

動画共有サービス上の興味を引く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 佃 洗撰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

概要: 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、自身が普段は聴かないクリエイターやジャンルの音楽動画の中から、キーワードやタグに基づく従来の検索を用いて、新たに興味をもって聴いてみたいと思える音楽動画を見つけるのは困難である。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、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を視聴してみたいと思うものが存在すると仮定し、そうした興味を引く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。本稿では、提示された時刻同期コメントのうち、興味を引くコメントを選ぶことでそれが投稿された音楽動画を探索できるシステムを実装した。また、選択されたコメントや視聴された音楽動画を分析する評価実験を実施して、提案するアプローチの有用性を検証した。

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] や音楽コンテンツ [3][4][5] をクエリとした検索や、ソーシャルタグ [6][7] を利用した検索手法が提案されてきた [8]。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言えない。

情報検索においては、意図しているものの検索だけでなく、予期せずにも価値あるものや好ましいものを発見すること（セレンディピティ [9]）も重視されている。セレンディピティを促進するシステムデザインはユーザ体験を向上させる可能性が示されており [10]、セレンディピティを考慮した音楽検索モデル [11] や、セレンディピティを誘発する音楽探索・推薦手法が提案されている [12][13]。音楽聴取においてセレンディピティの要素を取り入れることは、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

げたりというように、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、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 [14]。そのため、音楽動画であれば、「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、そのコメントが投稿された音楽動画の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、楽曲動画を視聴するきっかけを作れるという利点もある。

以上の点を踏まえて本研究では、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。本稿では、時刻同期コメントを提示し、ユーザが興味をもったコメントを選ぶとそのコメ

¹ 明治大学
Meiji University

² 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

ントが投稿された動画内の再生時刻から音楽動画を探索できるシステムを実装した。システムの利用実験を通して、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。

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 [7][15]、歌詞 [16][17]、ジャケット画像 [18][19] などに基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。例えば TagFlip [7] では、ユーザが選択したタグの組み合わせに応じて楽曲の再生リストが生成される。Lyric Jumper [17] 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 [18] 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くことができる。

タグ、歌詞、ジャケット画像のいずれのメタ情報も、楽曲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 [20][21][22]。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。本研究では、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] などが取り組まれてきた。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 [14] 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、視聴者が動画に抱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もつ、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アプローチを提案する。具体的には、時刻同期コメントをユーザに提示し、提示したコメントが選択された際に、そのコメントが投稿された動画内の時刻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から視聴できるようにすることで、ユーザは音楽動画を最初から最後まで視聴しなくとも、その音楽動画が自身の嗜好に合ったものか判断できるようになる。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音楽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. 実験

本研究では、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。本章では、まず使用するデータを説明し、次に実験用システムの詳細を述べ、最後に実験手順について説明する。

