肯定的なコメントや高く評価するコメントの中には、具体性が高いコメントとそうでないコメントが混在しており、どのようなコメントがより多くの人の興味を引くコメントであるのか、4 章の実験では十分に明らかにできなかった。そこで、我々は、興味を引くコメントが人々の間でどの程度一致するのかという新たな疑問をいだいた。

ユーザが個々に興味を引くコメントに基づいて動画を探索することの有用性は 4 章の実験により示されたので、興味を引くコメントの人々の間の一致度自体はその有用性には影響しないが、興味を引くコメントについてより深く理解し、高確率でユーザの興味を引くコメントを提示することができれば、提案アプローチに基づく音楽探索のユーザ体験を向上させることができると考えられる。興味を引くコメントについてより深く理解するために、本章では、以下の問いに答える。

RQ3 音楽動画を視聴したいと思う、興味を引くコメントは人々の間でどの程度一致するか。

5.1 実験設計

5.1.1 評価に使用したコメントセット

本実験では、4 章の実験で使用したコメントセットからサンプリングした各コメントセットに対して、興味を引くコメントを選択する作業を複数名に行ってもらうことで、興味を引くコメントの一致度合いを調べる。具体的なサンプリング方法を以下に述べる。まず、興味を引かないコメントだけを含むコメントセットが極力含まれないようにするため、4 章の実験での選択コメントを含む 4,480 件のコメントセットを抽出した。そのうえで、コメントの文字数による評価の一致度合いを検証するために、選択コメントの文字数が、6 から 10 文字、11 から 15 文字、16 から 20 文字、21 から 25 文字、26 から 49 文字のそれぞれの範囲にあるコメントセットを 80 件ずつ、計 400 件をランダムサンプリングし、実験に用いた。

5.1.2 実験手順

本実験では、実験参加者を等しい人数になるようランダムに 2 つのグループ (A グループ、B グループと呼ぶ) に分け、上記の 400 件のコメントセットを重複のない 200 件ずつに分割し、各グループの実験参加者が 200 件に対して評価を行った。200 件ずつに分割する際は、選択コメントの文字数が 6 から 10 文字のコメントセット 80 件からランダムに選択した 40 件を A グループ用、残りの 40 件を B グループ用とする、というように、選択コメントの文字数を考慮して分割した。

実験参加者は、4.1.2 項と同じ実験用システムを用いて、

200 件のコメントセットに対する評価を 10 日間かけて行った。4.1.3 項と同様、実験参加者は任意の Web ブラウザを通して実験用システムにアクセスし、ユーザ登録をしたうえでシステムを使用した。評価時には、システムが表示する 3 件のコメントのうち、興味を持ったコメントを 1 つのみ選んでもらい、コメントを選択した際は必ず音楽動画を視聴するよう指示した。コメントセット内の 3 件のいずれのコメントも興味を引かれなかった場合は、実験参加者は「他のコメントを見る」ボタンを押し、その場合は次のコメントセットをシステムが提示した。200 件のコメントセットを提示する順序は実験参加者ごとにランダムになるようにした。ただし、各コメントセット内の 3 件のコメントの順序は、4 章の実験参加者に提示した順序と同じにした。200 件のコメントセットの評価終了後、実験参加者は 4.1.3 項と同様の項目から構成されるアンケートに回答した。

5.2 実験結果

本実験の参加者は、4 章の実験参加者とは異なる、VOCALOID 楽曲を日常的に聴く 18 歳から 24 歳の学生 20 名（女性 4 名、男性 16 名）で、平均年齢は 20.7 歳であった。本実験の参加者は筆頭著者が所属する研究室外の学生であり、実験参加に対する報酬は 3,000 円分の Amazon ギフトカードであった。

実験参加者 20 名のうちランダムに選んだ 10 名を A グループとし、残りの 10 名を B グループとした。10 日間の評価後に各実験参加者の評価結果を調べたところ、表示されたコメントセットに対して、100 回以上連続して「他のコメントを見る」ボタンを押してスキップした参加者が、A グループでは 1 名、B グループでは 2 名存在した。これらの参加者によるデータは、分析対象として適切でないと考えて分析の対象から除外した。また、A グループでは 200 件のコメントセットのうち 3 件が、B グループでは 200 件のコメントセットのうち 9 件が正しく記録されていなかったため、分析対象から除いた。その結果本節では、A グループ 9 名による 197 件のコメントセットと、B グループ 8 名による 191 件のコメントセットを合わせた、計 388 件のコメントセットに対する評価データについて分析する。

