

歌声・伴奏・歌詞ごとの嗜好と許容度を考慮できる 楽曲推薦インタフェースの検討

中澤 一輝^{1,2,a)} 佃 洸² 中野 倫靖² 後藤 真孝²

概要: 楽曲に対する嗜好には、歌声・伴奏・歌詞といった音楽要素に対する嗜好の度合いがあると考えられる。しかし、楽曲の内容の類似度に基づく従来の楽曲推薦システムでは、ユーザは推薦された各楽曲が自分の嗜好に合うかどうかを楽曲単位で評価することが一般的であるため、音楽要素ごとのユーザの嗜好を十分に反映した推薦が難しい問題がある。また仮に、ある楽曲のある音楽要素に対するユーザの嗜好の度合いが高いことがわかったとしても、その音楽要素とどの程度まで類似した楽曲であればそのユーザに推薦すべきかわからない問題もある。本研究では、これら2つの問題を解決するための楽曲推薦インタフェースを提案する。具体的には、1つ目の問題を解決するために、歌声・伴奏・歌詞の3つの音楽要素を対象とし、各音楽要素に対してユーザが嗜好の度合いを評価可能にする。2つ目の問題を解決するために、各音楽要素に対して、ユーザがどの程度の類似度を嗜好に合う範囲として許容できるかの度合いを「許容度」として指定可能にする。これにより、許容度の低い音楽要素に対するユーザの嗜好を重視した楽曲推薦が可能になる。提案する楽曲推薦インタフェースを実装し、その有効性を検証するために評価実験を実施した。

1. はじめに

Spotify や Apple Music などの音楽ストリーミングサービスの普及で人々は膨大な数の楽曲に容易にアクセスできるようになった。しかし、膨大な数の楽曲からユーザが自分の嗜好に合う楽曲を探すことは困難である。この解決策の1つとして、ユーザの嗜好に合う楽曲を推薦する楽曲推薦システム [1], [2], [3] が活用されている。

楽曲推薦システムで主に用いられる方式として、協調フィルタリングと内容に基づくフィルタリングの2つが挙げられる [4]。協調フィルタリングは、他のユーザの聴取履歴などから得られる嗜好傾向を基に楽曲推薦を行う手法である [5], [6]。この手法を用いることで、楽曲の内容に関する特徴の解析を行うことなくユーザの嗜好に合う楽曲を精度よく推薦できる。しかし、新規ユーザや未評価の楽曲に対応しづらいコールドスタート問題 [7] や、人気のある楽曲へ推薦が偏ってしまう問題 [8] がある。一方、内容に基づくフィルタリングは、楽曲の音色などの音響特徴 [9] を統合したベクトル間の類似度に基づいて、ユーザが好きな楽曲に類似した楽曲を見つけることで、嗜好に合った楽曲を推薦できる手法 [10], [11], [12] であるため、再生数に依存しないことから、前述した問題は起きにくい。さらに、楽

曲の歌声 [13] や伴奏 [14], 歌詞 [15] といった楽曲を構成する要素に基づき推薦する手法も提案されている。本研究では、このような楽曲を構成する要素を「音楽要素 [16], [17]」と定義し、歌声・伴奏・歌詞の3つの要素を対象とする。内容に基づくフィルタリングにおけるユーザの嗜好は、楽曲単位で評価することが一般的である。しかし、「歌詞は嗜好に合うが、歌声は嗜好に合わない」や「歌声と伴奏は嗜好に合うが、歌詞は嗜好に合わない」のように、音楽要素に対する嗜好があると考えられる。楽曲単位の評価では、そうした情報は得られないため、ユーザの嗜好を十分に反映した推薦が行えないという問題がある (問題 1)。

ここで仮に、楽曲の音楽要素に対するユーザの嗜好が明示的に分かれば、音楽要素の類似度に基づいた楽曲推薦が可能になる。例えば、ある楽曲の歌詞が嗜好に合っていると評価したユーザに対しては、その楽曲と歌詞が類似した楽曲を推薦できる。しかし、どの程度類似した歌詞を持つ楽曲であれば、そのユーザの嗜好に合うかは自明ではない。楽曲を聴く際に歌詞に対するこだわりが強いユーザであれば、類似度の高い歌詞を持つ楽曲だけを嗜好に合っていると判断する一方で、歌詞に対するこだわりが弱いユーザであれば、類似度が低い歌詞を持つ楽曲であっても嗜好に合っていると判断するであろう。すなわち、音楽要素に対する嗜好が得られただけでは、ユーザがその音楽要素に関してどの程度の類似度までを嗜好に合う範囲として許容

¹ 筑波大学

² 産業技術総合研究所

^{a)} s2420785@u.tsukuba.ac.jp

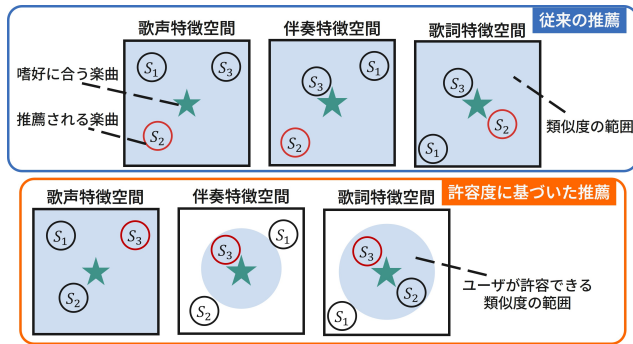


図 1: 各音楽要素の嗜好に基づいた推薦で、ユーザが許容できる類似度の範囲の指定が無い場合と、指定した場合。

できるかが分からないため、推薦すべき楽曲を適切に絞り込めないという問題がある（問題 2）。例えば図 1 の上段の例では、ユーザが各音楽要素に対して許容できる類似度の範囲の指定が無いため、楽曲 S_1 , S_2 , S_3 のいずれもが推薦候補となり、その中で嗜好に合う楽曲と最も類似度が高い楽曲 S_2 が推薦されてしまう。

本稿では、上記の 2 つの問題を解決するために、音楽要素ごとの嗜好と、各音楽要素に対してどの程度の類似度を許容できるかを指定可能な楽曲推薦インタフェースを提案する。具体的には、問題 1 を解決するために、音楽要素ごとに嗜好を評価できる仕組みを導入した。本研究では、歌声・伴奏・歌詞の 3 つの音楽要素に対する嗜好を評価できるようにした。これらの要素を選定した理由は、音楽に詳しくないユーザでも聴取を通じて比較的容易に判別できる特徴であると考えたためである。これにより、ユーザの嗜好に合う音楽要素と類似した音楽要素を持つ楽曲の推薦が可能になる。問題 2 を解決するために、音楽要素ごとにどの程度の類似度までを嗜好に合う範囲として許容できるかの度合い（本研究では「許容度」と呼ぶ）を指定できるようにした。本研究では、ユーザが指定した許容度に応じて、ユーザの嗜好に合う楽曲の範囲が各音楽要素でどのように変化するかを、ユーザがリアルタイムに確認できるように可視化した。これにより、ユーザの許容度が低い音楽要素（こだわりが強い音楽要素）の類似度を重視した楽曲の推薦が可能になる。

上記の 2 つの解決策を用いることで、図 1 の下段の例に示すように、ユーザが各音楽要素に対して許容できる類似度の範囲を指定できるようになり、上段で推薦されていた楽曲 S_2 は範囲外となる。その結果、ユーザの許容度を満たす楽曲 S_3 を推薦できるようになる。

