

PrefTolerance:

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

中澤 一輝^{*†} 佃 洸[†] 中野 倫靖[†] 後藤 真孝[†]

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、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。本論文では、これらの問題を解決するための楽曲推薦インタフェース PrefTolerance を提案する。具体的には、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。さらに、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。PrefTolerance は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とで、許容度の低さ（こだわりの強さ）を重視した楽曲推薦ができる。

1 はじめに

「歌詞は好みだけど歌声は好みでない」あるいは「歌声は好みだけど伴奏が好みでない」のように、音楽の好みは楽曲を構成する音楽要素 [13, 21] によって変わる。さらに、「歌詞はこだわらないので好みと違って許容できるけど、歌声は好みの曲しか聴きたくない」のように、音楽要素ごとにどの程度まで好みと違って許容できるかという度合い（許容度）は異なる。しかし、これまでの楽曲推薦システム [1, 20, 25] では、楽曲単位での嗜好に基づいており、許容度のような観点も考慮されていなかった。

そこで本論文では、歌声・伴奏・歌詞の3つの音楽要素ごとの嗜好（preference）と、各音楽要素に対する許容度（tolerance）をユーザが指定可能な楽曲推薦インタフェース PrefTolerance を提案する。ここでは「許容度」を、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と定義する。PrefTolerance は、ユーザが様々な楽曲の3つの音楽要素に対する嗜好を5段階で評価できる機能と、それぞれの要素に対する許容度をいつでも変更できる機能を備えており、これらの機能に基づいて楽曲推薦を行う。ユーザが許容度を直感的に指定できるようにするため、音楽要素ごとに別々に楽曲間の類似度を俯瞰して把握できる可視化を行う。さらに、ユーザが一度の操作で満足する許容度を決定することは困難であるため、スライダを用いてユーザが許容度を調整できるようにし、その

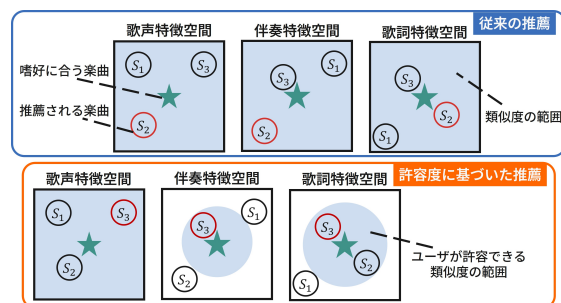


図 1. 各音楽要素の嗜好に基づいた推薦で、ユーザが許容できる類似度の範囲の指定が無い場合と、指定した場合。

変化に応じて嗜好に合う楽曲の範囲がどのように変わるかをインタラクティブに可視化する。

PrefTolerance により、単なる嗜好を超えた、ユーザのこだわりを反映した楽曲との出会いを実現できる。ユーザがある楽曲の歌詞を嗜好に合うと評価した場合、それと歌詞がどの程度類似した楽曲であれば、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（許容度が低い）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うと判断する一方で、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うと判断するであろう。すなわち、音楽要素ごとの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。例えば、図 1 の上段の例

Copyright is held by the author(s).

^{*} 筑波大学

[†] 産業技術総合研究所

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。この問題に対して、PrefTolerance では、図 1 の下段の例のように、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可能にする。これにより、ユーザは自身のこだわりが強い音楽要素の類似度を重視した楽曲推薦を体験することができる。

2 関連研究

本章では関連研究に対する本研究の位置づけを述べる。

2.1 推薦手法

楽曲推薦インタフェース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 [26]。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 [8, 18]。この手法は、楽曲の内容に関する特徴の解析を必要としない利点があるが、新規ユーザや未評価の楽曲に対応しづらい問題 [3] などがある。それに対して、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 [6, 14, 23] である。これは再生数に依存しないことから、前述した問題は起きにくく、PrefTolerance もこちらに位置づけられる。しかし、従来の内容に基づくフィルタリングでは楽曲の特徴全体に対するユーザの嗜好しか反映できなかった。PrefTolerance は、2次元空間上で楽曲間の類似度を可視化でき、かつ音楽に詳しくないユーザであっても比較的判別しやすい歌声・伴奏・歌詞についての音楽要素ごとに嗜好を反映できる利点を持つ。

