感情フィードバックを用いた オンライン自動楽曲ミキシングシステム

高橋 達哉[1,a)]　佐藤 俊樹[1,b)]

概要：本稿では，楽曲を選び，ミックスして観客を盛り上げる DJ プレイの自動化手法を検討する．DJ が楽曲を選ぶように複数の曲を推薦する研究や，ユーザーの不快感を減らすようにテンポを合わせながらミックスする研究，ユーザーのフィードバックを可視化して DJ プレイに反映する研究が行われてきた．しかし，現実の DJ の多くはユーザーからのフィードバックを受けてプレイ中，動的に楽曲を変更することがあることに対し，システム側はフィードバックを動的に活用できていなかったり，或いはユーザーからの能動的な動作を必要としている．そこで本研究では，web カメラを用いてユーザーの感情を認識し，動的に楽曲を変更しユーザーの幸福度をあげる方向に最適化するオンラインシステムを提案する．

1. はじめに

我々は普段，街中に流れる BGM や，サブスクリプション型の音楽サービスなど，様々な形で音楽を享受している．それらのほとんどは，予め決められたプレイリストから，決められた順番，或いはランダムに再生される．気分によって最適な楽曲に変更したり，楽曲間を滑らかにつなぐ（ミキシング）行為には幅広い知識と技術が要求される．

プレイリストの作成と，次にかける曲を選ぶ専門的な職業に DJ があり，また，DJ のプレイを自動化しようという研究がある．DJ のプレイが自動化することにより，我々はいつでもどこでも DJ プレイを体験することができる．自動化された DJ は，クラブで DJ の代わりにプレイすることにより，システムであることの利点を生かしてインターネット経由で巨大な楽曲群から最適な曲を選択することができる．また，自動化された DJ は大きな機材やプレイヤーとしての人が必要ないため，店舗やオフィスなどでも 24 時間常に DJ がミックスするような音楽体験を提供することができる．つまり，DJ の自動化には我々の音楽鑑賞体験をより向上させることを期待できる．

2. 関連研究

DJ プレイは大きく分けて，プレイリスト作成，プレイ中の選曲，曲間の滑らかな（或いは扇情的な）ミキシングの 3 つがある．ガウス過程回帰を用いてユーザーの嗜好を学習してプレイリストを作成する研究 [5] や，楽曲を 5 秒間の小区間に分割し，各小区間ごとの類似度をクロマワードを用いたトピックモデルで算出し，類似度の高い楽曲をミキシングしていく研究 [1]，ユーザーが不快と感じるのはミックスの手法ではなくテンポの過度な変更であることを調査で明らかにし，テンポの変更度合いとユーザーの不快度を関数として定義した上で，不快度が最小になるように最適化する手法についての研究 [2] などがある．

一方，Shamma ら [7] が行ったオンラインで活躍する複数の DJ に対する調査によれば，DJ は観客とのコミュニケーションを重要視しており，DJ プレイ中に観客の盛り上がり度合いを観察し，動的に次にかける曲を変更することがあることを明らかにした．

ユーザーからのフィードバックを取り入れる研究として，評価ボタンを押下することにより重み付けを変更する研究 [1] や，ユーザーからの評価やコメントを画面に表示して，リアルタイムにプレイを行うインタラクティブなインターフェースを開発した研究 [4] がある．しかし，いずれもユーザーからの能動的な動作を要求するため，本来の DJ パフォーマンスの鑑賞体験とは異なるものとなっている．ここでの本来の DJ パフォーマンスとは，DJ がユーザーの顔や動きなどから，盛り上がり具合を注意深く観察し，プレイに反映させていくことである．そこで本研究では，ユーザーは生体センサーの装着やコメントを入力することなく，カメラを用いてユーザーの感情を認識し，取得した情報をパラメーターとして，ユーザーの幸福度を最大化するように選曲を最適化していく手法を提案する．