5.2.1 選択されたコメントの一致度

分析対象となった 17 名の実験参加者において、表示された 3 つのうちどのコメントも興味を引かず「他のコメントを見る」ボタンが押された回数の平均は 11.4 で、標準偏差は 14.5 であった。実験参加者が選択したコメントの一致度は Krippendorff's alpha [42] を用いて求めた。この指標は -1 から 1 の間の値をとり、値が大きいほど実験参加者間の選択の一致度が高いことを表す。各グループ内で Krippendorff's alpha を求めたところ、A グループでは

表 3 A, B それぞれのグループで、4 章の実験で選択されたコメントを選んだ人数の割合のマクロ平均

Table 3 The macro-average of the proportion of participants in groups A and B who selected comments chosen in the experiment of Section 4.

文字数の範囲	6–10	11–15	16–20	21–25	26–49
割合	0.365	0.381	0.419	0.416	0.406

0.347、B グループでは 0.355 であった*7。また、一致度の高い実験参加者のペアが存在するか調べるため、各グループにおいて参加者を 2 名ずつ抽出し、すべての組合せに対して Krippendorff's alpha を求めた。その結果、両グループの中での Krippendorff's alpha の最大値は 0.476、最小値は 0.201 であった。一般的には、この値が 0.67 以上であれば中程度の一致度と見なされるため、これらの結果からは、実験参加者間の選択一致度は低く、興味を引くコメントには個人差があることが確認された。

次に、各コメントセットにおいて、4 章の実験で選択されたコメントが人の興味を引くコメントであると仮定し、その選択コメントを選んだ人数の割合を求めた。A, B それぞれのグループで選択コメントを選んだ人数の割合をコメントセットごとに算出し、そのマクロ平均を求めた結果、0.397 であった。さらに、選択コメントの文字数の範囲ごとに同様に割合を求めた結果を表 3 に示す。文字数が 6 から 10 文字と 11 から 15 文字のときの割合がそれぞれ 0.365 と 0.381 であるのに対して、文字数が 16 文字以上の場合はいずれも割合が 0.4 を超えていたことから、興味を引かれたコメントの文字数が多いほど、他の人もそのコメントに興味を引かれることが多いことが明らかになった。たとえば、「これから頑張れる (((「この歌詞ならもっと荘厳に仕上がるはずなのに・・・ミックスしたくなるぞ」「最後盛り上がることもうさくなくて音の調和が完璧にとれてて神」というコメントセットにおいて、2 つ目のコメントは 33 文字で 3 つ目のコメントは 30 文字と同程度に長いコメントであったが、9 名中 7 名の実験参加者が、選択コメントである「最後盛り上がることもうさくなくて音の調和が完璧にとれてて神」を選んでいった。

5.2.2 多くの実験参加者が選択したコメントの特徴

選択コメントを選んだ人数の割合が 0.5 以上であったコメントと、0.5 未満であったコメントそれぞれについて、MeCab を用いて形態素に分割し、表現の特徴が強く表れると思われる名詞と形容詞を対象として、各コメントに含まれる特徴的な語について分析した。選択コメントを選んだ人数の割合が 0.5 以上/未満のコメントにおいて各形態素が含まれる割合から、本実験で使用したすべての選択コメントにおいて各形態素が含まれる割合を引いた値が上位 20

*7 Krippendorff's alpha の計算には Castro により公開されているライブラリ [43] を使用した。

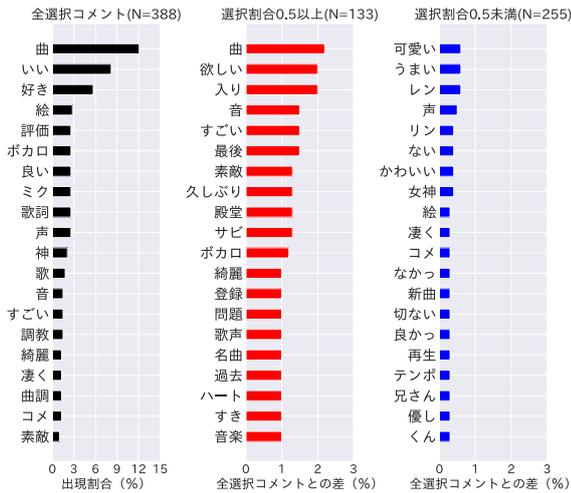


図 7 選択コメントを選んだ人数の割合が 0.5 以上であったコメントと、0.5 未満であったコメントそれぞれにおいて各形態素が含まれる割合から、すべての選択コメントにおいて同じ形態素が含まれる割合を引いた値が上位 20 の形態素。全コメントにおける出現割合と比較して、有意に出現割合が高かった形態素は存在しなかった ($p > .05$)