2.2 楽曲間の類似度とユーザの嗜好の可視化

楽曲間の類似度を俯瞰して把握できる可視化として、Hamasaki ら [5] 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。Vad ら [22] 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、楽曲のム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。Sasaki ら [19] 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によって可視化し探索できる LyricsRadar を提案した。これらの研究は、特定の音楽要素に対する可視化を実現しているが、PrefTolerance では、音楽要素ごとの類似度をそれぞれ 2次元空間で可視化できる点に特徴がある。

一方、楽曲に対するユーザの嗜好を可視化する研究として、Saito ら [17] 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次元空間に配置し、ユーザが「好き」または「好まない」とした楽曲を異なる色で可視化する MusiCube を提案した。この研究では、ユーザの好みを空間上の点の色と位置で把握できるが、より細かな好みの分布は明らかではない。これに対して PrefTolerance では、楽曲の評価と許容度に応じて推定される嗜好をグラデーションのように可視化することで、ユーザは好みの強弱や広がりを感じて把握できる。

2.3 許容度の指定

許容度と似た概念を扱う研究として、Bostandjiev ら [2] は、アーティストとウェブ情報に対して、ユーザがどの程度重視したいかをスライダーで入力し、アーティストを推薦する TasteWeights を提案した。Millecamp ら [12] 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ー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。これらの研究は、ユーザが楽曲の好みを判断する際に「歌詞の重要度が高い」といったように、音楽要素自体をどれぐらい重視するかの度合いを指定できるが、歌詞を重視しつつも「この歌詞に非常に似た歌詞だけを聴きたい」や「この歌詞とある程度似ている歌詞ならばよい」のような類似度の範囲指定には対応できない。それに対して PrefTolerance では、「音楽要素ごとにどの程度の類似度まで許容できるかの度合い」を表す許容度を用いるため、そうした場合にも対応できる。このように、音楽要素に対する重要度と、類似度に対する許容度は異なる概念である。

3 PrefTolerance

PrefTolerance のインタフェースを図 2 と図 3 に示す。図 2 は、使用開始直後にユーザが 1 曲目に聴く楽曲を選択する画面である。その後、ユーザが楽曲の聴取・評価・許容度の指定を行う図 3 の画面に変わる。これは、(1) 音楽要素ごとの空間、(2) 許容度スライダー、(3) 動画再生プレーヤ、(4) 評価エリア、(5) 探索・推薦タブの 5 つの機能で構成されている。

PrefTolerance は音源を入手可能な歌詞を伴う任意の楽曲が対象だが、今回の実装では、ニコニコ動画¹に投稿された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 曲とした。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、Songle Widget [4] を用いた、楽曲のサビ区間 (図 3 の動画再生プレーヤにおける下部のオレンジ色の四角) の試聴再生が容易にできるためである。

¹ <https://www.nicovideo.jp/>



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)。

3.1 操作手順

PrefTolerance 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、各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。
 - (a) 探索タブを押すと、ユーザが次に評価する楽曲がタブ内に 1 曲提示され、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。(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る楽曲が狭い範囲のみに偏った状態でユーザが繰り返し聴取・評価することを防ぐためである。適度に探索タブも利用することで、より適切に嗜好を反映した多様な楽曲

