

音楽を用いた感情誘導のための脳波からの感情予測

宮本 佳奈^{†,††} 田中 宏季^{†,††} 中村 哲^{†,††}

[†] 奈良先端科学技術大学院大学 〒 630-0192 奈良県生駒市高山町 8916-5

^{††} 理化学研究所 革新知能統合研究センター 〒 630-0192 奈良県生駒市高山町 8916-5

E-mail: [†]{miyamoto.kana.mk4,hiroki-tan,s-nakamura}@is.naist.jp

あらまし 日常生活において感情の誘導が必要な場面が存在する。音楽は感情を変化させる刺激の一つである。しかし、音楽聴取時の感情には個人差があるため、感情誘導には個人に適応した音楽を利用することが望まれる。そこで我々は、脳波から感情を予測し、個人の状態を考慮して音楽を生成する感情誘導システムを提案した。本稿では、脳波を用いた感情予測に Convolutional Neural Network (CNN) を利用することを提案する。モデルの予測値と実測値の RMSE を計算した結果、CNN を用いたモデルはベースラインモデルの線形回帰より RMSE が有意に低くなった。CNN を利用した脳波からの感情予測は感情誘導システムの構築に役立つと考えられる。

キーワード 脳波, 感情予測, 感情誘導

EEG-based Emotion Prediction for Emotion Induction using Music

Kana MIYAMOTO^{†,††}, Hiroki TANAKA^{†,††}, and Satoshi NAKAMURA^{†,††}

[†] Nara Institute of Science and Technology 8916-5 Takayamacho, Ikoma, Nara, 630-0192 Japan

^{††} Center for Advanced Intelligence Project, RIKEN 8916-5 Takayamacho, Ikoma, Nara, 630-0192 Japan

E-mail: [†]{miyamoto.kana.mk4,hiroki-tan,s-nakamura}@is.naist.jp

Abstract There are situations where we need to induce emotions in our daily life. Music is one of stimuli that can change emotions. Since there are individual differences in emotions they feel, it is important to use customized music. Therefore, we proposed an emotion induction system that predicts emotions using electroencephalogram (EEG) and generates music based on the emotions. In this paper, we propose using convolutional neural network (CNN) for emotion prediction from EEG. As a result of calculating RMSE between the predicted and measured values of the models, the CNN-based models showed significantly lower RMSE than a linear regression model. The emotion prediction from EEG using CNN is expected to be useful for constructing the emotion induction system.

Key words Electroencephalogram, emotion prediction, emotion induction

1. はじめに

感情を適切に誘導することはメンタルヘルスの向上に有効である [1]~[3]。日常生活ではさまざまな感情が誘発され、それらの感情にはそれぞれ役割がある [4], [5]。ポジティブな感情は思考と行動の範囲を広げるため、全体的な認知や処理能力を高める効果がある。ネガティブな感情は思考と行動の範囲を狭めるため、局所的な認知や処理能力を高める効果がある。このことから場面に適した感情を誘導することが注目されている。

音楽は感情を誘発する効果がある刺激の一つである。音楽を構成するテンポや音の大きさなどが聴取者の感情を変化させることがこれまでの研究により示されている [6]。しかし、同じ音楽を聴いてもすべての人が同じ感情になるわけではなく、誘発される感情には個人差がある [7]。したがって感情を誘導する際には個人の状態に合わせた音楽を使用する必要がある。

生体信号から個人の状態を取得する研究は盛んに行なわれている。その中でも時間分解能が高い脳波は、Human Computer Interaction (HCI) への利用が期待されている。脳波は脳から生じる電気的な活動を記録したものである。脳波のノイズ除去などの前処理や感情を予測するモデルの学習方法などが活発に議論されている [8], [9]。