本研究では評価実験として、提案する楽曲推薦インタフェースを実装し、ユーザごとの各音楽要素に対する許容度、各音楽要素の評価に対する満足度および最終的な推薦結果に対するユーザの満足度を調査した。その結果、楽曲を音楽要素ごとに評価することは効果的であることに加

え、ユーザ・音楽要素ごとに許容度は異なることが明らかになった。

2. 楽曲推薦インタフェースが満たすべき要件

提案するインタフェースを実現するために必要な要件は以下の 5 つである。

- 要件 1** 歌声・伴奏・歌詞の各音楽要素をユーザが評価できる機能
- 要件 2** 音楽要素ごとに、楽曲同士の類似度をユーザが俯瞰して把握できる可視化機能
- 要件 3** ユーザが指定する許容度に応じて、嗜好に合う楽曲の範囲の変化をユーザがリアルタイムに把握できる可視化機能
- 要件 4** ユーザに評価してもらう楽曲をインタフェースが探索する機能
- 要件 5** 音楽要素ごとのユーザの嗜好の度合いと許容度を満たす楽曲をインタフェースが推薦する機能

要件 1 は、1 章で述べたように、歌声・伴奏・歌詞の 3 つの音楽要素に対するユーザの嗜好を取得できる必要があるため設定した。さらに、ユーザが各音楽要素に対する許容度を直感的に指定できる必要があるため、要件 2 と要件 3 を定めた。要件 2 は、ユーザがある楽曲と他の楽曲がどの程度類似しているかを視覚的に確認できる必要があるため設定した。要件 3 は、ユーザが指定した許容度に応じて、推薦候補の楽曲が変化することを視覚的に把握できる必要があるため設定した。要件 4 は、楽曲推薦はユーザが評価した楽曲に基づいて行われるため、インタフェースがユーザに評価してもらう楽曲を探索する必要があるため設定した。一般的に、楽曲の探索というとユーザが自身の嗜好に合う楽曲を探索することを指すが、本研究では「インタフェースがユーザに評価してほしい楽曲を探索する」ことを意味する。例えば歌詞の許容度が低いユーザが、ある楽曲の歌詞を好きと評価した場合、その楽曲と十分に類似していない他の楽曲まで好きかどうかは明らかではないため、楽曲の探索機能により、それらの楽曲の中から次にユーザに評価してもらう楽曲を探索し、ユーザに提示する。要件 5 は、1 章で述べたように、ユーザの許容度が低い音楽要素に対して嗜好に合う楽曲をインタフェースが推薦する必要があるため設定した。

3. 関連研究

本章では、音楽要素に関する研究と楽曲推薦・探索インタフェースに関する関連研究について述べる。

3.1 音楽要素に関する研究

音楽要素に着目した関連研究の 1 つとして、Tsukuda

ら [16] は、人が初めて聴く楽曲を好きかどうか判断する際に、どのような音楽要素を重要視するかを調査した。その結果、歌声や歌詞といった要素が重要視されていることが明らかになった。Watanabe ら [18] は、歌詞の単語と、音響特徴、アーティスト名をユーザが任意に組み合わせて楽曲を柔軟に探索できる Query-by-Blending を提案した。これらのことから、音楽要素に着目した楽曲推薦手法は、ユーザの嗜好をより詳細かつ柔軟に考慮できる可能性がある。

3.2 楽曲推薦・探索インタフェース

ユーザが音楽情報に対してどの程度重視するかを指定できる推薦インタフェースに関する研究が行われてきた。Bostandjiev ら [19] は、アーティストとウェブ情報に対して、ユーザがどの程度考慮したいかをスライダで入力し、アーティストを推薦する TasteWeights を提案した。Millecamp ら [20] は、楽曲推薦において重要視したい属性をユーザがスライダもしくはレーダーチャートで指定できるインタフェースを提案した。

また、音楽情報を可視化する研究も行われてきた [21]。Liang ら [22] は、楽曲のジャンルについて、2次元空間上の等高線プロットで表現し推薦するインタフェースを提案した。Takahashi ら [23] らは、楽曲の楽器編成を色を用いた円グラフで可視化し探索できる INSTRUDIVE を提案した。倉持ら [24] は、音楽動画のコメントにおけるコンテキスト (ユーザの状況) および感情を、棒グラフとドーナツチャートで可視化し推薦する MusicCommentVisualizer を提案した。Tsukuda ら [25] は、ユーザと楽曲を2次元空間上に可視化し探索できる機能を持つ Kiite World を提案した。

楽曲の歌声や歌詞に着目したインタフェースも提案されてきた。Hamasaki ら [26] は、同曲異唱楽曲を2次元空間上で可視化し探索できる機能を持つ Songrium を提案した。Sasaki ら [27] は、歌詞のトピックおよび類似度をレーダーチャートと2次元空間上のマッピングから可視化し探索できる LyricsRadar を提案した。Tsukuda ら [28] は、楽曲の歌詞のトピックをドーナツチャートで可視化し、アーティストを推薦する機能を持つ Lyric Jumper を提案した。

楽曲の音響特徴に着目したインタフェースも提案されてきた。Lamere ら [29] は、音響特徴の類似度を基に楽曲を3次元空間に配置し、楽曲のアルバムアートを用いて可視化し推薦するインタフェースを提案した。Vad ら [30] は、音響特徴の類似度を基に楽曲を2次元空間上の点として配置し、楽曲のモードやテンポなどの情報を点の色を用いて可視化するインタフェースを提案した。また、Saito ら [31] は、楽曲のテンポや音量といった5種類の音響特徴の中から、ユーザが任意に2つの特徴を選択し、それらを軸とした2次元空間に配置する MusiCube を提案した。MusiCube は、ユーザが楽曲を評価できる機能を持ち、高評価または

低評価した楽曲を異なる色で可視化する。MusiCube を用いることで、ユーザが設定した空間における色の分布から嗜好が類似したユーザを見つけることができると示唆している。

4. 音楽要素ごとの嗜好と許容度を考慮できる楽曲推薦インタフェース

本章では、2章で述べた要件を満たすようなインタフェースの概要と操作手順を説明したうえで、インタフェースを実現するための詳細を述べる。

4.1 インタフェースの概要

本研究で作成したインタフェースを図2と図3に示す。図2は、ユーザが1曲目に聴く楽曲の選択する画面である。図3は、ユーザが楽曲の聴取・評価・許容度の指定を行う画面である。画面の上部では、歌声 (VOCAL)・伴奏 (ACCOMPANIMENT)・歌詞 (LYRICS) の3つの音楽要素ごとの嗜好と許容度に応じて、嗜好に合う楽曲の範囲がどのように変化するかを可視化する。これにより、ユーザは嗜好に合う楽曲と類似する楽曲との関係を各音楽要素で視覚的に把握しながら、好みに合う楽曲を見つけることができる。ユーザが入力した音楽要素ごとの嗜好と許容度を基にして、インタフェースは楽曲の推薦およびユーザに評価してもらう楽曲の探索を行う。

インタフェースは「音楽要素ごとの空間」、「許容度スライダ」、「動画再生プレーヤ」、「評価エリア」、「探索・推薦タブ」の5つの機能で構成されており、Django を用いて実装した。