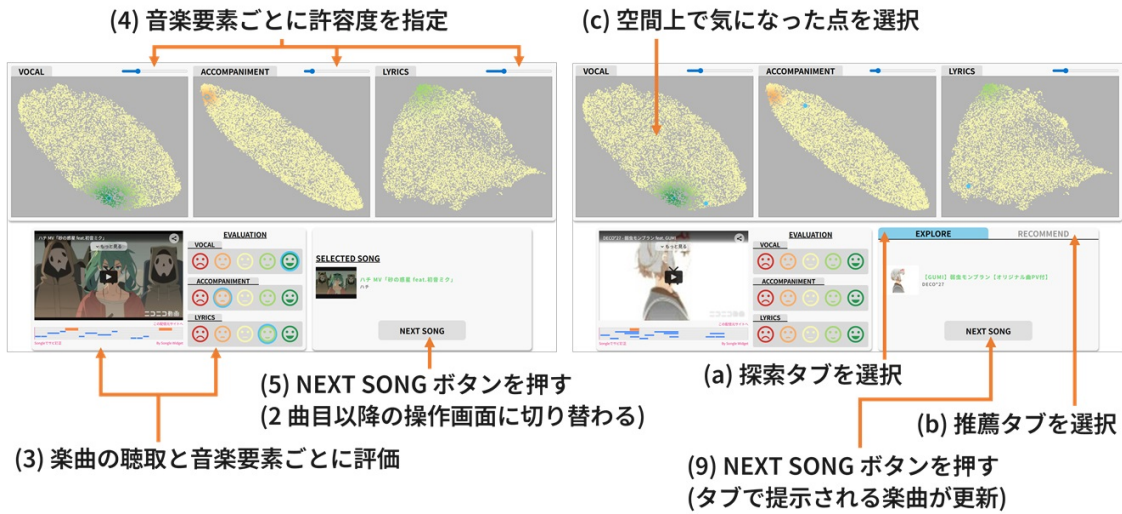


図 4. 1 曲目の操作画面 (左) と, 2 曲目以降の操作画面 (右) における操作手順.

を推薦することが可能となる.

3.2 音楽要素ごとの空間

図 3 の画面上部における, 歌声 (VOCAL)・伴奏 (ACCOMPANIMENT)・歌詞 (LYRICS) の 3 つの音楽要素の空間は以下の方法で作成した. 歌声と伴奏に関しては, 既存の音源分離技術 HT Demucs [16] を用いて各楽曲を歌声と伴奏に分離した. その後, 自己教師あり学習で事前学習済みの音楽特徴抽出モデル MERT [7] 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, 事前学習済みの文埋め込みモデル Sentence-BERT [15] を用いて, 各楽曲の歌詞から 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, 既存の非線形次元圧縮技術 UMAP [11] 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 [24] 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長がある. 例えば, 歌詞の許容度が高いユーザーの場合, 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て採用した. この可視化により, ユーザは音楽要素ごとに, 楽曲同士の類似度を俯瞰して把握できる.

3.3 評価エリア

PrefTolerance の図 3 における評価エリアでは, 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うかを入力する. 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

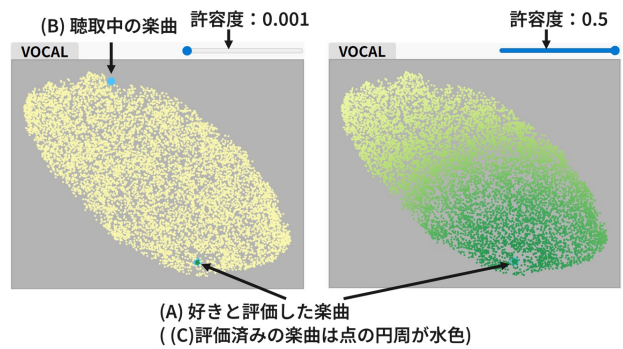


図 5. 歌声空間における評価値の例 (左図が許容度が 0.001 の場合, 右図が許容度が 0.5 の場合).

ト尺度を用いた. 3 段階では粗すぎ, 7 段階では多数の楽曲を評価する際に負担を感じるので 5 段階とした. ユーザが評価を直感的に入力できるように, フェイス・スケール [10] を用いた.

