

動画共有サービス上の時刻同期コメントを視聴のきっかけに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 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: 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、ユーザが興味をも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から音楽動画を視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高い確率で出会えることや、ユーザが興味をも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。

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言えない。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、「このギターカッコいい」「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。時刻同期コメントは気軽に沢山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利点がある。さらに、そうした気軽さに加えて、時刻同期コメントは同じユーザが一つの動画に複数回

¹ 明治大学

² 産業技術総合研究所

^{a)} kinoshita@nkmr-lab.org

投稿することも多く、それらがある種のライフログのように蓄積されているため、そうした膨大な時刻同期コメントを活用できる。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が期待できる。

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、ユーザが興味をも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量的および定性的に示した。
- 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当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。例えば TagFlip [10] 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 [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してしまうため、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。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 [22][23]、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 [24][25]、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 [26] 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 [27][28] 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 [29]。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 [30]、視聴者の反応に基づく音楽動画ランキングの生成 [31]、音楽動画の印象の分類 [32] 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、その一つが選択された際に、それ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 [33] 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 [34] 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 [35] 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。

4.1.2 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、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

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

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

*3 <https://songrium.jp>

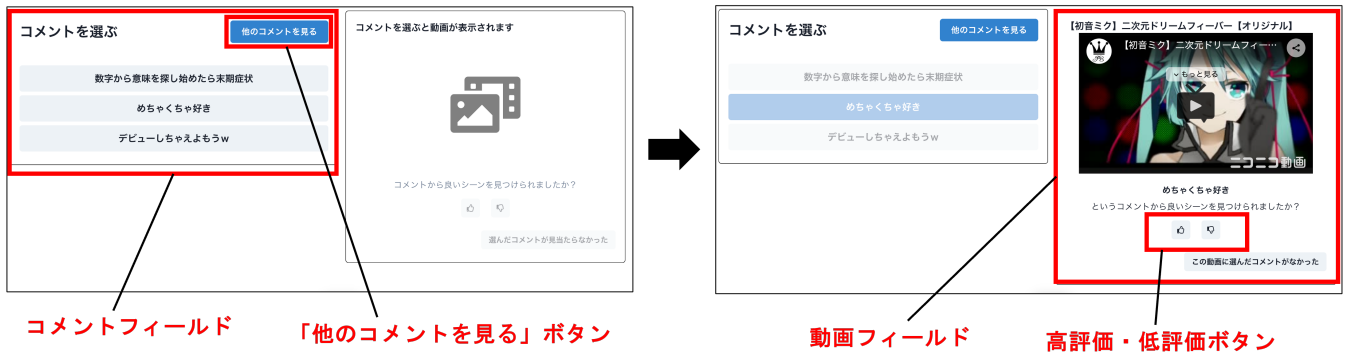


図 1 表示された3つのコメントから興味を引かれたコメントを選択すると、そのコメントが投稿されたシーンから音楽動画を視聴できる実験用システム。

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。

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。3件の中に興味を引くコメントが一つもなかった場合は、実験参加者は「他のコメントを見る」のボタンを押す、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。

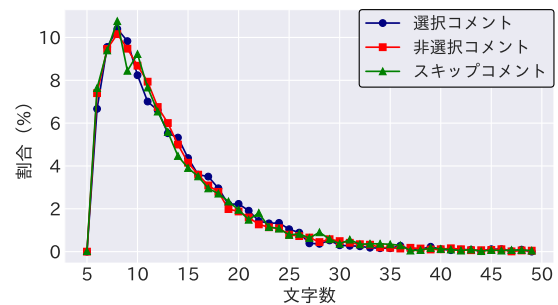


図 2 選択・非選択・スキップコメントの文字数の分布。

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。

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

*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 [36]」の「大分類」を採用した。

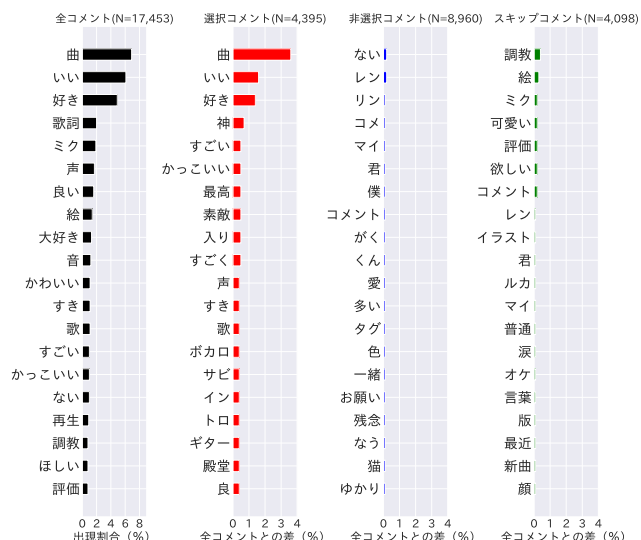


図 3 全コメントにおいて出現割合上位 20 の形態素と、選択・非選択・スキップのそれぞれのコメントにおいて各形態素が含まれる割合から、全コメントにおいて同じ形態素が含まれる割合を引いた値が上位 20 の形態素。