[1] 北陸先端科学技術大学院大学
[a)] s2010103@jaist.ac.jp
[b)] tsato@jaist.ac.jp

© 2021 Information Processing Society of Japan

3. 提案

3.1 提案概要

本稿では，web カメラ越しにユーザーの感情を認識し，幸福度が高まるように最適化を行い，自動で選曲，ミキシングして音楽を配信し続けるオンラインインターフェースを提案する．提案手法では，初期値として楽曲間の類似度マトリクスを作成し，ユーザーのフィードバックにより類似度のバイアス項を更新することで最適化を図る．

3.2 プレイリスト

プレイリストに入れる楽曲は SoundCloud

`https://soundcloud.com/`

から Creative Commons の楽曲を 5 つのジャンル（Chill, EDM, FutureBass, Healing, Lofi-Hiphop）から 20 曲ずつ，計 100 曲を収集した．DJ のミックスにおいて，ミックスする 2 曲のテンポを合わせることは，ユーザーの快適度を保つための重要な要素の一つである．他方で，テンポを過度に変更することはユーザーの不快度を増すという先行研究の結果もある．本研究においては，ユーザーの感情がテンポの変更に左右されることを避けるため，テンポの変更は行っていない．

3.3 楽曲の類似度

各楽曲を先頭から 10 秒のフラグメント T_{head} と，20 秒目から 10 秒のフラグメント T_{tail} の 2 つに分割し，分割したフラグメントの Log-melspectrogram を入力とし，楽曲のジャンルを推定するように Convolutional Neural Network(CNN) を用いて学習を行う．図 1 にジャンルが lofi-hiphop の楽曲のうちの一つの Log-melspectrogram を示す．

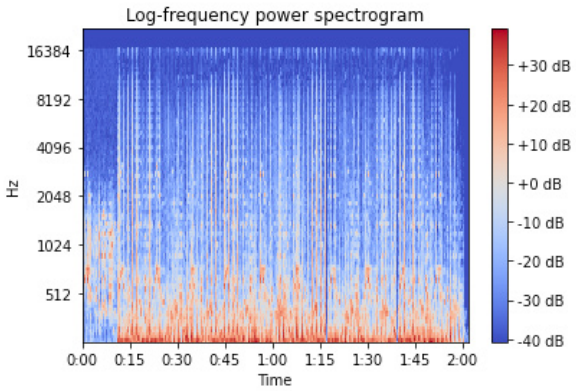

図 1 ジャンルが lofi-hiphop の楽曲の Log-melspectrogram

学習済みモデルの Forward 時の最終層の出力を楽曲の特徴量とし，コサイン類似度で各フラグメントの類似度を算出する．これは，楽曲のジャンルを分けるための特徴量は，楽曲そのものの特性をよく表しているだろうというアイディアに基づいている [6]．分類モデルに CNN を選択したのは，周辺情報を畳み込むことにより時間，周波数に対して頑健に識別できるためである．楽曲間のコサイン類似度のマトリクスをヒートマップで可視化したものが図 2 である．この図の縦軸は T_{tail_t}，横軸は T_{head_t}（$0 < t \leq 100$）であり，それぞれが同じ順番に並んでいる．つまり 1 行目の 2 列目は，T_{tail_1} と T_{head_2} のコサイン類似度を表す．この図より，類似度が高い 5 つの正方形の領域は，20 曲ずつの各ジャンルの類似度を示していることから，類似度マトリクスの初期値においては同一ジャンル内の楽曲の特徴ベクトルはコサイン類似度も高いといえる．