このような背景より、脳波から取得した感情を反映させた音楽を利用して、感情を変化させる研究が行われてきた [10]。Sourina らは音楽データベースの中から感情誘導に使用する音楽を参加者の脳波から予測した感情を用いて選択することを提案した [11], [12]。Sourina らの提案手法では、誘導したい目標感情を 6 種類の感情の中から設定する。また、脳波から予測する感情も 6 種類である。目標感情と脳波から予測した感情が一致するように音楽データベースの中から音楽を選択することが提案されている。我々はこの手法に 2 つの課題があると考える。

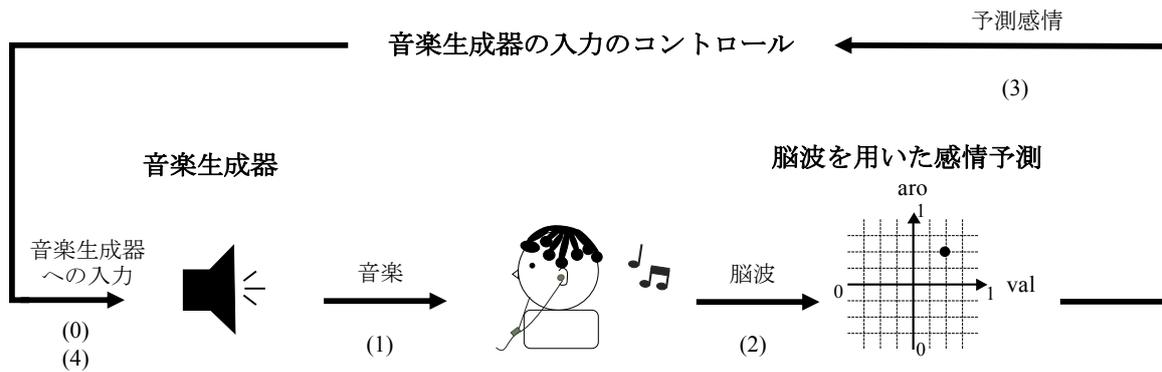


図1 個人に適応した音楽を生成する感情誘導システム

1つ目は使用できる音楽はデータベースに収録されているものに限定されることである。音楽を聴いた際の感情の個人差により、音楽データベースに参加者の感情を変化させることに有効な音楽が含まれていない可能性がある。2つ目は取り扱う感情の種類が少ないことである。この研究では目標感情に設定できる感情や脳波から予測できる感情は6種類である。感情を離散的に扱っているため、音楽聴取による感情の変化が観察しづらく、目標の感情も柔軟に設定できない。

先行研究より、我々は回帰モデルを用いて脳波から感情を予測し、予測した感情を利用して感情誘導に適した音楽をリアルタイムに生成するシステムを提案する。提案システムは3つの要素から構成される。1つ目は感情を誘発する音楽を作る音楽生成器である。本研究では感情を valence と arousal の二次元座標で表す。音楽生成器は valence と arousal の値が入力され、入力値に近い感情を誘発できる音楽を自動的にリアルタイムに作る。この音楽生成器の設計方法や有用性の評価は我々の先行研究で示している [13]。2つ目は脳波を用いた感情予測である。この要素では、音楽を聴いている際の参加者の脳波から短時間で高精度に感情を予測する。3つ目は音楽生成器のコントロールである。音楽生成器は入力した感情に近い感情が誘発できる音楽を作るが、感情の個人差により音楽生成器に入力する値は参加者に合わせた調整が必要である。この要素は脳波から予測した感情を基に音楽生成器の入力を変更する役割を担う。これらの要素を組み合わせたシステムによって、個人に適応した感情誘導を目指す。本稿では2章で感情誘導システムの概要、3章で脳波を用いた感情予測モデルの検討について述べる。これらの内容は、我々の先行研究をまとめたものである [13], [14]。

2. 感情誘導システム

提案する感情誘導システムの概要を図1に示した。システムは感情を valence と arousal のそれぞれで0から1の間の連続値で扱う。システムは以下の手順で感情を誘導する。(0) 誘導したい目標感情の valence と arousal を音楽生成器に入力する。(1) 感情を変化させるための音楽が音楽生成器で作られる。(2) 音楽を聴取している間の脳波から感情を予測する。(3) 目標感情と予測した感情の差を計算し、次のループで音楽生成器へ入力する valence と arousal の値を計算する。(4) 計算した valence