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ために、MeCab [37] 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の含まれる割合を引いたときの値を求めた。例えば、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言える。

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。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く、選ばれる回数が多かったと言える。また、選択コメントでは「神」

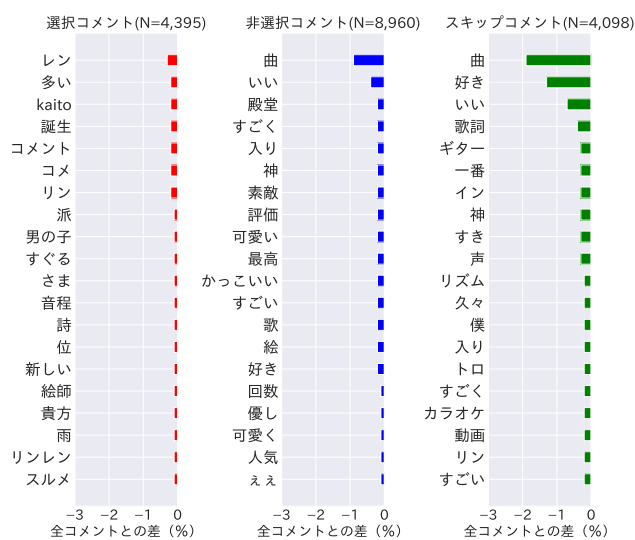


図 4 選択・非選択・スキップのそれぞれのコメントにおいて各形態素が含まれる割合から、全コメントにおいて同じ形態素が含まれる割合を引いたときの値が下位 20 の形態素。

「すごい」「最高」の値も高く、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、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」の値が高かった。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、例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づらい形態素であると言える。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」の値が低く、そのような語を含むコメントは選ばれづらかったことがわかる。これは、図 3 に示したように、非選択・スキップコメントにおいてバーチャルシンガーの名前を指す語の出現割合が他と比べて高かったことと一致している。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」が含まれていた。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」「ギター」「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。これらの語の値が小さいということは、そのような語を含むコメントは選ばれやすかったと言える。これについても、図3に示したように、楽曲に対して言及しているコメントや特に高く評価するコメントは選ばれやすかったという結果と一致していた。

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当てはまるコメントも含まれていることがわかった。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。

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節で述べた10種類の各ジャンルに対応するタグを手で抽出し、「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め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

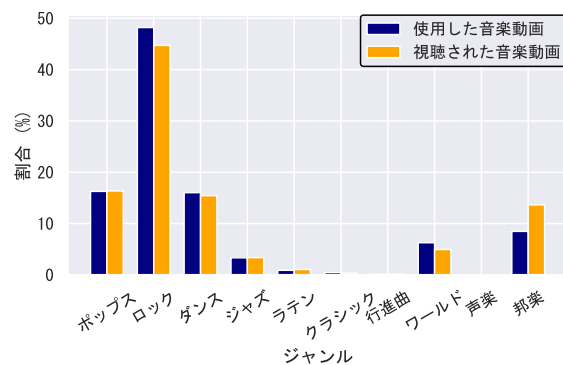


図5 実験に使用した音楽動画全体の音楽ジャンルの分布と、実験で視聴された音楽動画の音楽ジャンルの分布。

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けられる確率を表す。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が

*5 <https://kiite.jp>

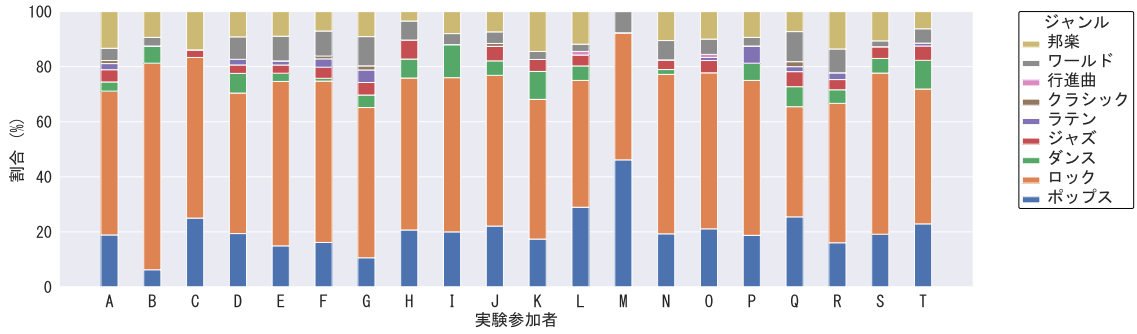


図 6 各実験参加者が視聴した音楽動画のジャンルの内訳.