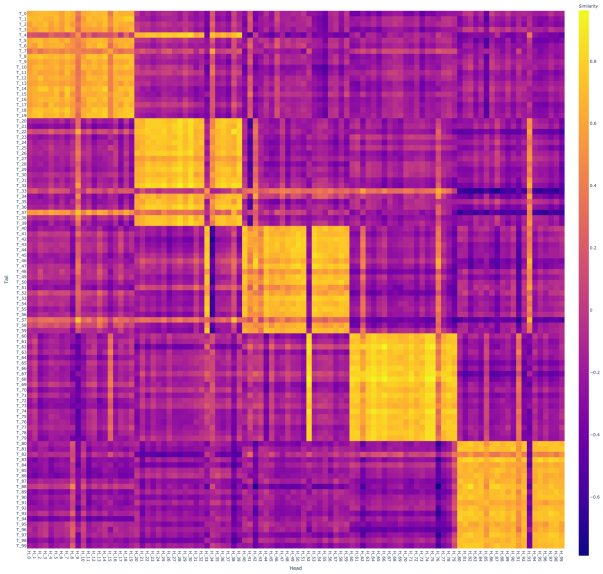

図 2 類似度マトリクスの可視化

3.4 選曲方法

一回の DJ プレイ中にプレイする曲数を N，i 番目に再生中の楽曲の先頭のフラグメントを T^i_{head}，末尾のフラグメントを T^i_{tail} とする（$0 < i \leq N$）．i 番目の楽曲から i+1 番目の楽曲選ぶ際は，T^i_{tail} と T^{i+1}_{head} のコサイン類似度とバイアス項を加算した値が一番大きいものを選ぶ．なお，類似度マトリクス作成時は，全てのバイアス項は 0 である．バイアスについては，類似度マトリクスと同じサイズの行列であるバイアスマトリクスとして初期化する．

3.5 感情認識システム

ユーザーの感情認識には Microsoft のクラウドサービスである Azure Face API を用いる．Khanal ら [3] が，Azure Face API は他のクラウド感情認識サービスと比較し，真正面の顔に対しては一番精度高く認識できていることを示している．Azure Face API では，anger, contempt, disgust,

© 2021 Information Processing Society of Japan

fear, happiness, neutral, sadness, surprise の 8 種類の感情を全ての値を足すと 1 になる値で取得することができる．この Azure Face API を用いて，web カメラで写している人の顔から感情を認識してデータベースに蓄積する web アプリケーションを構築した（図 3）．このシステムでは利用時に名前を入力し，利用者ごとの感情がデータとして蓄積される．

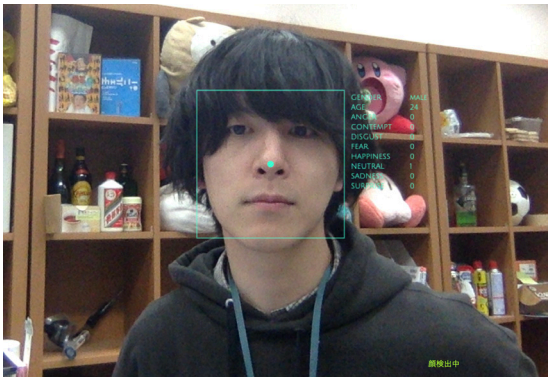

図 3　感情認識システムにて顔から感情を認識している様子

3.6　バイアスの更新

感情は 1 秒おきに取得し，i 番目の楽曲の再生開始後 10 秒内に取得した感情値を平均し，happiness が negative な感情（anger, contempt, disgust, fear, sadness のいずれかの最大値）より大きい場合，T_{tail}^{i+1} に係る全ての T_{head}，及び T_{head}^{i+1} に係る全ての T_{tail} の類似度に対し，それぞれのコサイン類似度と happiness 値の積で，対応するバイアスマトリクスの要素に加える．

逆に，negative な感情が happiness より大きい場合，ユーザーは不快であると考え，コサイン類似度とバイアス項を加算した値が一番小さい T_{head} を選曲する．これは，類似している曲を選択してもユーザーの幸福度をあげることはできないため，最も似ていない曲を選ぶことで改善を期待することを意味する．そして加算の時と同様に全ての関係する楽曲の類似度マトリクスを減算する．このように類似度マトリクスを更新して，ユーザーの感情に応じてバイアスの更新を行う．

4.　実験・評価

提案手法が，ユーザーにとって新しく気分を向上させるインターフェースであるか，また，楽曲が最適化されていくことを感じることができるかを明らかにするための実験を行った．実験では，複数人に対してオンライン上で実施した．各自 PC で感情認識サイトに接続してもらいながら web カメラを ON にし，類似度マトリクスを最適化しない手法（従来的手法）と，随時最適化する手法（以下，提案手法）で両方を聴取させた．画面には図 4 に示すように，再生中の楽曲，次の候補として示されている楽曲，現在視聴している被験者の名前とその感情をリアルタイムにプロットしている．音声及び映像の配信は，全て YouTube Live で行った．