と arousal の値を音楽生成器へフィードバックする。この(1)から(4)を繰り返す。

3. 脳波を用いた感情予測

本章では、感情誘導システムの構築に必要な脳波を用いた感情予測について述べる。音楽生成器で作られた音楽を聴取している際の脳波から感情を高精度に予測する方法を検討した。

3.1 データ収集

本実験は、奈良先端科学技術大学院大学の研究倫理委員会の承認を受けて行われた。インフォームドコンセントに同意した健康な20名(10名の男性, 10名の女性)が収録に参加した。参加者の年齢は、19歳から37歳(平均24.3歳)であった。

データ収集で使用した刺激は音楽生成器で作られた1曲が20秒の音楽である。参加者は合計41曲の音楽を聴取した。参加者は音楽の聴取後に誘発された感情の valence と arousal を Self-Assessment Mannequin (SAM) [15] を使い、それぞれ0から1までの9段階で回答した。音楽のサンプルは以下のURLで公開されている (<https://sites.google.com/view/music-generator>)。

我々はCGX社製Quick-30で収録した脳波を1秒ごとに切り分け、バンドパスフィルタを使い5つの周波数帯に分けた。図2に示した脳波計の14チャンネルごとに5つの周波数帯の脳波波形の分散の対数を計算した。得られた70次元の特徴量から、脳波計の電極の位置を考慮した行列を作成した。この前処理により、データを1曲あたり20サンプル獲得した。データ収集に関する詳細な情報は我々の先行研究に示している [13]。