Fig. 7 The top 20 morphemes obtained by subtracting the proportion of each morpheme appearing in all selected comments from the proportion of each morpheme appearing in comments with a selection rate of 0.5 or higher, and in comments with a selection rate of less than 0.5. No morphemes were found to have significantly higher occurrence ratios compared to all comments ($p > .05$).

の形態素を図 7 に示す。出現割合に統計的に有意な差があるか χ^2 検定を行った。その際、多重比較を行うため、 p 値にはボンフェローニ補正を適用し、有意水準は 0.05 を用いた。選択割合が 0.5 以上のコメントでは、「すごい」「素敵」「殿堂入り」(「殿堂」および「入り」)「名曲」といった高い評価を表す語の値が比較的高く、「サビ」や「最後」といった音楽動画の特定のシーンに関する語の値も高い傾向がみられたが、いずれの形態素も全コメントにおける出現割合と比較して有意差は認められなかった ($p > .05$)。一方で、選択割合が 0.5 未満のコメントでは、「可愛い」「うまい」といった肯定的な語がみられたもののその種類は少なく、音楽動画の特定のシーンに関連する語はみられなかった。

以上のように、多くの人の興味を引くコメントの傾向は示された一方で、最も一致度の高い実験参加者同士であっても Krippendorff's alpha の値は 0.476 であったことから、ユーザにはそれぞれ固有の興味を引かれるコメントのタイプが存在すると考えられる。実験後のアンケートでも、「歌の動画でこんなコメントを打つ機会なんてあるのか?と思わせるようなコメント」のときにそのコメントが投稿された音楽動画を視聴したいと思ったと答えた実験参加者がいた一方で、「このギターすごいやこの曲いいなどのときに見たいと思った」と回答した実験参加者も存在し、興味を引かれると感じる基準は様々だった。本章の冒頭でも述

べたように、そうした基準が多様で、結果的に興味を引くコメントの一致率が低くても、本研究の学術的価値が損なわれるわけではなく、人によって多様な基準を持っていることを明らかにしたこと自体が有用な知見である。

今後はそうした多様性があることをふまえて、自然言語処理モデルを用いるなどしてコメントに対してさらなる分析を行い、ユーザそれぞれの興味を引くコメントのタイプの特定を目指す。ユーザの好みに応じたコメントの推薦を行えるようにすることで、提案アプローチに基づく音楽動画探索の満足度の向上を実現する。

6. おわりに

本稿では、ユーザが普段は聴かないタイプの音楽動画の中から、新たに聴いてみたいと思える音楽動画に出会えるようにするため、そのユーザが興味を持った時刻同期コメントをきっかけとして音楽動画を視聴するアプローチを提案した。実験を通して、ランダムに提示されたコメントの中から、ユーザが興味を持ったコメントを選択して音楽動画を視聴することで、そのユーザが普段は聴かないジャンルでかつ嗜好に合う音楽動画に出会えることが分かり、提案アプローチの有効性が示された。また、ユーザは歌声や歌詞、動画を高く評価する語や音楽動画の特定のシーンに関連する語を含むコメントに興味を持つ傾向にあることが明らかになったと同時に、同一のコメントセットを提示されたユーザが選択したコメントはユーザ間で異なり、多様であることも明らかになった。

本稿で用いた実験用システムではコメントフィールドに表示するコメントをランダムに選択しているが、個々のユーザにとって興味を引くコメントの性質をより明確にすることで、そうしたコメントを推定可能な手法を提案し、それぞれのユーザに合った興味を引くコメントを高い確率で表示できるような音楽動画探索システムを実現していく予定である。また本稿では、普段視聴しないタイプの音楽動画を判定する際にジャンルに着目したが、今後はクリエイターや曲調、テンポなどに着目し、多様な観点から普段聴かない楽曲との出会いを支援可能であるかも検証したい。