ユーザーがある楽曲のある音楽要素に対して評価すると即座に, 図 5 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 5 (A)) に変化する. これにより, ユーザは評価した楽曲が各空間上のどの位置にあり, どのような評価となったかを視覚的に確認できる. また, 評価前で聴取中の楽曲の点を水色 (B), 評価済みの楽曲は点の円周を水色 (C) で表現する.

3.4 許容度スライダを用いた音楽要素ごとの空間とのインタラクション

ユーザーは, 図 3 の許容度スライダを左右に動かして各空間に対する許容度を指定できる. 各空間にお

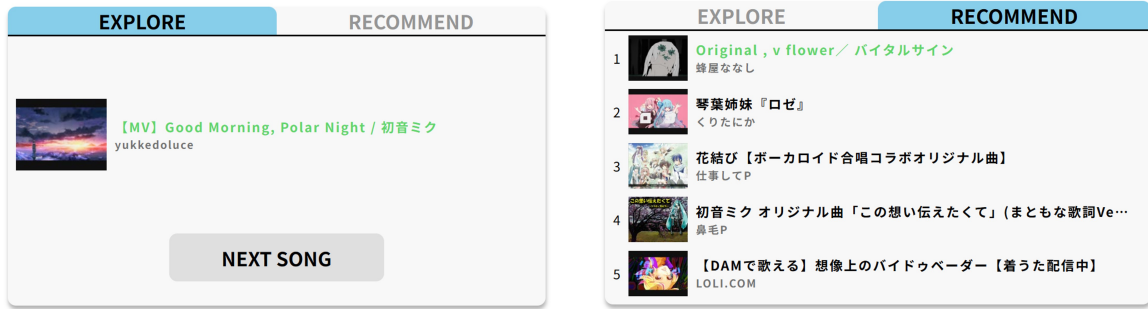


図 6. 探索結果を提示する画面 (左, 探索タブ) と推薦結果を提示する画面 (右, 推薦タブ). タブ内には楽曲のサムネイル画像と楽曲名, クリエイター名が表示され, 選択中の楽曲は楽曲名が緑色で表示される.

ける許容度および評価済み楽曲の位置と評価値に基づき, 未評価楽曲の点の色がグラデーションのように変化する. このグラデーションは, 未評価楽曲の各音楽要素に対する評価の推定値から決められる. 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, 未評価の楽曲 i , 評価済み楽曲 j , 歌声を V , 伴奏を A , 歌詞を L とし, 音楽要素の集合を $E = \{V, A, L\}$ としたとき, 音楽要素 $e \in E$ における i の評価の推定値 Z_{e_i} を次式とする.

$$Z_{e_i} = \frac{w_{e_j}}{2\pi\sigma_{e_j}^2} \exp\left(-\frac{(x_{e_i}-X_{e_j})^2+(y_{e_i}-Y_{e_j})^2}{2\sigma_{e_j}^2}\right) \quad (1)$$

ここで, 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 (X, 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. 2次元ガウス分布は共分散行列の非対角成分が 0 で, 対角成分が同一の値 σ^2 とした. 許容度は, $0.001 \sim 0.5$ 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} の値を加算した合計値を用いた. 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 5 の左, 許容度が 0.5 の場合を図 5 の右に示す. 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 (許容度が大きい右図の方が, より遠くまで影響を与える). この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.

3.5 探索・推薦タブ

ユーザがまだ評価していない領域から繰り返し点を選んで評価するのは煩雑なため, もっと効果的に探索できるように「探索」タブ機能を設けた. 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, 本論文では「インターフェースがユーザに評価してほしい楽曲を探索する」(次に評価すると効果的な楽曲をインターフェースがユーザに提示する)ことを意味する. ユーザが図 3 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 6: 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) 楽曲を候補とする. 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 (本論文では $\tau = 0.3$ 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形に更新され, より的確な推薦が可能となる.

一方, 「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 6: 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.