4.2 操作手順

本節では、提案インタフェースの操作手順について述べる。1曲目に関する操作は以下のとおりである。

- (1) ユーザは1曲目に聴く楽曲を選択する (図2)。
- (2) 「START」ボタンを押すと図2から図4の1曲目の操作画面に切り替わる。
- (3) 選択した楽曲を動画再生プレーヤで聴取し、歌声・伴奏・歌詞の音楽要素がどの程度嗜好に合うかを5段階で評価する。ただし、ユーザは全要素を評価する必要はなく、評価されていない要素は、中間の値 (黄色) が自動的に割り当てられる。
- (4) 音楽要素ごとの許容度を、許容度スライダからそれぞれ指定する。
- (5) 図4の1曲目の操作画面の右下にある「NEXT SONG」ボタンを押すことで、図4の2曲目以降の操作画面に切り替わる。

2曲目以降の操作は以下のとおりである。

- (6) 以下の3種類の楽曲の中から、次に聴取する楽曲をユーザが選択する。



図 2: ユーザが 1 曲目に聴く楽曲を選択する画面 (セレクトボックスを押すと、選択候補の楽曲がプルダウンで表示され、その中から選ぶか、最上部の検索ボックスから楽曲名を入力して選択することができる)。

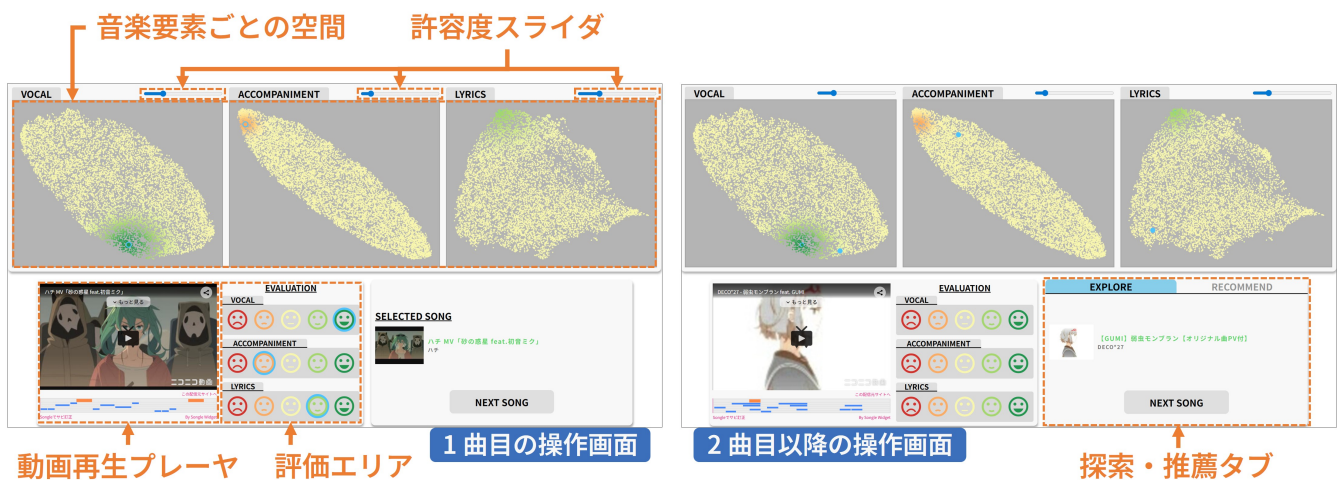


図 3: 選択した楽曲の聴取・評価・許容度の指定を行う画面 (左図が 1 曲目の操作画面、右図が 2 曲目以降の操作画面)。

- (a) 探索タブを押すと、ユーザが次に評価する楽曲が 1 曲タブ内に提示され、動画再生プレーヤで再生される。
- (b) 推薦タブを押すと、ユーザの嗜好に合う楽曲が 5 曲タブ内に提示され、その中の 1 曲が動画再生プレーヤで再生される。
- (c) 音楽要素ごとの空間上で気になった点を押すことで、その点に該当する楽曲が動画再生プレーヤで再生される。
- (7) (3) と (4) の操作を行う。
- (8) 図 4 の 2 曲目以降の操作画面の探索タブ内にある「NEXT SONG」ボタンを押すことで、探索・推薦タブで提示される楽曲が更新される。ただし、探索タブで提示される楽曲を評価しなければ、探索・推薦タブで提示される楽曲は更新されない。

以上の (6)～(8) をユーザは繰り返し操作しながら、楽曲の評価と許容度の入力を行う。加えて、探索・推薦タブで提示する楽曲を更新するには、(8) の操作を行う。(8) において「NEXT SONG」ボタンを利用できる条件として、

探索タブで提示される楽曲を評価したときに限定した理由は、推薦タブで提示された楽曲のみをユーザが繰り返し聴取・評価することを防ぐためである。このような偏った評価が生じた場合、例えば歌詞の許容度が低いユーザが、ある楽曲の歌詞を好きと評価したとしても、その楽曲と十分に類似していない他の楽曲は推薦されないため、ユーザがそうした楽曲を好きかどうかをインターフェースは判断できない。そのため、探索タブで提示される楽曲を評価したときという条件を設けることで、ある楽曲と十分に類似していない他の楽曲までも考慮した推薦が可能である。

4.3 評価エリア

本節では、2 章の要件 1 (各音楽要素をユーザが評価) を満たすための機能について説明する。評価エリアでは、ユーザは聴取している楽曲の歌声・伴奏・歌詞の各音楽要素がどの程度嗜好に合っているかを入力する。評価に関しては「嫌い」、「少し嫌い」、「どちらでもない」、「少し好き」、「好き」に対応した 5 段階のリッカート尺度を用いた。この尺度を用いた理由は、3 段階の尺度と比べて、ユーザからの