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	0.0633	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

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。提案アプローチの有用性が実験を通して示されたので、今後は上記の「高評価したコメントや動画の保存機能」を始めとする、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な Web サービスとして公開したいと考えている。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。

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 節の分析により、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、我々は分析を行う中で、同じような内容のコメントであっても、選択・非選択・スキップのうちの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き、興味を引くコメントが人々の間でどの程度一致するかという新たな疑問を抱いた。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 4.2.2 節の定量評価と 4.2.3 節のアンケート調査で示されたので、

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節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節と同様、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、興味をもったコメントを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節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。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つのうちの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 [38]を用いて求めた。この指標は-1から1の間の値を取り、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0.347、Bグループでは0.355であった*6。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人ずつ抽出し、すべての組み合わせに対して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みなされるため、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して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

*6 Krippendorff's alpha の計算には Castro により公開されているライブラリ [39] を使用した。

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.

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

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. 例えば, 「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る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る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 の形態素を図 7 に示す. 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かった. 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. これらのことから, 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は, より多くの人の興味を引くと言える.

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

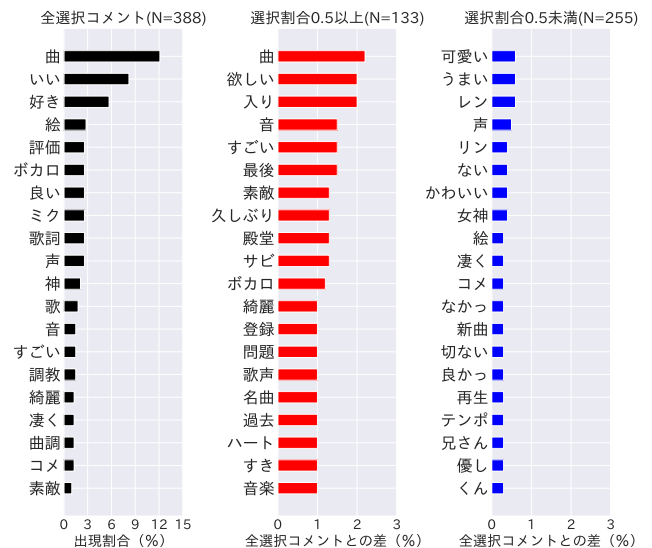


図 7 選択コメントを選んだ人数の割合が 0.5 以上であったコメントと, 0.5 未満であったコメントそれぞれにおいて各形態素が含まれる割合から, すべての選択コメントにおいて同じ形態素が含まれる割合を引いた値が上位 20 の形態素.

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踏まえて, ユーザそれぞれの興味を引かれるコメントのタイプを特定して, それに応じたコメントを推薦するなどして, 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目指す.

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もった時刻同期コメントをきっかけとしてそれが投稿されたシーンから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, ユーザが興味をも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高い確率で出会えることがわかり, 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も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、今後はクリエータや曲調、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。