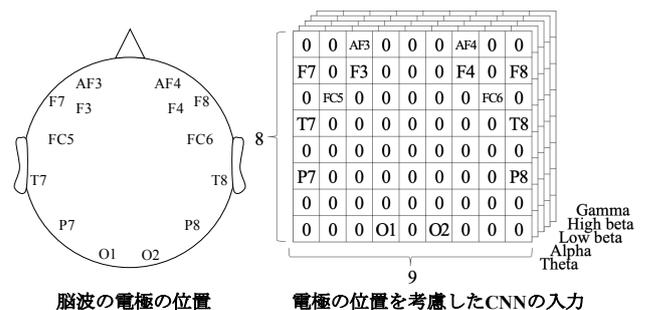


図2 CNNの入力

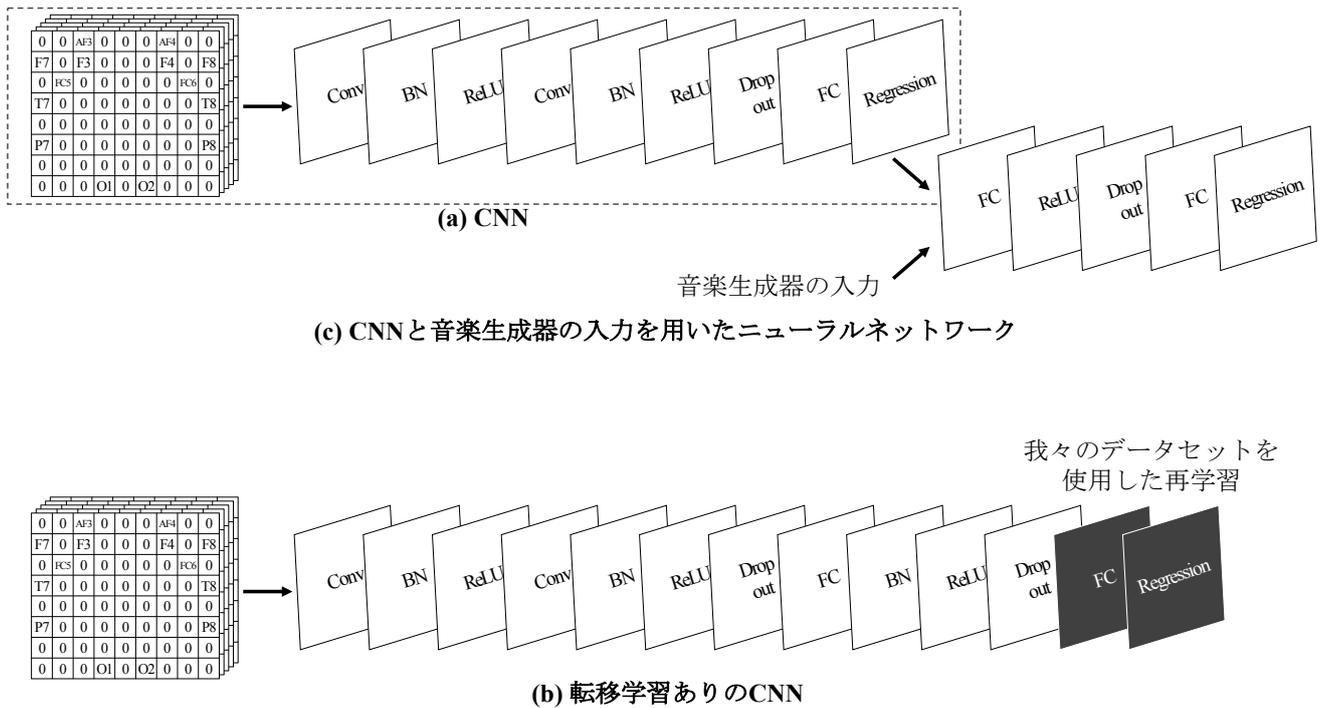


図3 提案モデル

3.2 感情予測モデル

本実験では、音楽聴取中の脳波を用いて誘発された感情を予測する回帰モデルを検証した。ベースラインモデルは先行研究[10]の類似モデルである線形回帰とした。我々はより高精度な感情予測のため、脳波計の電極の位置関係に基づいた学習を行うことを提案した。提案したモデルはCNN、転移学習ありのCNN、CNNと音楽生成器の入力を用いたニューラルネットワークの3種類である。脳波の現れ方は、参加者によって異なることが知られている[16]。個人性を考慮するため、参加者ごとに4種類のモデルを学習した。

3.2.1 線形回帰

Ehrlichらは音楽生成器で作られた音楽を聴取している間の脳波から感情を予測するため、Linear Discriminant Analysis (LDA)とシグモイド関数を利用した[10]。本実験ではLDAとシグモイド関数の類似のモデルとして線形回帰を選択し、ベースラインモデルとした。学習に用いたデータは、Z-scoreで正規化された。

3.2.2 CNN

CNNは特徴量の抽出に有効であり、画像認識だけでなく脳波を扱う研究においても利用されている[17],[18]。CNNを用いることで電極の位置関係に基づいた感情予測が期待できる。本実験では、先行研究[19],[20]を参考に図2の電極配置と5つの周波数帯に基づいた行列を作成し、CNNの入力とした。行列のサイズは $8 \times 9 \times 5$ である。

CNNの構造を図3(a)に示した。構造は、畳み込み層(フィルタサイズ: 2×2 , スライド幅: 1)、バッチ正規化層、ReLU層、畳み込み層(フィルタサイズ: 2×2 , スライド幅: 1)、バッチ正規化層、ReLU層、ドロップアウト層(ドロップアウト率: 0.5)、