参考文献

[1] Kageyama, T., Mochizuki, K. and Takashima, Y.: Melody Retrieval with Humming, *Proc. 1993 International Computer Music Conference, ICMC '93*, pp.349-351 (1993).
 [2] Ghias, A., Logan, J., Chamberlin, D. and Smith, B.C.: Query by Humming: Musical Information Retrieval in an Audio Database, *Proc. 3rd ACM International Conference on Multimedia, MM '95*, pp.231-236 (1995).
 [3] Wang, C.-C., Jang, J.-S.R. and Wang, W.: An Improved Query by Singing/Humming System Using Melody and Lyrics Information, *Proc. 11th International Society for Music Information Retrieval Conference, ISMIR '10*, pp.45-50 (2010).

- [4] Tsai, W.-H., Yu, H.-M. and Wang, H.-M.: A Query-By-Example Technique for Retrieving Cover Versions of Popular Songs with Similar Melodies, *Proc. 6th International Conference on Music Information Retrieval, ISMIR '05*, pp.183–190 (2005).
- [5] Itoyama, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H.G.: Instrument Equalizer for Query-by-Example Retrieval: Improving Sound Source Separation Based on Integrated Harmonic and Inharmonic Models, *Proc. 9th International Conference on Music Information Retrieval, ISMIR '08*, pp.133–138 (2008).
- [6] Casey, M., Veltkamp, R., Goto, M., Leman, M., Rhodes, C. and Slaney, M.: Content-Based Music Information Retrieval: Current Directions and Future Challenges, *Proc. IEEE*, Vol.96, No.4, pp.668–696 (2008).
- [7] Shang, L., Zhang, D., Shen, J., Marmion, E.L. and Wang, D.: CCMR: A Classic-Enriched Connotation-Aware Music Retrieval System on Social Media with Visual Inputs, *Social Network Analysis and Mining*, Vol.11, pp.1–14 (2021).
- [8] Park, J., Shin, H., Oh, C. and Kim, H.Y.: “Is Text-Based Music Search Enough to Satisfy Your Needs?” A New Way to Discover Music with Images, *Proc. 2024 CHI Conference on Human Factors in Computing Systems, CHI '24*, pp.1–21 (2024).
- [9] Levy, M. and Sandler, M.: Music Information Retrieval Using Social Tags and Audio, *IEEE Trans. Multimedia*, Vol.11, No.3, pp.383–395 (2009).
- [10] Kamalzadeh, M., Kralj, C., Möller, T. and Sedlmair, M.: TagFlip: Active Mobile Music Discovery with Social Tags, *Proc. 21st International Conference on Intelligent User Interfaces, IUI '16*, pp.19–30 (2016).
- [11] Knees, P., Schedl, M. and Goto, M.: Intelligent User Interfaces for Music Discovery, *Trans. International Society for Music Information Retrieval*, Vol.3, No.1, pp.165–179 (2020).
- [12] 佃 洗撰, 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*, Vol.52, No.12, pp.3471–3482 (2011).
- [13] Wang, J.-C., Shih, Y.-C., Wu, M.-S., Wang, H.-M. and Jeng, S.-K.: Colorizing Tags in Tag Cloud: A Novel Query-by-Tag Music Search System, *Proc. 19th ACM International Conference on Multimedia, MM '11*, pp.293–302 (2011).
- [14] Baur, D., Steinmayr, B. and Butz, A.: SongWords: Exploring Music Collections Through Lyrics, *Proc. 11th International Society for Music Information Retrieval Conference, ISMIR '10*, pp.531–536 (2010).
- [15] Tsukuda, K., Ishida, K. and Goto, M.: Lyric Jumper: A Lyrics-Based Music Exploratory Web Service by Modeling Lyrics Generative Process, *Proc. 18th International Society for Music Information Retrieval Conference, ISMIR '17*, pp.544–551 (2017).
- [16] Stober, S. and Nürnberg, A.: MusicGalaxy: A Multi-focus Zoomable Interface for Multi-facet Exploration of Music Collections, *Proc. 7th International Conference on Exploring Music Contents, CMMR '10*, pp.273–302 (2010).