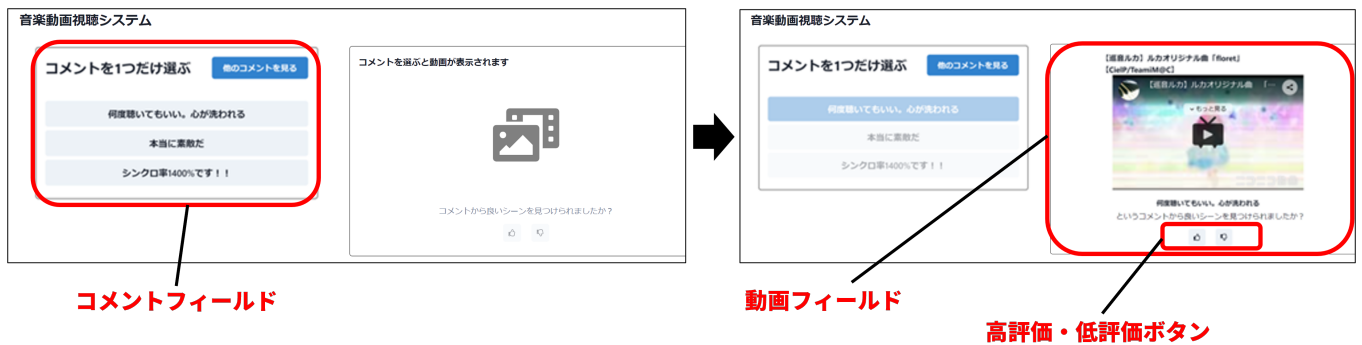


図 1 表示された 3 つのコメントから興味を引かれたコメントを選択すると、そのコメントが投稿されたシーンから音楽動画を探索できる実験用システム。

4.1 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。

4.2 実験用システム

前節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成り、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、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が表示される。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。

本実験では、表示されたコメントすべてをクリックすることで、実際に視聴したいと思うようなコメントの特徴を分析できなくなる可能性を考慮し、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。

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

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

*3 <https://songrium.jp>

4.3 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 1 日最低 30 回評価した。3 件のコメントから 1 件を選んだ場合だけでなく、コメントを選ばずに「他のコメントを見る」のボタンを押して新しくコメントを表示した場合も評価回数に含めた。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をした。

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。

5. 結果

実験参加者は、筆頭著者が所属する研究室の学生 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等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、本章ではこの 85 件を除いた 4,395 件の選択コメントを分析に使用する。

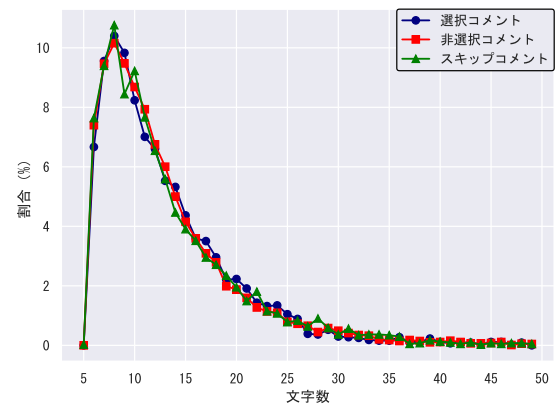


図 2 選択・非選択・スキップコメントの文字数の分布。

5.1 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、大きな違いは見られなかった（例えば、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。そのために、3 種類のそれぞれから 4,000 件のコメントをランダムに選択し、MeCab [38] 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、各形態素が含まれるコメントを数え、出現頻度の上位 20 の形態素を求めた結果を図 3 に示す。選択・非選択・スキップのいずれにおいても「曲」「いい」「好き」の出現頻度が高かったが、特に選択コメントでは他の 2 種類のコメントよりもこれらの出現頻度が高かった。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いコメントであると言える。また、選択コメントにのみ「最高」「神」といった語が上位に出現し、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わかる。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、上記の結果とも一致していた。一方で、「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も得られており、例えば歌声合成パラメータ調整に関連した「調教」を含むスキップコメントはこれに該当すると考えられる。