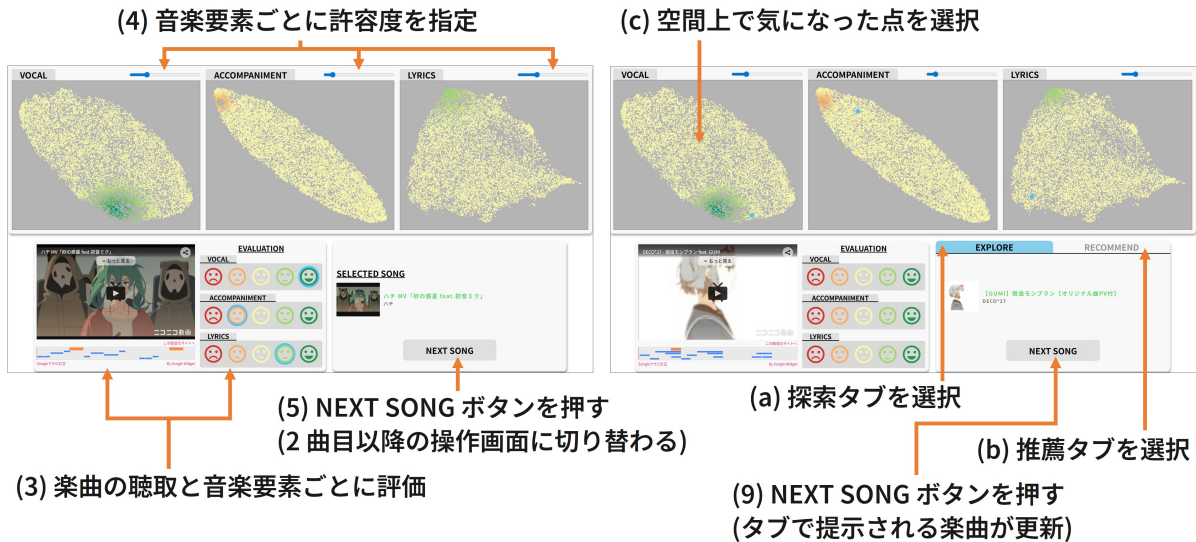


図 4: 1 曲目の操作画面と、2 曲目以降の操作画面における操作手順。

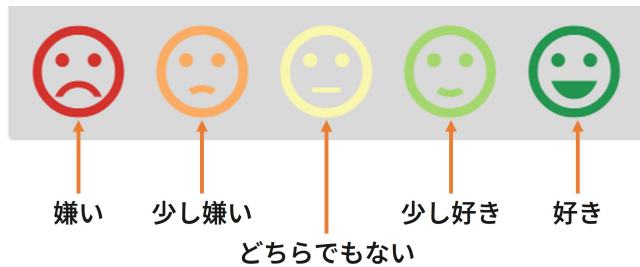


図 5: 評価の入力ボタンの詳細。

詳細な評価を得ることができると考えたためである。加えて、ユーザは何度も評価を行うため、7 段階のような細かな尺度はユーザ体験を損なう恐れがあると考え、使用しなかった。また、評価の入力ボタンには、図 5 のフェイス・スケール [32] を利用した。フェイス・スケールを用いた理由は、評価値の入力がより直感的に理解しやすいと判断したためである。

4.4 音楽要素ごとの空間

本節では、2 章の要件 2(各音楽要素で楽曲同士の類似度を可視化)を満たすための機能を説明する。

本研究では、動画共有サービス「ニコニコ動画*1」に投稿されている歌声合成楽曲の再生数上位 1 万曲のうち、歌詞の解析が可能であった 9103 曲を用いた。歌声合成楽曲を利用した理由は、様々なジャンルの楽曲が投稿されていることに加え、4.6 節で述べるように、楽曲のサビ区間の情報を利用した楽曲の再生が可能のためである。

歌声と伴奏に関しては、HT Demucs [33] を用いて各楽曲を歌声と伴奏に分離したうえで、MERT [34] を用いてそれぞれ 1024 次元のベクトルを取得し、歌声と伴奏のベクトル空間を作成した。歌詞に関しては、Sentence-BERT [35]

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

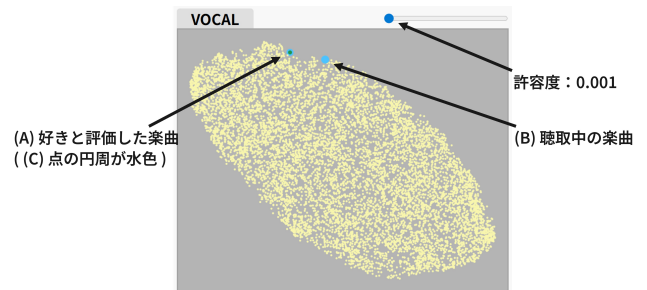


図 6: 音楽要素ごとの空間と評価値の例 (歌声空間で許容度が 0.001 の場合)。

を用いて各楽曲の歌詞を 384 次元のベクトルを取得し、歌詞のベクトル空間を作成した。そして、各ベクトル空間の標準化を行い、UMAP [36] を用いて 2 次元ベクトル空間に変換した後、横軸、縦軸ともに 0 から 1 の範囲を取るように正規化を行うことで各空間上の楽曲の座標値を決定した。UMAP は、t-SNE [37] と比べて、空間の局所的な構造だけでなく、大域的な構造も考慮できるという特徴がある。例えば、歌詞の許容度が高いユーザの場合、歌詞空間上でより広範囲の楽曲まで考慮する必要があるため、大域的な構造も考慮できる特徴を持つ UMAP が有効であると考える採用した。

ユーザがある楽曲のある音楽要素に対して評価すると、図 6 のように、その音楽要素の空間内で、その楽曲に対応する点がユーザの評価に応じて着色される。例えば、ある楽曲の歌声に関して「好き」(緑色のとても喜んでいる顔)と入力した際は、歌声の空間上でその楽曲に該当する点の色が緑色(図 6 (A))に変化する。また、評価前で聴取中の楽曲の点は水色(B)で表現し、評価済みの楽曲になった際は点の円周を水色(C)で表現した。

4.5 未評価楽曲の評価の推定値と許容度スライダ

本節では、2章の要件3(許容度を基に、嗜好に合う楽曲の変化を可視化する機能)を満たすために、各空間における許容度および評価済み楽曲の位置と評価値に基づいて、未評価楽曲の各音楽要素に対する評価の推定値を求める。例えばユーザが歌詞を「好き」と評価した楽曲があったときに、歌詞の空間内ではその楽曲との距離が遠くなるにつれて、楽曲の歌詞に対する好みの度合いは徐々に下がると考えられる。そのため、評価済み楽曲の位置と評価値を基に、各空間における未評価楽曲の評価の推定値は2次元ガウス分布に従うと仮定する。ここで、歌声を V 、伴奏を A 、歌詞を L とし、音楽要素の集合を $E = \{V, A, L\}$ とすると、音楽要素 $e \in E$ における未評価の楽曲 i の評価の推定値 Z_{e_i} を次式から求める。

$$Z_{e_i} = \frac{w_e}{2\pi\sigma_e^2} \exp\left(-\frac{(x_e - \mu_{x_e})^2 + (y_e - \mu_{y_e})^2}{2\sigma_e^2}\right). \quad (1)$$

ここで、ガウス分布は相関がなく(共分散行列の非対角成分を0)かつ、対角成分が同一の値 σ とした。 w は評価値に対応する重み(「嫌い」は -1 , 「少し嫌い」は -0.5 , 「どちらでもない」は 0 , 「少し好き」は 0.5 , 「好き」は 1)、 (x, y) は未評価の楽曲の座標(横軸が x , 縦軸が y)、平均 (μ_x, μ_y) は評価済み楽曲の座標、標準偏差 σ は、ユーザから入力された許容度である。許容度は、ユーザがインタフェース上のスライダから $0.001 \sim 0.5$ の範囲を 0.001 刻みで指定可能であり、初期値は 0.05 である。許容度 σ が大きい場合、評価済み楽曲との類似度が低い(距離が遠い)楽曲に対しての Z_{e_i} が高くなる。

評価済み楽曲が複数ある場合は、評価済み楽曲からそれぞれ求められた Z_{e_i} の値を加算して、その合計値に応じて、空間上の点の色をグラデーションのように変化する表現をした。ただし、合計値が 1 より大きいときは 1 , -1 より小さいときは -1 の色とした。例えば、ユーザがある楽曲の歌声に対して「好き」と評価して、許容度が 0.001 の場合を図6、許容度が 0.5 の場合は図7に示す。これらを比較すると、実際に評価した楽曲以外に対しての、評価の推定値が異なる(許容度が大きい図7の方が、より遠くまで影響を与える)。

これにより、評価値と許容度に応じた Z_{e_i} の変化は、リアルタイムに更新され、ユーザは即座にその効果を見ることができる。ユーザは、このような空間上の点の色や位置などを参考にしながら、気になった点(楽曲)をクリックして楽曲動画を再生したり、その楽曲の音楽要素に対する嗜好を評価したりできる。