- [17] 平良木智悠, 山内正人, 砂原秀樹: 音楽に馴染みのない人を対象としたジャケットを活かした音楽発見サポートシステムの提案, *マルチメディア, 分散協調とモバイルシンポジウム 2017 論文集*, Vol.2017, pp.1477–1483 (2017).
- [18] Hong, J., Deng, H. and Yan, Q.: Tag-Based Artist Similarity and Genre Classification, *Proc. 2008 IEEE International Symposium on Knowledge Acquisition and Modeling Workshop, KAM '08*, pp.628–631 (2008).
- [19] Tsaptsinos, A.: Lyrics-Based Music Genre Classification Using a Hierarchical Attention Network, *Proc. 18th International Society for Music Information Retrieval Conference, ISMIR '17*, pp.694–701 (2017).
- [20] Oramas, S., Barbieri, F., Nieto, O. and Serra, X.: Multimodal Deep Learning for Music Genre Classification, *Trans. International Society for Music Information Retrieval*, Vol.1, No.1, pp.4–21 (2018).
- [21] Holm, J., Lehtiniemi, A. and Eronen, A.: Evaluating an Avatar-Based User Interface for Discovering New Music, *Proc. 9th International Conference on Mobile and Ubiquitous Multimedia, MUM '10*, pp.1–10 (2010).
- [22] Sergio, O., Luis, E.-A., Aonghus, L., Xavier, S. and Horacio, S.: Exploring Customer Reviews for Music Genre Classification and Evolutionary Studies, *Proc. 17th International Society for Music Information Retrieval Conference, ISMIR '16*, pp.150–156 (2016).
- [23] Yee, W.G., Yates, A., Liu, S. and Frieder, O.: Are Web User Comments Useful for Search?, *Proc. 7th Workshop on Large-scale Distributed Systems for Information Retrieval, LSDS-IR '09*, pp.63–70 (2009).
- [24] Chelaru, S.V., Orellana-Rodriguez, C. and Altingovde, I.S.: Can Social Features Help Learning to Rank YouTube Videos?, *Proc. 13th International Conference on Web Information Systems Engineering, WISE '12*, pp.552–566 (2012).
- [25] Hong, W., Jiang, S., Wang, H. and Shi, J.: Weighted-Based Summarization of Music Comments, *Proc. 8th International Conference on Computer Science & Education, ICCSE '13*, pp.140–143 (2013).
- [26] Liang, H., Liu, J., Xiang, Y., Du, J., Zhou, L., Pan, S. and Lei, W.: DiVa: An Iterative Framework to Harvest More Diverse and Valid Labels from User Comments for Music, *Proc. 31st ACM International Conference on Multimedia, MM '23*, pp.6223–6233 (2023).
- [27] 倉持友哉, 濱崎雅弘, 中野倫靖: MusicComment-Visualizer: 音楽動画へのコメントの可視化に基づく音楽推薦・鑑賞インタフェース, *コンピュータソフトウェア*, Vol.42, No.1, pp.97–114 (2025).
- [28] 中村聡史, 山本岳洋, 後藤真孝, 濱崎雅弘: 視聴者反応と音響特徴量に基づくサムネイル動画の生成手法, *情報処理学会論文誌データベース (TOD)*, Vol.6, No.3, pp.148–158 (2013).
- [29] Xian, Y., Li, J., Zhang, C. and Liao, Z.: Video Highlight Shot Extraction with Time-Sync Comment, *Proc. 7th International Workshop on Hot Topics in Planet-Scale MOBILE Computing and Online Social Networking, HOTPOST '15*, pp.31–36 (2015).
- [30] 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, *情報処理学会誌*, Vol.53, No.5, pp.466–471 (2012).
- [31] Nakamura, S. and Tanaka, K.: Video Search by Impression Extracted from Social Annotation, *Proc. 10th International Conference on Web Information Systems Engineering, WISE '09*, pp.401–414 (2009).
- [32] Tsukuda, K., Masahiro, H. and Goto, M.: SmartVideo-Ranking: Video Search by Mining Emotions from Time-Synchronized Comments, *Proc. 2016 IEEE 16th International Conference on Data Mining Workshops, ICDMW '16*, pp.960–969 (2016).
- [33] Yamamoto, T. and Nakamura, S.: Leveraging Viewer Comments for Mood Classification of Music Video Clips, *Proc. 36th International ACM SIGIR Conference on*

Research and Development in Information Retrieval, SIGIR '13, pp.797-800 (2013).