出現頻度が特に高かった「曲」「いい」「好き」に着目し、

*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 [37]」の「大分類」を採用した。

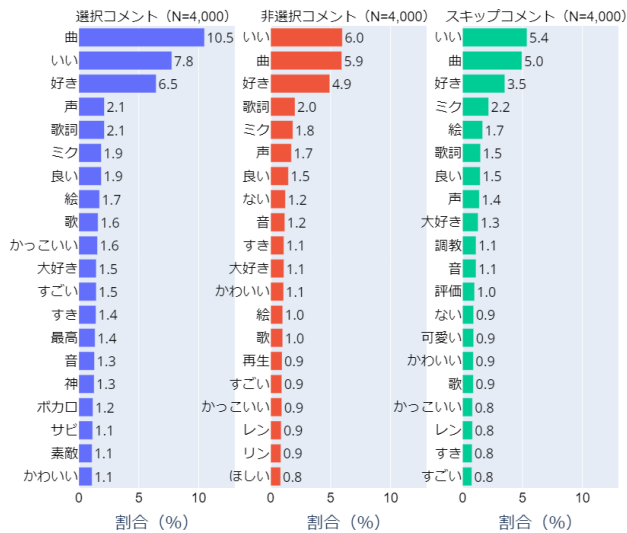


図 3 コメント内の出現割合上位 20 の形態素。

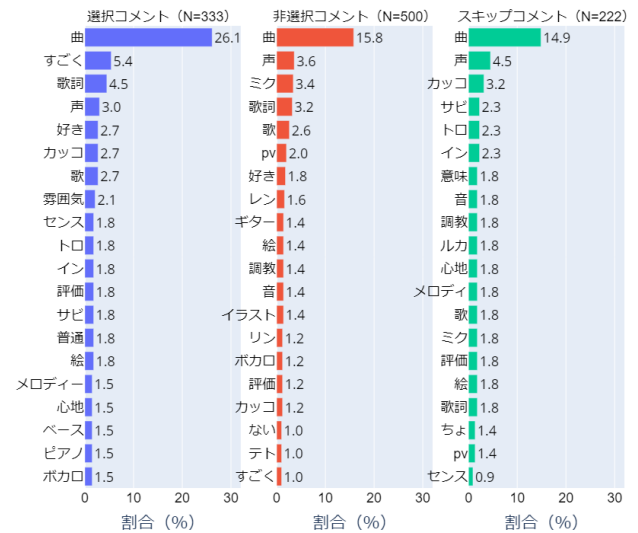


図 5 形態素「いい」を含むコメント内の出現割合上位 20 の形態素。

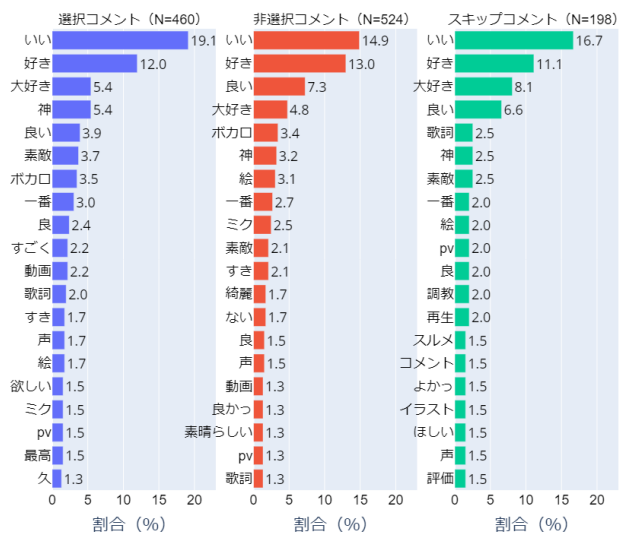


図 4 形態素「曲」を含むコメント内の出現割合上位 20 の形態素。

コメントに含まれる語をより詳細に分析した。例えば「曲」であれば、選択・非選択・スキップの各コメント集合において「曲」の形態素を含むコメント内での出現頻度が上位 20 の形態素を求めた。「曲」「いい」「好き」のそれぞれに対する結果を図 4 から図 6 に示す。図 4 の「曲」の選択コメントを見ると、「神」「最高」といった語の出現確率が非選択・スキップのコメントよりも高く、楽曲自体あるいは楽曲の何らかの側面を高く評価するコメントが選択される傾向にあった。図 5 の「いい」の結果を見ると、「歌詞」や「声」など、「いい」と言っている対象となる語の出現確率が高く、特に選択コメントでは「雰囲気」「センス」といった抽象的な側面や「ベース」「ピアノ」などの楽器に関する語の出現確率が高かった。これらは、アンケートで得られた「抽象的なコメント」や「楽器名が書いてあるコメント」が表示されたときに音楽動画を視聴したいと思ったという回答と一致していた。図 6 の「好き」では、好きな対象を表す語と、好きな程度を表す語の両方が上位に含ま