4.6 動画再生プレーヤ

動画の再生は、ブラウザ内に埋め込みができる Songle Widget [38] を用いた。Songle Widget では、能動的音楽鑑賞サービス Songle [39] により解析された楽曲のサビ区間

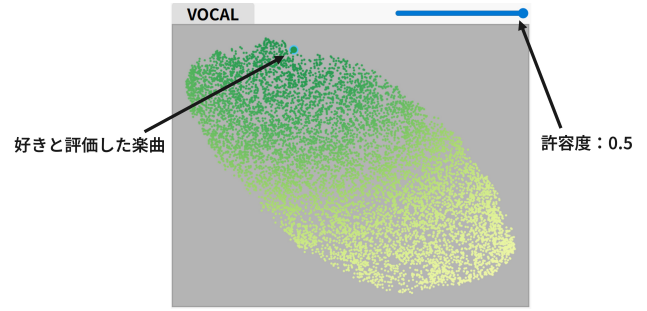


図7: 歌声を好きと評価した楽曲と許容度が 0.5 の場合。

がプレーヤの下部に表示されており、ユーザはサビ区間を選択することで簡単にサビを聴くことができる。

4.7 探索・推薦タブ

本節では、2章の要件4(楽曲探索)と要件5(楽曲推薦)を満たすための機能について説明する。ユーザがインタフェース上の「探索」タブを押すと、自身とその周辺が未評価の楽曲が1曲提示される(図8:左)。そのような楽曲の選定のために、未評価の楽曲の中から、評価の推定値 Z_{e_i} の絶対値が 0 に近い(評価済み楽曲から遠い)楽曲を対象とする。具体的には、 $-\tau \leq Z_{e_i} \leq \tau$ となる楽曲数が最も多い音楽要素を対象とし(本研究では $\tau = 0.3$ とした)、その範囲に含まれる楽曲の中からランダムに1曲を選択する。このような楽曲に対して評価が得られれば、 Z_{e_i} の多くが評価済み楽曲に基づくようになり、より信頼性の高い推薦が可能となる。

「推薦」タブを押すと、音楽要素ごとの嗜好の度合いと許容度に基づいた楽曲が提示される(図8:右)。具体的には、未評価の楽曲 i に対して、 Z_{e_i} の平均値 R_i を次式から求める。

$$R_i = \frac{Z_{V_i} + Z_{A_i} + Z_{L_i}}{3}. \quad (2)$$

R_i が高い順に上位5曲を推薦楽曲とし、推薦タブ上で1位から昇順に提示する。例えば、ユーザが歌詞の許容度が低く、ある楽曲を「好き」と評価した場合、評価した楽曲との類似度が高い楽曲の Z_{L_i} が高くなり、推薦で重視されやすい。そのため、 R_i を用いることで、ユーザは許容度の低い音楽要素に対して自身の嗜好を重視した楽曲推薦を受けることができる。また、提示された楽曲は、1位の楽曲から順に動画再生プレーヤで再生されるが、ユーザが気になった楽曲をクリックすることでその楽曲から再生することも可能である。

5. 評価実験

ユーザごとの各音楽要素に対する許容度の違いや、各音楽要素の評価に対する満足度および最終的な推薦結果に対するユーザの満足度を調査し、提案インタフェースの有効性を検証する。

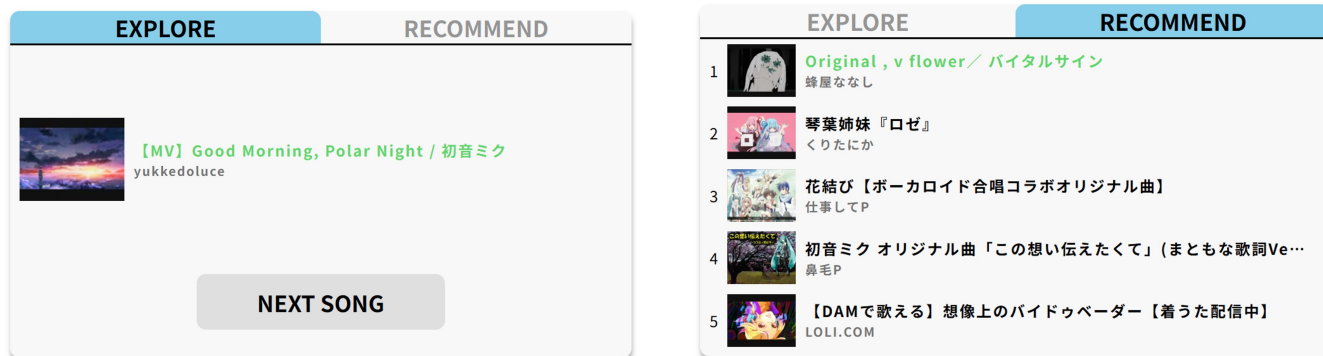


図 8: 探索結果を提示する画面 (左, 探索タブ) と推薦結果を提示する画面 (右, 推薦タブ).

タブ内には楽曲のサムネイル画像と楽曲名, クリエイター名が表示され, 選択中の楽曲は楽曲名が緑色で表示される.

5.1 実験方法

実験参加者は, 「日常的に音楽鑑賞するが, 専門的な教育を受けたことはない人」に該当する, 筆頭著者 1 名 (参加者 A とする) と, 男性の学生 2 名 (参加者 B, C とする) の合計 3 名である. 本実験は実験者と 1 対 1 の対面形式で実施した. 参加者はまず, 実験者からインタフェースの概要と使用方法の説明を受け, 操作練習を行った後にインタフェースを使用した. インタフェースの使用にあたっては, 以下の 2 つの条件を設けた. 1 つ目は, インタフェースの探索・推薦タブで提示される楽曲が更新された際, 参加者は推薦された楽曲を 1 曲以上評価することである. これは, 最終的な推薦結果に対する満足度を実験終了時に回答してもらうために必要である. 2 つ目は, 参加者が 1 曲目に選択する楽曲は, 既に聴取したことがあり, 音楽要素のいずれかについて「好き」または「少し好き」と評価できる楽曲とすることである. これは, 参加者の嗜好を取得することを目的としている. また, 楽曲の再生位置については指定せず, 参加者が任意の位置から再生して聴取できるようにした. インタフェースの使用時間は最大 30 分とし, 参加者が満足する推薦結果を得られた時点で使用を終了できるようにした.

インタフェースの使用後, 参加者はアンケートに回答した. アンケートは, 音楽要素ごとの評価に対する満足度および最終的な推薦結果に対する満足度を, 7 段階のリッカート尺度で選択する項目と, 自由記述欄で構成されており, ウェブ上のアンケートフォームを用いて行った.