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

推薦タブ上で R_i が高い順に, 1 位から 5 位の上位 5 曲を推薦楽曲として提示する. 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, 嗜好の度合いが高い音楽要素のう

ち、許容度の低さ（こだわりの強さ）を重視した楽曲推薦を受けることができる。また、提示される楽曲は、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。

4 ケーススタディ

本章では、PrefToleranceの体験を探るために、男性の学生2名（参加者A、Bとする）と1対1の対面形式によるケーススタディを行った。参加者はまず、インター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ーフェースを使用した。インター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、インターフェースの探索・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インターフェースの使用終了時に回答してもらうためである。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、参加者の嗜好を反映して使用を開始してもらうためである。また、楽曲の再生位置については指定せず、参加者が楽曲の先頭やサビなど任意の位置から再生して聴取できるようにした。インター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。インターフェースの使用後、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のそれぞれに対して7段階（1～7で大きいほど満足度が高い）で回答する項目と、インターフェースを使用した感想を述べる自由記述欄の構成とした。

ケーススタディを行った結果、インターフェースの使用時間は参加者Aは26分、参加者Bは30分だった。評価した楽曲数は、参加者Aは18曲、参加者Bは20曲だった。音楽要素ごとの評価に対する満足度は、参加者Aは6、参加者Bは5と高かった。最終的な推薦結果に対する満足度も参加者Aは6、参加者Bは4と高かった。自由記述欄では参加者A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」という回答が得られた。

参加者Aがインターフェースの使用を終えたときのそれぞれ空間を図7に示す。図7について、それぞれの空間上の点の色で「好き」に該当する緑色の点の分布が多く見られたが、歌声空間と歌詞空間では「少し嫌い」に該当する点の分布が見られた。参加者Aのインターフェースの使用を終えるまでの許容度の推移を図8に示す。図8について、3曲目までには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、歌詞の許容度は0.050から0.213と変化して

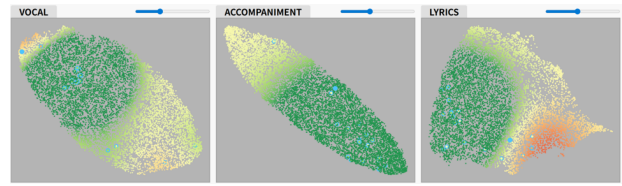


図7. 参加者AがPrefToleranceの使用を終えたときのそれぞれの空間。

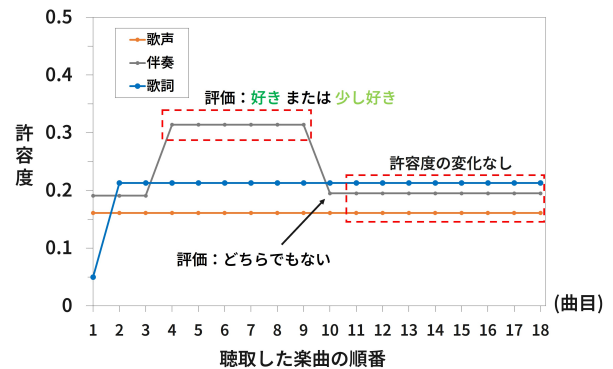


図8. 参加者Aにおける許容度の推移。

いた。10曲目の伴奏の許容度は、0.314から0.195に低下しており、このときの伴奏の評価は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」であった。11曲目以降の許容度は、各音楽要素で変化していなかった。

5 おわりに

本論文の主な貢献を以下に述べる。

- ユーザが歌声・伴奏・歌詞の音楽要素ごとの嗜好と、各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」を指定できる楽曲推薦インターフェース PrefTolerance を提案した。
- ユーザが満足する許容度を指定可能にするため、音楽要素ごとの空間の可視化と、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応じて、好みに合う楽曲の範囲がどのように変化するかを点の色のグラデーションで表現するインタラクションを実現した。
- ユーザの音楽要素ごとのこだわりを重視することができる推薦機能と、効率よくユーザの嗜好を反映することができる探索機能を実現した。