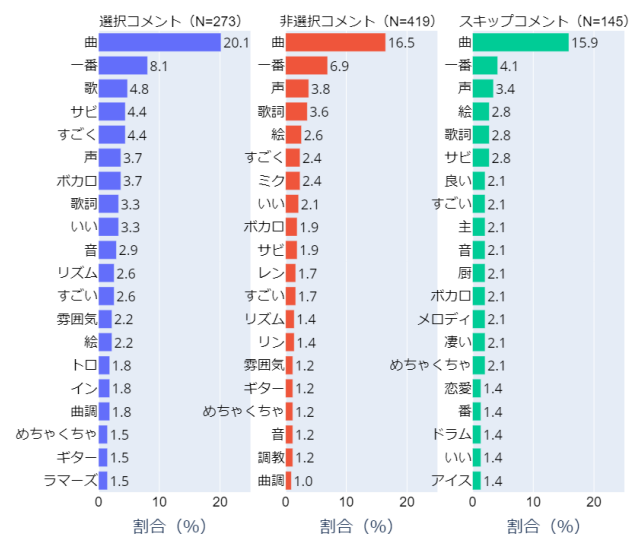


図 6 形態素「好き」を含むコメント内の出現割合上位 20 の形態素。

れていた。選択コメントでは好きな対象として「サビ」や「イントロ」（「イン」および「トロ」といった楽曲のパートに関する語の出現確率が高いという特徴があった。

以上のように、選択コメントにはある程度の傾向は見られたものの、全体的には選択・非選択・スキップで同じ語が上位 20 件に含まれることが多かった。全く同じコメントが選択コメントとスキップコメントの両方に含まれている事例も複数あったことから、興味を引かれるコメントには個人差があることが原因の一つであると考えられる。そのため今後は、システムのコメントフィールドに、複数の実験参加者間で共通の 3 つのコメントを表示することで、選ばれるコメントの一致度合いを検証するなど、より詳細な実験を行う予定である。また、本稿では選択コメントの一般的な傾向を調べるための分析を行ったが、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

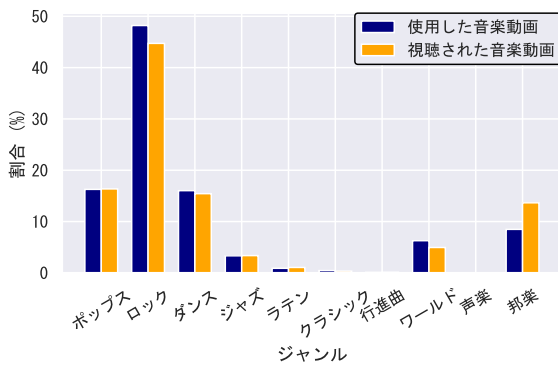


図 7 実験に使用した音楽動画全体の音楽ジャンルの分布と、実験で視聴された音楽動画の音楽ジャンルの分布。

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のみ当てはまるコメントも含まれていた。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。

5.2 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、歌声合成楽曲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、4.3節で述べた10種類の各ジャンルに対応するタグを人手で抽出し、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、実験で使用した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7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概ね一致していた。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8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数は6.4であった。本実験で使用した音楽動画の多くはジャンルがロックであったため(図7)、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みなした。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 $V_{unfam+fav}^u$ とする。 V^u に対する $V_{unfam+fav}^u$ 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げられる確率を表す。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ため、きっかけとして十分に機能していると言える。

さらに、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関わらず、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言える。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。

5.3 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ぬ負担があった。そのため、このシステムのようにコメントから選ぶと全て視聴しなければいけな