5.2 実験結果

表 1 に各参加者のインタフェースの使用時間, 評価した楽曲数, 最終的な許容度の値, およびアンケートの回答結果を示す. ただし, 著者によるアンケートの回答は, 結果の客観性を損なう可能性があるため除外した. インタフェースの使用時間に関して, 参加者 A と B は実験で設定した 30 分よりも早い時間に終了し, 評価した楽曲数はそれぞれ 16 曲と 18 曲であった. 参加者 C は 30 分が経過

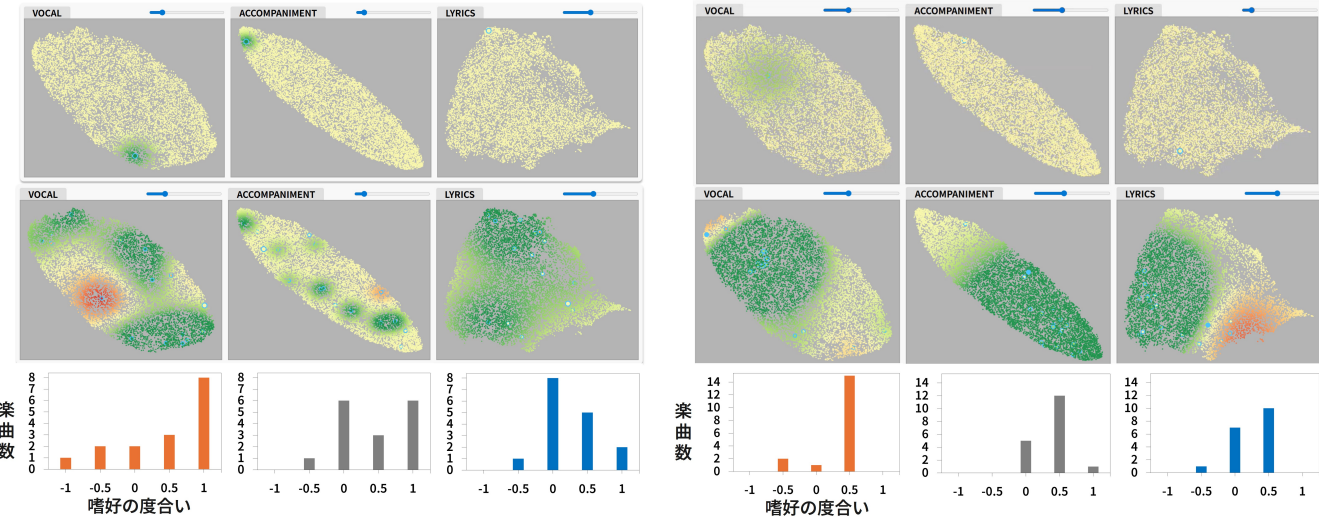
したため使用を終了し, 評価した楽曲数は 20 曲であった. 最終的な許容度の値は参加者によって異なり, 例えば歌詞に対する許容度は参加者 B が 0.213 と最も高く, 参加者 C が 0.057 と最も低かった. また, 音楽要素ごとによって許容度は異なり, 例えば参加者 A の伴奏に対する許容度は 0.054, 歌詞に対する許容度は 0.162 であった. 音楽要素ごとの評価に対する満足度は, それぞれの参加者で 5 以上と高かった. 最終的な推薦結果に対する満足度も平均値 5 と高かった. また, 自由記述欄では参加者 B から「最終的な推薦結果の上位 3 曲は自分の好みに合っていた」という回答や参加者 C から「それぞれの空間上の点がどのような理由で似ているか明らかだと使いやすい」「許容度の指定が少し難しかった」という回答があった.

図 9 に各参加者が 1 曲目を評価したときの空間, インタフェースの使用を終了したときの空間, 評価された楽曲の各音楽要素における嗜好の度合いの分布を示す. 図 9 の上段と中段の図から, 各参加者で色や色の範囲に差がある. 色に関しては, 例えば図 9(i) の中段の歌詞空間は「好き」に該当する緑色の点が多かったが, 図 9(iii) の中段の歌詞空間は「嫌い」に該当する色が多かった. 色の範囲に関しては, 図 9(i) の中段の伴奏空間は, 小さい緑色の点の広がりが見られたが, 図 9(ii) の中段の伴奏空間は緑の大きな広がりが見られた. 評価された楽曲の各音楽要素における嗜好の度合いの分布は, 例えば図 9(iii) の下段で伴奏に対しては「好き」「少し好き」という評価が多かったが, 歌声に対しては「嫌い」「少し嫌い」という評価も多かった. 図 9(iii) の下段で歌詞に対しては「好き」「少し好き」という評価がやや多かったが, 中段の歌詞空間では「好き」「少し好き」の評価の推定値 Z_{e_i} に該当する緑色の分布は見られなかった. 図 9(iii) の下段で歌声に対して「好き」という評価はなかったが, 中段の歌声空間では Z_{e_i} が「好き」に該当する点の色になっていた.

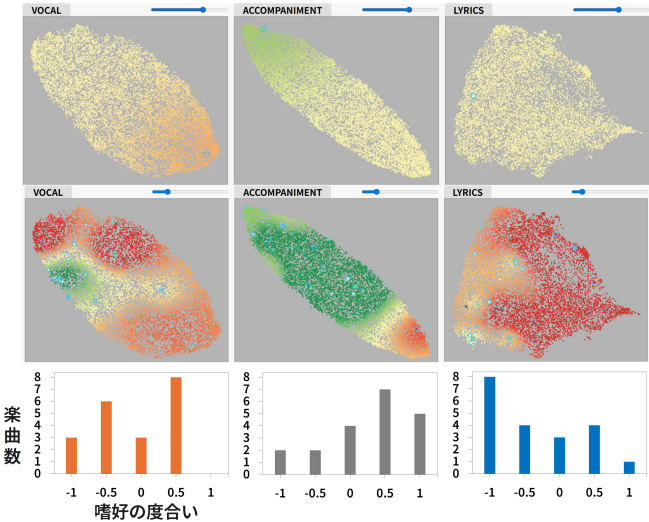
図 10 に各参加者における音楽要素ごとの許容度の推移を示す. まず, 各参加者で 3 曲目までにいずれかの音楽要素で許容度の変化が起きていた. 例えば, 図 10(ii) にお

表 1: 参加者のインタフェースの使用時間と、評価した楽曲数、最終的な許容度、満足度.

実験参加者	インタフェースの使用時間 (分)	評価した楽曲数	最終的な許容度			満足度	
			歌声	伴奏	歌詞	音楽要素ごとの評価	最終的な推薦結果
A (著者)	25	16	0.102	0.054	0.162	-	-
B	26	18	0.161	0.195	0.213	6	6
C	30	20	0.092	0.086	0.057	5	4



(i) 参加者 A. (ii) 参加者 B.



(iii) 参加者 C.

図 9: 各参加者が 1 曲目を評価したときの空間 (上段), インタフェースの使用を終了したときの空間 (中段), 嗜好の度合いの分布 (下段). 嗜好の度合いは「嫌い」を -1 , 「少し嫌い」を -0.5 , 「どちらでもない」を 0 , 「少し好き」を 0.5 , 「好き」を 1 とする.

ける歌詞の許容度は 0.050 から 0.213 と変化していた. また, 図 10(i) における歌声の許容度は, 5 曲目に 0.124 から 0.102 に低下した. このとき歌声に対する評価は「嫌い」であった. 図 10(ii) における 10 曲目の伴奏の許容度は, 0.314 から 0.195 に低下していた. 図 10(ii) の伴奏では, 4 曲目から 9 曲目まで「好き」または「少し好き」の評価で, 10 曲目は「どちらでもない」という評価であった. 参加者

が後半に聴取した楽曲に対しては, 各参加者で音楽要素ごとの許容度に変化は見られなかった. 例えば図 10(iii) の各音楽要素で 17 曲目以降の許容度は, どの音楽要素でも変化していなかった.