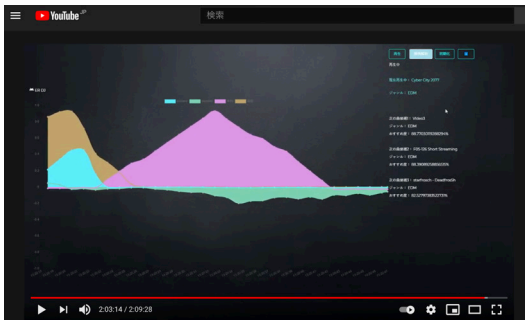

図 4　実験時の配信画面

実験開始時に被験者に説明を行ったのちに，あらかじめ決められた 1 曲目をかける．最適化しない手法では，そのままコサイン類似度に従い一番類似しているものを選択し，楽曲の 20 秒目から 30 秒目にかけて次の曲にクロスフェードしてミキシングしていく．提案手法では，曲が変更された後 10 秒間の感情の平均を用いて，バイアスマトリクスの更新を行う．この場合の感情の平均とは，全被験者の 10 秒間の平均である．なお実験中は，両手法とも同じ曲を 2 回かけることはしないという制約を設けた．

評価方法として，聴取後にアンケートを実施した．最適化しない手法と提案手法を聞き比べた結果の，定性的な評価を行える入力項目を用意した．

5.　結果

図 5 は，更新後のバイアスマトリクスをヒートマップとして可視化したものである．この図の縦軸と横軸及び示す色は，図 2 と一緒である．

この図から，EDM の部分が正のバイアスが強く，逆に healing や lofihiphop の負のバイアスが強いことが見て取れる．被験者は，落ち着いたビートよりも，アップテンポなビートを求めている傾向があることがわかる．

今回は，現実でもオンラインでも DJ のプレイを鑑賞をしたことがある 4 名を被験者とした．アンケートでは，提案手法の方が，体験としての未知さ，楽しさについては既存手法を大きく上回る結果となった．さらに，作業用 BGM としての良さ，フィードバックの手軽さについても良い評価を得た．提案手法が感情に合わせて楽曲が最適化されているように感じるかについては，様々な意見が出た．そもそも実験時間が短いために好みの通りに最適化されたかどうか判断するに至らなかったという意見や，lofihiphop ジャンルの時に退屈な表情をすることにより，アップテンポな EDM に変更された時にとても感情の高まりを感じたという意見があった．また，気に入ってる曲はより長く，気に

© 2021 Information Processing Society of Japan

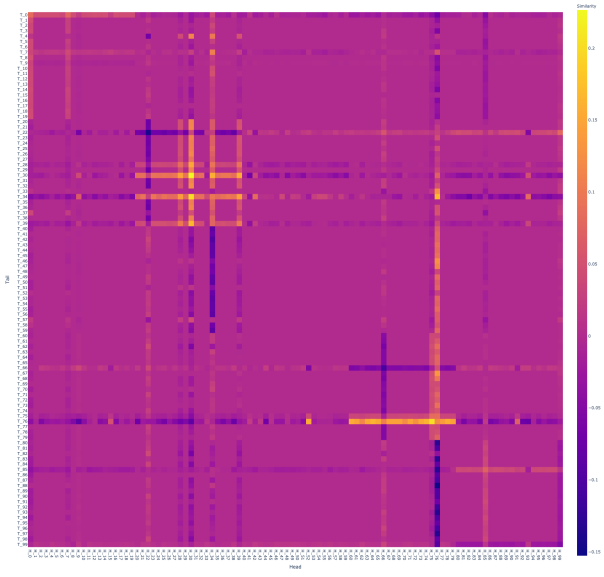

図 5 実験後のバイアスマトリクスの可視化

入らない曲はより短くした方がいいという要望もあった．

今後の課題としては，より長期的な最適化の方策を決めることや，大衆の平均ではなく個人のパーソナリティによる最適化，また，はじめにユーザーの好みがわからない状態でのコールドスタート問題をどのように解決するか，などを検討する必要がある．