*5 <https://kiite.jp>

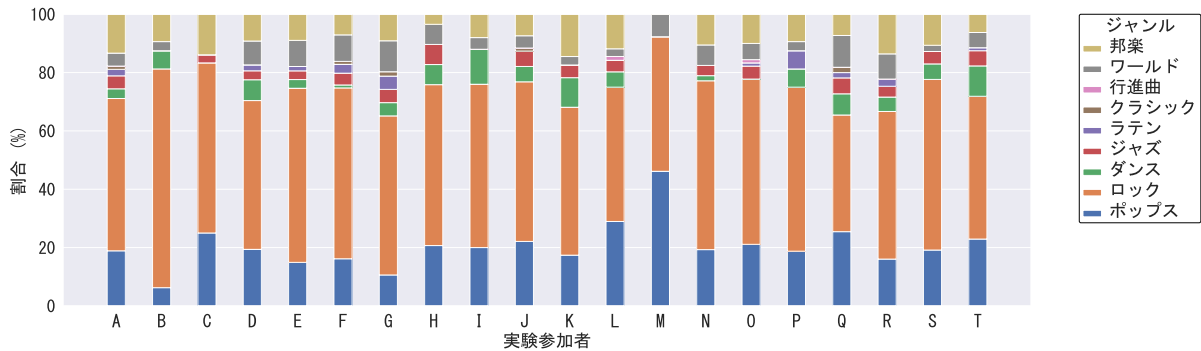


図 8 各実験参加者が視聴した音楽動画のジャンルの内訳。

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

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

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

い気持ちが無くなり、動画選択の負荷が軽くなった。」「コメントから動画の内容をある程度予想しながら使っていたが、予想外のコンテンツに出会うことが多く、良い意味で期待を裏切られるのが面白かった。」「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断するのが難しい。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。

アンケートで得られた実験用システムの改善点をふまえて、今後は自身が閲覧したコメントの振り返り機能や高評価したコメントの保存機能を実装する予定である。この機能を実装することで、ユーザはコメントが投稿されたシーンのみを聴いて良いと思った場合に高評価して保存し、時間があるときにまとめて高評価したコメントの音楽動画を楽しむという使い方が可能になる。また、本稿はニコニコ動画に投稿された VOCALOID 楽曲を対象に実験を行ったが、今後は様々な音楽動画も使用して提案するアプローチの有用性を検証したい。

6. おわりに

本稿では、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うものを発見できるようにするために、動画に投稿された時刻同期コメントの中で興味を引くコメントをきっかけとして用いるアプローチを提案した。実験を通して、楽曲自体や楽曲のある側面に対して高い評価を述べているコメントが興味を引く傾向にあることが明らかになった一方で、人によって好みのコメントは異なり、さらなる詳細な調査が必要であることも示された。ただし、興味を引かれたコメントを選択することで、普段聴かないジャンルでかつ自身の嗜好に合う音楽動画を高い確率で見つけられており、提案するアプローチの有効性は示された。

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、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。また、現状の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、人の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、人の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する予定である。