今後は、ユーザの許容度を自動的に推定する機能を設けることで、許容度の指定が得意でないユーザでも扱えるインターフェースの実現を目指していきたい。また、許容度を用いない手法との比較実験を行い、PrefToleranceの有効性を調査していきたい。

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] B. Amiri, N. Shahverdi, A. Haddadi, and Y. Ghahremani. Beyond the Trends: Evolution and Future Directions in Music Recommender Systems Research. *IEEE Access*, 12:51500–51522, 2024.
- [2] S. Bostandjiev, J. O’Donovan, and T. Höllerer. TasteWeights: a visual interactive hybrid recommender system. In *Proceedings of the 6th ACM Conference on Recommender Systems (ACM RecSys 2012)*, pp. 35–42, 2012.
- [3] Y. Deldjoo, M. Schedl, and P. Knees. Content-driven music recommendation: Evolution, state of the art, and challenges. *Computer Science Review*, 51:100618, 2024.
- [4] M. Goto, K. Yoshii, and T. Nakano. Songle Widget: Making animation and physical devices synchronized with music videos on the web. In *Proceedings of the 2015 IEEE International Symposium on Multimedia (IEEE ISM 2015)*, pp. 85–88, 2015.
- [5] M. Hamasaki, M. Goto, and T. Nakano. Songrium: Browsing and listening environment for music content creation community. In *Proceedings of the 12th Sound and Music Computing Conference (SMC 2015)*, pp. 23–30, 2015.
- [6] D. Kostrzewa, J. Chrobak, and R. Brzeski. Attributes Relevance in Content-Based Music Recommendation System. *Applied Sciences*, 14(2), 2024.
- [7] Y. Li, R. Yuan, G. Zhang, Y. Ma, X. Chen, H. Yin, C. Xiao, C. Lin, A. Ragni, E. Benetos, N. Gyenge, R. Dannenberg, R. Liu, W. Chen, G. Xia, Y. Shi, W. Huang, Z. Wang, Y. Guo, and J. Fu. MERT: Acoustic Music Understanding Model with Large-Scale Self-supervised Training. In *The Twelfth International Conference on Learning Representations (ICLR 2024)*, 2024.
- [8] X. Liu, Z. Yang, and J. Cheng. Music recommendation algorithms based on knowledge graph and multi-task feature learning. *Scientific Reports*, 14(1):2055, 2024.
- [9] B. Logan. Music Recommendation from Song Sets. In *Proceedings of the 5th International Society for Music Information Retrieval Conference (ISMIR 2004)*, pp. 425–428, 2004.
- [10] C. D. Lorish and R. Maisiak. The face scale: A brief, nonverbal method for assessing patient mood. *Arthritis & Rheumatism*, 29(7):906–909, 1986.
- [11] L. McInnes, J. Healy, and J. Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv preprint arXiv:1802.03426*, 2018.
- [12] M. Millecamp, N. N. Htun, Y. Jin, and K. Verbert. Controlling Spotify Recommendations: Effects of Personal Characteristics on Music Recommender User Interfaces. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization (ACM UMAP 2018)*, pp. 101–109, 2018.
- [13] T. Nakano, K. Yoshii, and M. Goto. Musical Similarity and Commonness Estimation Based on Probabilistic Generative Models of Musical Elements. *International Journal of Semantic Computing (IJSC)*, 10(01):27–52, 2016.
- [14] M. Pulis and J. Bajada. Siamese Neural Networks for Content-based Cold-Start Music Recommendation. In *Proceedings of the 15th ACM Conference on Recommender Systems (ACM RecSys 2021)*, pp. 719–723, 2021.
- [15] N. Reimers and I. Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *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.
- [16] S. Rouard, F. Massa, and A. Défossez. Hybrid Transformers for Music Source Separation. In *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP 2023)*, 2023.
- [17] Y. Saito and T. Itoh. MusiCube: A Visual Music Recommendation System featuring Interactive Evolutionary Computing. In *Proceedings of the 2011 Visual Information Communication-International Symposium (ACM VINCI)*, pp. 1–6, 2011.
- [18] D. Sánchez-Moreno, A. B. G. González, M. D. M. Vicente, V. F. L. Batista, and M. N. M. García. A collaborative filtering method for music recommendation using playing coefficients for artists and users. *Expert Systems with Applications*, 66:234–244, 2016.
- [19] S. Sasaki, K. Yoshii, T. Nakano, M. Goto, and S. Morishima. LyricsRadar: A Lyrics Retrieval System Based on Latent Topics of Lyrics. In *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, pp. 585–590, 2014.
- [20] M. Schedl, P. Knees, B. McFee, and D. Bogdanov. Music Recommendation Systems: Techniques, Use Cases, and Challenges. In *Recommender Systems Handbook*, pp. 927–971. Springer, 2022.
- [21] K. Tsukuda, T. Nakano, M. Hamasaki, and M. Goto. Unveiling the Impact of Musical Factors in Judging a Song on First Listen: Insights From a User Survey. In *Proceedings of the 24th International Society for Music Information Retrieval Conference (ISMIR 2023)*, pp. 561–570, 2023.
- [22] B. Vad, D. Boland, J. Williamson, R. Murray-Smith, and P. B. Steffensen. Design and Evaluation of a Probabilistic Music Projection In-