5.3 考察

実験参加者の人数が少ないために得られた結果の一般性

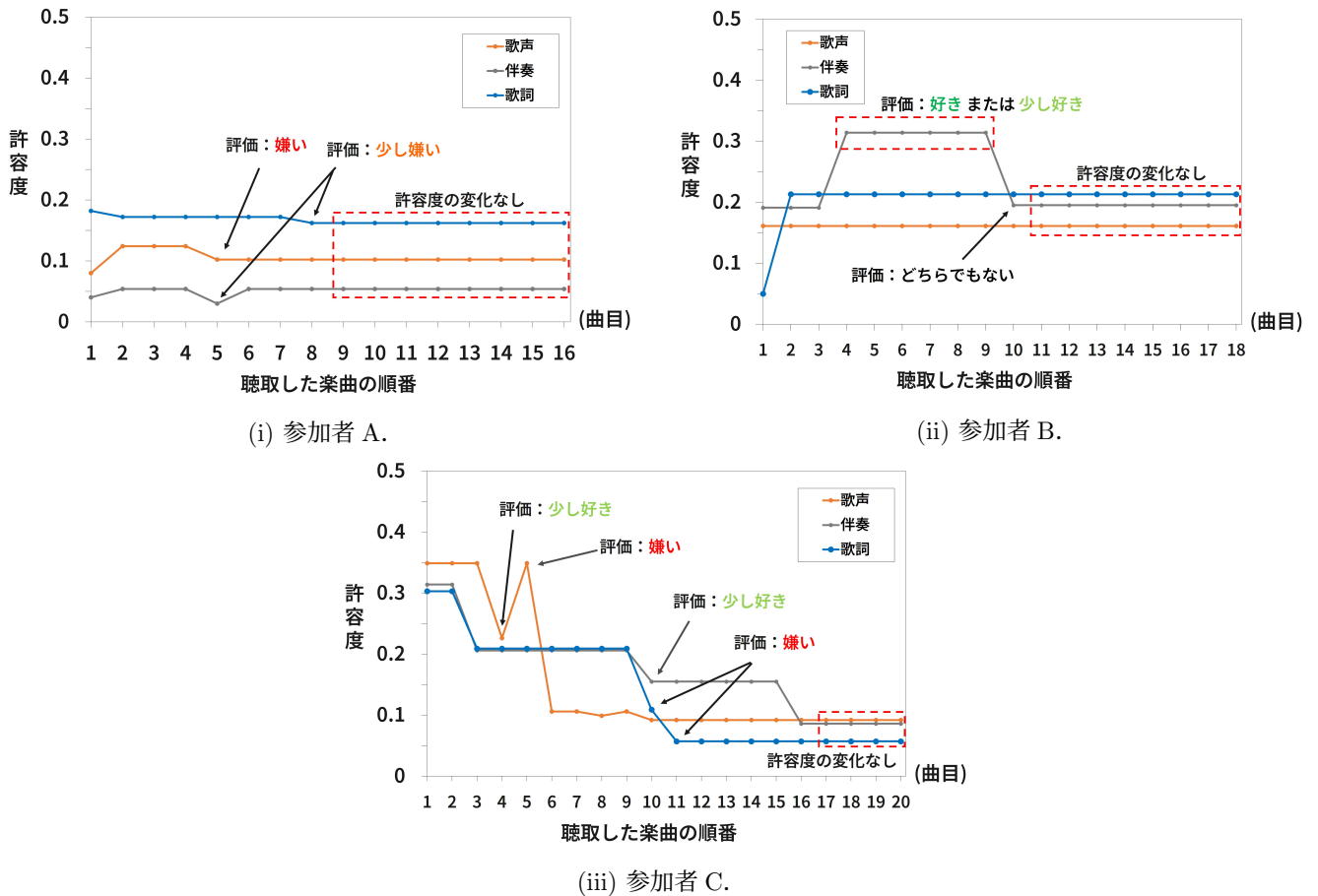


図 10: 各参加者における音楽要素ごとの許容度の推移。

には限界があるが、アンケート結果での満足度が高かったことから、ユーザが音楽要素ごとに楽曲を評価することが有効であることが考えられる。各参加者における最終的な音楽要素ごとの許容度の結果から、ユーザ・音楽要素ごとに許容度は異なることが明らかになった。図 9 に対する結果から、提案インタフェースでは評価の推定値 Z_{e_i} を加算したときの値に応じた色を可視化していたため、一度も「好き」と評価していない要素で「好き」に該当する緑色の点が表示されていた。そのため、 Z_{e_i} の推定に改善の余地があると考えられる。さらに、図 10 の結果では、いずれの実験参加者も許容度の変更はできていたものの、前述のように参加者 C の自由記述欄には「許容度の指定が少し難しかった」という回答もあり、今後は許容度を設定しやすくしたり自動推定したりする発展が考えられる。

6. おわりに

本研究では、音楽要素ごとの嗜好と許容度をユーザが指定可能な楽曲推薦インタフェースを提案した。音楽要素は、歌声・伴奏・歌詞の 3 つを対象とし、それぞれの音楽要素でどの程度の類似度までを嗜好に合う範囲として許容できるかの度合いの「許容度」をインタフェース上で指定できるようにした。評価実験として、ユーザごとの各音楽要素に

対する許容度、各音楽要素の評価に対する満足度および最終的な推薦結果に対するユーザの満足度を調査し、提案インタフェースの有効性を検証した。その結果、楽曲を音楽要素ごとに評価することは効果的であることに加え、ユーザ・音楽要素ごとに許容度は異なることが明らかになった。

今後は、ユーザの許容度を自動的に推定することで、よりの確に許容度を扱えるインタフェースを目指していきたい。

謝辞 本研究の一部は JST CREST JPMJCR20D4 と JSPS 科研費 JP25H01174 の支援を受けた。

参考文献

- [1] 奥 健太: 楽曲推薦システム—プレイリスト, コンテキスト, インタクション—, 人工知能学会誌, Vol. 34, No. 3, pp. 300–308 (2019).
- [2] Schedl, M., Knees, P., McFee, B. and Bogdanov, D.: Music Recommendation Systems: Techniques, Use Cases, and Challenges, *Recommender Systems Handbook*, Springer, pp. 927–971 (2022).
- [3] Amiri, B., Shahverdi, N., Haddadi, A. and Ghahremani, Y.: Beyond the Trends: Evolution and Future Directions in Music Recommender Systems Research, *IEEE Access*, Vol. 12, pp. 51500–51522 (2024).
- [4] 吉井和佳, 後藤真孝: 音楽情報処理技術の最前線: 7. 音楽推薦システム, 情報処理学会, Vol. 50, No. 8, pp. 751–755