参考文献

[1] Kageyama, T., Mochizuki, K. and Takashima, Y.: Melody Retrieval with Humming, *Proc. of 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. of MM '95*, pp. 231–236 (1995).
- [3] 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. of ISMIR '05*, pp. 183–190 (2005).
- [4] 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. of ISMIR '08*, pp. 133–138 (2008).
- [5] Casey, M., Veltkamp, R., Goto, M., Leman, M., Rhodes, C. and Slaney, M.: Content-Based Music Information Retrieval: Current Directions and Future Challenges, *Proceedings of the IEEE*, Vol. 96, No. 4, pp. 668–696 (2008).
- [6] Levy, M. and Sandler, M.: Music Information Retrieval Using Social Tags and Audio, *IEEE Transactions on Multimedia*, Vol. 11, No. 3, pp. 383–395 (2009).
- [7] Kamalzadeh, M., Kralj, C., Möller, T. and Sedlmair, M.: TagFlip: Active Mobile Music Discovery with Social Tags, *Proc. of IUI '16*, pp. 19–30 (2016).
- [8] Knees, P., Schedl, M. and Goto, M.: Intelligent User Interfaces for Music Discovery, *Transactions of the International Society for Music Information Retrieval*, Vol. 3, No. 1, pp. 165–179 (2020).
- [9] Rosenau, M. J.: Serendipity, *Journal of Bacteriology*, Vol. 29, No. 2, pp. 91–98 (1935).
- [10] Sun, T., Zhang, M. and Mei, Q.: Unexpected Relevance: An Empirical Study of Serendipity in Retweets, *Proc. of ICWSM '21*, pp. 592–601 (2021).
- [11] Schedl, M., Hauger, D. and Schnitzer, D.: A Model for Serendipitous Music Retrieval, *Proc. of CaRR '12*, pp. 10–13 (2012).
- [12] Taramigkou, M., Bothos, E., Christidis, K., Apostolou, D. and Mentzas, G.: Escape the Bubble: Guided Exploration of Music Preferences for Serendipity and Novelty, *Proc. of RecSys '13*, pp. 335–338 (2013).
- [13] Saito, T., Sato-Shimokawara, E. and Chen, L.-H.: Fusion-Based Music Recommender System Using Music Affective Space Based on Serendipity, *Proc. of TAAI '22*, pp. 171–176 (2022).
- [14] 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).
- [15] 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. of MM '11*, pp. 293–302 (2011).
- [16] Baur, D., Steinmayr, B. and Butz, A.: SongWords: Exploring Music Collections Through Lyrics, *Proc. of ISMIR '10*, pp. 531–536 (2010).
- [17] Tsukuda, K., Ishida, K. and Goto, M.: Lyric Jumper: A Lyrics-Based Music Exploratory Web Service by Modeling Lyrics Generative Process, *Proc. of ISMIR '17*, pp. 544–551 (2017).
- [18] Stober, S. and Nürnbergger, A.: MusicGalaxy: A Multi-focus Zoomable Interface for Multi-facet Exploration of Music Collections, *Proc. of CMMR '10*, pp. 273–302 (2010).
- [19] 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).
- [20] Hong, J., Deng, H. and Yan, Q.: Tag-Based Artist Similarity and Genre Classification, *Proc. of KAM '08*, pp. 628–631 (2008).
- [21] Tsaptsinos, A.: Lyrics-Based Music Genre Classification Using a Hierarchical Attention Network, *Proc. of ISMIR '17*, pp. 694–701 (2017).
- [22] Oramas, S., Barbieri, F., Nieto, O. and Serra, X.: Multimodal Deep Learning for Music Genre Classification, *Transactions of the International Society for Music Information Retrieval*, Vol. 1, No. 1, pp. 4–21 (2018).
- [23] Yee, W. G., Yates, A., Liu, S. and Frieder, O.: Are Web User Comments Useful for Search?, *Proc. of LSDS-IR '09* (2009).
- [24] Chelaru, S. V., Orellana-Rodriguez, C. and Altingovde, I. S.: Can Social Features Help Learning to Rank YouTube Videos?, *Proc. of WISE '12*, pp. 552–566 (2012).
- [25] Hong, W., Jiang, S., Wang, H. and Shi, J.: Weighted-Based Summarization of Music Comments, *Proc. of 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. of MM '23*, pp. 6223–6233 (2023).
- [27] 倉持友哉, 濱崎雅弘, 中野倫靖: MusicCommentVisualizer: 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, *Proc. of WISS '23* (2023).
- [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. of 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 (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. of 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. of ICDMW '16*, pp. 960–969 (2016).
- [33] Yamamoto, T. and Nakamura, S.: Leveraging Viewer Comments for Mood Classification of Music Video Clips, *Proc. of SIGIR '13*, pp. 797–800 (2013).
- [34] Goto, M.: SmartMusicKIOSK: Music Listening Station with Chorus-Search Function, *Proc. of UIST '03*, pp. 31–40 (2003).
- [35] Kenmochi, H. and Ohshita, H.: VOCALOID - Commercial Singing Synthesizer Based on Sample Concatenation, *Proc. of 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. of WikiSym '13*, pp. 1–10 (2013).
- [37] Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Music Genre Database and Musical Instrument Sound Database, *Proc. of ISMIR '03*, pp. 229–230 (2003).
- [38] Kudo, T., Yamamoto, K. and Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis, *Proc. of EMNLP '04*, pp. 230–237 (2004).