- terface. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015)*, pp. 134–140, 2015.
- [23] A. van den Oord, S. Dieleman, and B. Schrauwen. Deep content-based music recommendation. In *Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS 2013)*, pp. 2643–2651, 2013.
- [24] L. Van der Maaten and G. Hinton. Visualizing Data using t-SNE. *Journal of machine learning research (JMLR)*, 9(11):2579–2605, 2008.
- [25] 奥 健太. 楽曲推薦システム—プレイリスト, コンテキスト, インタラクション—. *人工知能学会誌*, 34(3):300–308, 2019.
- [26] 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*, 50(8):751–755, 2009.

未来ビジョン

本研究が目指す未来は大きく2つある。

まず、音楽推薦において「許容度」という新たな概念を人々に浸透させていくことで、単に「好き」や「嫌い」といった評価を伝え合うだけでなく、音楽要素ごとの嗜好と許容度を可視化した2次元空間を見せ合う形で、互いの好みや許容度を直感的に共有するようなコミュニケーションが可能になることを期待している。この空間を互いに見せ合うことにより、「あなたは歌声に対してこだわりが強い」、「あなたと私で伴奏の好みが全然違う」、「あなたと比べて私は歌詞に対して全然こだわっていない」といった違いを感覚的に理解しやすくなる。さらに、互いの嗜好や許容度を比較・探索するようなインタラクションも可能になるので、例えば自分と近い聴き方をする人を見つけてその人のイチ推しの楽曲を聴かせてもらうような

音楽との出会いも実現できる。

次に、許容度の概念は音楽に限らず様々な可能性を持っており、他のメディアコンテンツや日常生活にも応用されていくことを期待している。例えばメディアコンテンツでは、アニメやドラマなどの映像作品の推薦システムを実現する際にも、物語や映像表現、演出のトーンなどの特徴に対して、ユーザが自分の嗜好と許容度を設定できるような機能が考えられる。日常生活でも、「食の辛さや甘さなどに対する嗜好と許容度」や「服の色や素材、デザインなどに対する嗜好と許容度」など、多様な場面で好みと許容度を探索・可視化できるような発展が考えられる。

このように本研究をきっかけに「許容度」の概念が広まることで、音楽のみならず日常生活の様々な体験において、嗜好や許容度が可視化されて活用されやすくなり、相互理解がより円滑になる未来が切り拓ける。