謝辞 本研究の一部は JSPS 科研費 JP22K12135 の助成を受けたものです。

参考文献

- [1] Kageyama, T. et al.: Melody Retrieval with Humming, *ICMC '93*, pp. 349–351 (1993).
- [2] Ghias, A. et al.: Query by Humming: Musical Information Retrieval in an Audio Database, *MM '95*, pp. 231–236 (1995).
- [3] Wang, C.-C. et al.: An Improved Query by Singing/Humming System Using Melody and Lyrics Information, *ISMIR '10*, pp. 45–50 (2010).
- [4] Tsai, W.-H. et al.: A Query-By-Example Technique for Retrieving Cover Versions of Popular Songs with Similar Melodies, *ISMIR '05*, pp. 183–190 (2005).
- [5] Itoyama, K. et al.: Instrument Equalizer for Query-by-Example Retrieval: Improving Sound Source Separation Based on Integrated Harmonic and Inharmonic Models, *ISMIR '08*, pp. 133–138 (2008).
- [6] Casey, M. et al.: Content-Based Music Information Retrieval: Current Directions and Future Challenges, *Proceedings of the IEEE*, Vol. 96, No. 4, pp. 668–696 (2008).
- [7] Shang, L. et al.: 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. et al.: “Is Text-Based Music Search Enough to Satisfy Your Needs?” A New Way to Discover Music with Images, *CHI '24*, pp. 1–21 (2024).
- [9] 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).
- [10] Kamalzadeh, M. et al.: TagFlip: Active Mobile Music Discovery with Social Tags, *IUI '16*, pp. 19–30 (2016).
- [11] Knees, P. et al.: Intelligent User Interfaces for Music Discovery, *Transactions of the International Society for Music Information Retrieval*, Vol. 3, No. 1, pp. 165–179 (2020).
- [12] 佃洗撰 他: 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. et al.: Colorizing Tags in Tag Cloud: A Novel Query-by-Tag Music Search System, *MM '11*, pp. 293–302 (2011).
- [14] Baur, D. et al.: SongWords: Exploring Music Collections Through Lyrics, *ISMIR '10*, pp. 531–536 (2010).
- [15] Tsukuda, K. et al.: Lyric Jumper: A Lyrics-Based Music Exploratory Web Service by Modeling Lyrics Generative Process, *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, *CMMR '10*, pp. 273–302 (2010).
- [17] 平良木智悠 他: 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. et al.: Tag-Based Artist Similarity and Genre Classification, *KAM '08*, pp. 628–631 (2008).
- [19] Tsaptsinos, A.: Lyrics-Based Music Genre Classification Using a Hierarchical Attention Network, *ISMIR '17*, pp. 694–701 (2017).
- [20] Oramas, S. et al.: Multimodal Deep Learning for Music Genre Classification, *Transactions of the International Society for Music Information Retrieval*, Vol. 1, No. 1, pp. 4–21 (2018).
- [21] Holm, J. et al.: Evaluating an Avatar-Based User Interface for Discovering New Music, *MUM '10*, pp. 1–10 (2010).
- [22] Yee, W. G. et al.: Are Web User Comments Useful for Search?, *LSDS-IR '09* (2009).
- [23] Chelaru, S. V. et al.: Can Social Features Help Learning to Rank YouTube Videos?, *WISE '12*, pp. 552–566 (2012).
- [24] Hong, W. et al.: Weighted-Based Summarization of Music Comments, *ICCSE '13*, pp. 140–143 (2013).
- [25] Liang, H. et al.: DiVa: An Iterative Framework to Harvest More Diverse and Valid Labels from User Comments for Music, *MM '23*, pp. 6223–6233 (2023).
- [26] 倉持友哉 他: 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, *WISS '23* (2023).
- [27] 中村聡史 他: 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).
- [28] Xian, Y. et al.: Video Highlight Shot Extraction with Time-Sync Comment, *HOTPOST '15*, pp. 31–36 (2015).
- [29] 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).
- [30] Nakamura, S. and Tanaka, K.: Video Search by Impression Extracted from Social Annotation, *WISE '09*, pp. 401–414 (2009).
- [31] Tsukuda, K. et al.: SmartVideoRanking: Video Search by Mining Emotions from Time-Synchronized Comments, *ICDMW '16*, pp. 960–969 (2016).
- [32] Yamamoto, T. and Nakamura, S.: Leveraging Viewer Comments for Mood Classification of Music Video Clips, *SIGIR '13*, pp. 797–800 (2013).
- [33] Goto, M.: SmartMusicKIOSK: Music Listening Station with Chorus-Search Function, *UIST '03*, pp. 31–40 (2003).
- [34] Kenmochi, H. and Ohshita, H.: VOCALOID - Commercial Singing Synthesizer Based on Sample Concatenation, *INTERSPEECH '07*, pp. 4009–4010 (2007).
- [35] Hamasaki, M. and Goto, M.: Songrium: A Music Browsing Assistance Service Based on Visualization of Massive Open Collaboration Within Music Content Creation Community, *WikiSym '13*, pp. 1–10 (2013).
- [36] Goto, M. et al.: RWC Music Database: Music Genre Database and Musical Instrument Sound Database, *ISMIR '03*, pp. 229–230 (2003).
- [37] Kudo, T. et al.: Applying Conditional Random Fields to Japanese Morphological Analysis, *EMNLP '04*, pp. 230–237 (2004).
- [38] Krippendorff, K.: Computing Krippendorff’s Alpha-Reliability (2011).
- [39] Castro, S.: Fast Krippendorff: Fast Computation of Krippendorff’s Alpha Agreement Measure, <https://github.com/pln-fing-udelar/fast-krippendorff> (2017).