とはいえ，ユーザーの感情に合わせて楽曲を更新することは，ユーザーの関心を引きつけ，また作業を中断せずに音楽体験の質を向上させる新たなインターフェースとして有効である可能性を見ることができた．

6. アプリケーション

この感情認識をする DJ のアプリケーション例として，車のオーディオシステムが考えられる．運転中は両手が塞がった状態であり，自分自身で楽曲を選ぶことは注意を散漫にさせる恐れがあるが，運転者の表情から感情を読み取り最適化することにより安全にその時の気分に合わせた楽曲を選ぶことができる．

それ以外にも，自宅での作業中の BGM として，無自覚的な感情を読み取り最適化することはもちろん，自覚的に顔の表情だけでフィードバックを送ることで，作業の集中力を妨げることなく最適な音楽鑑賞をすることができる．

7. おわりに

本研究ではユーザーの感情をフィードバックとした選曲最適化システムを提案した．このシステムではカメラに顔が映ってさえいればユーザーの能動的なフィードバックを必要としないため，作業中などでも集中力を害することなく利用することができるという利点がある．今後の展望として，我々の音楽体験が全く新しいものへと変革していくために，今は顔から感情認識というパラメータしか用いていないが，他にもユーザーの場所的情報や，性別，年齢などの多くの特徴量を用いたり，選曲だけではなく BPM の変更度合い，ミックスの仕方，エフェクトのかけ方なども最適化し，より一層表現力豊かな自動 DJ システムを作成することを検討する．

参考文献

[1] Hirai, T., Doi, H. and Morishima, S.: Latent Topic Similarity for Music Retrieval and Its Application to a System that Supports DJ Performance, *Journal of Information Processing*, Vol. 26, pp. 276–284 (online), DOI: 10.2197/ipsjjip.26.276 (2018).

[2] ISHIZAKI, H.: Full-automatic DJ Mixing System with Optimal Tempo Adjustment Based on Measurement Function of User Discomfort, *Proc. of ISMIR2009*, pp. 135–140 (online), available from ⟨https://ci.nii.ac.jp/naid/10028092808/⟩ (2009).

[3] Khanal, S. R., Barroso, J. a., Lopes, N., Sampaio, J. and Filipe, V.: Performance Analysis of Microsoft's and Google's Emotion Recognition API Using Pose-Invariant Faces, *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion*, DSAI 2018, New York, NY, USA, Association for Computing Machinery, p. 172–178 (online), DOI: 10.1145/3218585.3224223 (2018).

[4] Laursen, L. F., Goto, M. and Igarashi, T.: A Multi-Touch DJ Interface with Remote Audience Feedback, *Proceedings of the 22nd ACM International Conference on Multimedia*, MM '14, New York, NY, USA, Association for Computing Machinery, p. 1225–1228 (online), DOI: 10.1145/2647868.2655065 (2014).

[5] Platt, J. C., Burges, C. J. C., Swenson, S., Weare, C. and Zheng, A.: Learning a Gaussian Process Prior for Automatically Generating Music Playlists, *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, NIPS'01, Cambridge, MA, USA, MIT Press, p. 1425–1432 (2001).

[6] Qosmo: AI DJ Project — A dialogue between AI and a human, `http://qosmo.jp/projects/ai-dj-human-dj-b2b/`.

[7] Shamma, D. A., Churchill, E. F., Bobb, N. and Fukuda, M.: Spinning Online: A Case Study of Internet Broadcasting by DJs, *Proceedings of the Fourth International Conference on Communities and Technologies*, C&T '09, New York, NY, USA, Association for Computing Machinery, p. 175–184 (online), DOI: 10.1145/1556460.1556486 (2009).

© 2021 Information Processing Society of Japan