- [34] Goto, M.: SmartMusicKIOSK: Music Listening Station with Chorus-Search Function, *Proc. 16th Annual ACM Symposium on User Interface Software and Technology, UIST '03*, pp.31-40 (2003).
- [35] Kenmochi, H. and Ohshita, H.: VOCALOID - Commercial Singing Synthesizer Based on Sample Concatenation, *Proc. 8th Annual Conference of the International Speech Communication Association, INTERSPEECH '07*, pp.4009-4010 (2007).
- [36] Hamasaki, M. and Goto, M.: Songrium: A Music Browsing Assistance Service Based on Visualization of Massive Open Collaboration Within Music Content Creation Community, *Proc. 9th International Symposium on Open Collaboration, WikiSym '13*, pp.1-10 (2013).
- [37] 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, 情報処理学会研究報告, Vol.2008, No.50, pp.37-42 (2008).
- [38] 土方嘉徳, 早川卓弥: 実時間と動画時間から抽出した面白い動画コメント 100 連発, Web インテリジェンスとインタラクション研究会 予稿集, Vol.6, pp.37-40 (2015).
- [39] Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Music Genre Database and Musical Instrument Sound Database, *Proc. 4th International Conference on Music Information Retrieval, ISMIR '03*, pp.229-230 (2003).
- [40] Kudo, T., Yamamoto, K. and Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis, *Proc. 2004 Conference on Empirical Methods in Natural Language Processing, EMNLP '04*, pp.230-237 (2004).
- [41] 竹川和毅, 土方嘉徳, 西田正吾: 内容に基づく音楽データの探索・推薦システム, 人工知能学会論文誌, Vol.23, No.5, pp.330-343 (2008).
- [42] Krippendorff, K.: *Computing Krippendorff's Alpha-Reliability*, Technical Report, University of Pennsylvania (2011).
- [43] Castro, S.: Fast Krippendorff: Fast Computation of Krippendorff's Alpha Agreement Measure, available from <https://github.com/pln-fing-udelar/fast-krippendorff> (2017).



木下 裕一朗 (学生会員)

2001 年生。2024 年明治大学総合数理学部先端メディアサイエンス学科卒業。現在、同大学大学院先端数理科学研究科博士前期課程在学中。音楽動画探索や、画像によるスポーツのネタバレ防止等の研究に従事。学士(理学)。



佃 洸撰 (正会員)

2014 年京都大学大学院情報学研究科博士後期課程修了。博士(情報学)。現在、産業技術総合研究所主任研究員。主に音楽コンテンツおよびユーザ生成コンテンツを対象とした情報推薦・情報探索の研究に従事。日本データベース学会会員。



渡邊 研斗 (正会員)

2018 年東北大学大学院情報科学研究科博士後期課程修了。博士(情報科学)。現在、産業技術総合研究所主任研究員。自然言語処理、音楽情報処理、歌詞情報処理の研究に従事。言語処理学会会員。



中塚 貴之 (正会員)

2021 年早稲田大学大学院先進理工学研究科博士後期課程修了。博士(工学)。現在、産業技術総合研究所研究員。主にメディアコンテンツを対象とした情報検索・創作支援のためのマルチモーダル情報処理研究に従事。



中野 倫靖 (正会員)

2008 年筑波大学大学院図書館情報メディア研究科博士後期課程修了。博士(情報学)。現在、産業技術総合研究所研究グループ長。日本音響学会会員。2009 年情報処理学会山下記念研究賞(音楽情報科学研究会)、2013 年 Sound and Music Computing Conference (SMC 2013) The Best Paper Award, 2018 年 IEEE Pacific Visualization Symposium (PacificVis 2018) Honorable Mention Poster Award 等受賞。



後藤 真孝 (正会員)

1998年早稲田大学大学院理工学研究科博士後期課程修了。博士(工学)。現在、産業技術総合研究所上級首席研究員。2009~2017年にIPA未踏IT人材発掘・育成事業PM, 2016~2022年にJST ACT-I「情報と未来」研究総括を兼任。現在、JST創発PO, 早稲田大学客員教授, 統計数理研究所客員教授, 筑波大学連携大学院教授等を兼任。日本学士院学術奨励賞, 日本学術振興会賞, ドコモ・モバイル・サイエンス賞基礎科学部門優秀賞, 市村学術賞, FIT船井業績賞等, 77件受賞。本会フェロー。



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先端メディアサイエンス学科准教授, 2018年同教授, 現在に至る。サーチとインタラクションや, ネットバレ防止技術, 平均手書き文字等の研究活動に従事。博士(工学)。