- (2009).
- [5] Sánchez-Moreno, D., González, A. B. G., Vicente, M. D. M., Batista, V. F. L. and García, M. N. M.: A collaborative filtering method for music recommendation using playing coefficients for artists and users, *Expert Systems with Applications*, Vol. 66, pp. 234–244 (2016).
 - [6] Liu, X., Yang, Z. and Cheng, J.: Music recommendation algorithms based on knowledge graph and multi-task feature learning, *Scientific Reports*, Vol. 14, No. 1, p. 2055 (2024).
 - [7] Deldjoo, Y., Schedl, M. and Knees, P.: Content-driven music recommendation: Evolution, state of the art, and challenges, *Computer Science Review*, Vol. 51, p. 100618 (2024).
 - [8] Kowald, D. and Lalic, E.: Popularity Bias in Collaborative Filtering-Based Multimedia Recommender Systems, *Advances in Bias and Fairness in Information Retrieval (BIAS 2022)*, Communications in Computer and Information Science, Vol. 1610, pp. 1–11 (2022).
 - [9] Logan, B.: Music Recommendation from Song Sets, *Proceedings of the 5th International Society for Music Information Retrieval Conference (ISMIR 2004)*, pp. 425–428 (2004).
 - [10] van den Oord, A., Dieleman, S. and Schrauwen, B.: Deep content-based music recommendation, *Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS 2013)*, pp. 2643–2651 (2013).
 - [11] Pulis, M. and Bajada, J.: Siamese Neural Networks for Content-based Cold-Start Music Recommendation, *Proceedings of the 15th ACM Conference on Recommender Systems (ACM RecSys 2021)*, pp. 719–723 (2021).
 - [12] Kostrzewa, D., Chrobak, J. and Brzeski, R.: Attributes Relevance in Content-Based Music Recommendation System, *Applied Sciences*, Vol. 14, No. 2 (2024).
 - [13] Yang, J.: Personalized Song Recommendation System Based on Vocal Characteristics, *Mathematical Problems in Engineering*, Vol. 2022, No. 1, p. 3605728 (2022).
 - [14] Hashizume, Y., Li, L. and Toda, T.: Music Similarity Calculation of Individual Instrumental Sounds Using Metric Learning, *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (IEEE APSIPA ASC 2022)*, pp. 33–38 (2022).
 - [15] Gossi, D. and Gunes, M. H.: Lyric-Based Music Recommendation, *Complex networks VII: Proceedings of the 7th Workshop on Complex Networks (CompleNet 2016)*, pp. 301–310 (2016).
 - [16] Tsukuda, K., Nakano, T., Hamasaki, M. and Goto, M.: Unveiling the Impact of Musical Factors in Judging a Song on First Listen: Insights From a User Survey, *Proceedings of the 24th International Society for Music Information Retrieval Conference (ISMIR 2023)*, pp. 561–570 (2023).
 - [17] Nakano, T., Yoshii, K. and Goto, M.: Musical Similarity and Commonness Estimation Based on Probabilistic Generative Models of Musical Elements, *International Journal of Semantic Computing (IJSC)*, Vol. 10, No. 01, pp. 27–52 (2016).
 - [18] Watanabe, K. and Goto, M.: Query-by-Blending: A Music Exploration System Blending Latent Vector Representations of Lyric Word, Song Audio, and Artist, *Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR 2019)*, pp. 144–151 (2019).
 - [19] Bostandjiev, S., O'Donovan, J. and Höllerer, T.: TasteWeights: a visual interactive hybrid recommender system, *Proceedings of the 6th ACM Conference on Recommender Systems (ACM RecSys 2012)*, pp. 35–42 (2012).
 - [20] Millecamp, M., Htun, N. N., Jin, Y. and Verbert, K.: Controlling Spotify Recommendations: Effects of Personal Characteristics on Music Recommender User Interfaces, *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization (ACM UMAP 2018)*, pp. 101–109 (2018).
 - [21] 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).
 - [22] Liang, Y. and Willemsen, M. C.: The role of preference consistency, defaults and musical expertise in users' exploration behavior in a genre exploration recommender, *Proceedings of the 15th ACM Conference on Recommender Systems (ACM RecSys 2021)*, pp. 230–240 (2021).
 - [23] Takahashi, T., Fukayama, S. and Goto, M.: INSTRUDIVE: A Music Visualization System Based on Automatically Recognized Instrumentation, *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR 2018)*, pp. 561–568 (2018).
 - [24] 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, コンピュータ ソフトウェア, Vol. 42, No. 1, pp. 97–114 (2025).
 - [25] Tsukuda, K., Takahashi, T., Ishida, K., Hamasaki, M. and Goto, M.: Kiite World: Socializing Map-Based Music Exploration Through Playlist Sharing and Synchronized Listening, *Proceedings of the 31st International Conference on MultiMedia Modeling (MMM 2025)*, Lecture Notes in Computer Science, Vol. 15521, Springer, pp. 197–211 (2025).
 - [26] Hamasaki, M., Goto, M. and Nakano, T.: Songrium: Browsing and listening environment for music content creation community, *Proceedings of the 12th Sound and Music Computing Conference (SMC 2015)*, pp. 23–30 (2015).
 - [27] Sasaki, S., Yoshii, K., Nakano, T., Goto, M. and Morishima, S.: LyricsRadar: A Lyrics Retrieval System Based on Latent Topics of Lyrics, *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, pp. 585–590 (2014).
 - [28] Tsukuda, K., Ishida, K. and Goto, M.: Lyric Jumper: A Lyrics-Based Music Exploratory Web Service by Modeling Lyrics Generative Process, *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017)*, pp. 544–551 (2017).
 - [29] Lamere, P. and Eck, D.: Using 3D Visualizations to Explore and Discover Music, *Proceedings of the 8th International Society for Music Information Retrieval Conference (ISMIR 2007)*, pp. 173–174 (2007).
 - [30] Vad, B., Boland, D., Williamson, J., Murray-Smith, R. and Steffensen, P. B.: Design and Evaluation of a Probabilistic Music Projection Interface, *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015)*, pp. 134–140 (2015).
 - [31] Saito, Y. and Itoh, T.: MisiCube: A Visual Music Recommendation System featuring Interactive Evolutionary Computing, *Proceedings of the 2011 Visual Information Communication-International Symposium (ACM VINCI)*, pp. 1–6 (2011).

- [32] Lorish, C. D. and Maisiak, R.: The face scale: A brief, nonverbal method for assessing patient mood, *Arthritis & Rheumatism*, Vol. 29, No. 7, pp. 906–909 (1986).
- [33] Rouard, S., Massa, F. and Défossez, A.: Hybrid Transformers for Music Source Separation, *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP 2023)* (2023).
- [34] Li, Y., Yuan, R., Zhang, G., Ma, Y., Chen, X., Yin, H., Xiao, C., Lin, C., Ragni, A., Benetos, E., Gyenge, N., Dannenberg, R., Liu, R., Chen, W., Xia, G., Shi, Y., Huang, W., Wang, Z., Guo, Y. and Fu, J.: MERT: Acoustic Music Understanding Model with Large-Scale Self-supervised Training, *The Twelfth International Conference on Learning Representations (ICLR 2024)* (2024).
- [35] Reimers, N. and Gurevych, I.: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP 2019)*, pp. 3982–3992 (2019).
- [36] McInnes, L., Healy, J. and Melville, J.: UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, *arXiv preprint arXiv:1802.03426* (2018).
- [37] Van der Maaten, L. and Hinton, G.: Visualizing Data using t-SNE, *Journal of machine learning research (JMLR)*, Vol. 9, No. 11, pp. 2579–2605 (2008).
- [38] Goto, M., Yoshii, K. and Nakano, T.: Songle Widget: Making animation and physical devices synchronized with music videos on the web, *Proceedings of the 2015 IEEE International Symposium on Multimedia (IEEE ISM 2015)*, pp. 85–88 (2015).
- [39] Goto, M., Yoshii, K., Fujihara, H., Mauch, M. and Nakano, T.: Songle: A Web Service for Active Music Listening Improved by User Contributions, *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, pp. 311–316 (2011).