全結合層(出力次元: 2)、回帰出力層である。最適化手法はAdamを適用した。学習率は0.001、バッチサイズは64、学習エポックは100である。

3.2.3 転移学習ありのCNN

転移学習は、あるドメインで事前に学習したモデルを別ドメインのデータを使って更新する方法である。少数データで事前学習済みモデルを更新することができるため、一度の収録で多くのデータを収録することが難しい脳波を扱う研究において、転移学習の利用が注目されている。しかし、先行研究では事前学習済みモデルとして、脳波を学習データに含まない画像認識用のモデルが利用された[21]。我々は、事前学習の際にも脳波を利用の方が感情予測に効果があると考え、Dataset for Emotion Analysis using EEG, Physiological and Video Signals (DEAP)[22]を用いた事前学習済みモデルを利用する。DEAPデータセットはミュージックビデオを視聴している間の脳波と感情の主観評価を収録したデータセットである。本実験の刺激に映像は含まれていないが、本実験もDEAPデータセットも刺激に音楽が含まれているため、DEAPデータセットを事前学習に使用した。DEAPデータセットには32名分のデータが収録されており、刺激であるミュージックビデオの長さは1分である。このミュージックビデオの再生前には3秒間の無音状態が含まれている。DEAPデータセットの脳波のサンプリング周波数は128Hzであり、4Hzから45Hzのバンドパスフィルタが適用されている。このDEAPデータセットに対して、我々のデータセットと同じ前処理を行なった。音楽の長さの違いより、我々のデータセットではデータを1曲あたり20サンプル獲得したが、DEAPデータセットでは1曲あたり60サンプル獲得した。

CNNの構造を図3(b)に示した。構造は、畳み込み層(フィ

表 1 4 種類のモデルによる予測値と実測値の RMSE

参加者	線形回帰		CNN		CNN+転移学習		CNN+音楽生成器の入力		音楽生成器の入力	
	val	aro	val	aro	val	aro	val	aro	val	aro
1	0.285	0.330	0.261	0.282	0.254	0.265	0.124	0.150	0.140	0.125
2	0.307	0.234	0.287	0.217	0.281	0.200	0.180	0.144	0.174	0.216
3	0.276	0.286	0.247	0.229	0.233	0.223	0.137	0.130	0.140	0.140
4	0.118	0.164	0.104	0.140	0.100	0.130	0.097	0.128	0.266	0.301
5	0.334	0.283	0.312	0.250	0.309	0.240	0.205	0.143	0.192	0.152
6	0.244	0.285	0.222	0.266	0.214	0.249	0.223	0.169	0.293	0.122
7	0.403	0.410	0.353	0.372	0.346	0.367	0.143	0.188	0.122	0.149
8	0.259	0.308	0.217	0.277	0.206	0.268	0.228	0.197	0.402	0.244
9	0.213	0.290	0.192	0.254	0.186	0.243	0.157	0.128	0.238	0.128
10	0.074	0.166	0.073	0.152	0.062	0.147	0.053	0.122	0.256	0.360
11	0.165	0.233	0.151	0.217	0.141	0.209	0.115	0.157	0.232	0.186
12	0.216	0.291	0.195	0.263	0.182	0.262	0.166	0.220	0.305	0.253
13	0.218	0.318	0.208	0.302	0.200	0.289	0.135	0.222	0.177	0.201
14	0.193	0.227	0.174	0.193	0.165	0.188	0.172	0.134	0.277	0.244
15	0.252	0.239	0.228	0.210	0.222	0.202	0.143	0.177	0.175	0.306
16	0.081	0.165	0.067	0.148	0.050	0.140	0.068	0.137	0.293	0.244
17	0.464	0.379	0.411	0.369	0.407	0.357	0.155	0.169	0.146	0.137
18	0.178	0.189	0.152	0.152	0.147	0.152	0.102	0.139	0.220	0.250
19	0.077	0.222	0.067	0.190	0.059	0.183	0.057	0.134	0.253	0.213
20	0.384	0.367	0.365	0.299	0.369	0.295	0.367	0.282	0.347	0.284
平均	0.237	0.269	0.214	0.239	0.207	0.230	0.151	0.164	0.232	0.213
標準偏差	0.108	0.071	0.099	0.068	0.101	0.067	0.071	0.041	0.075	0.070

ルタサイズ: 2x2, スライド幅: 1), バッチ正規化層, ReLU 層, 畳み込み層 (フィルタサイズ: 2x2, スライド幅: 1), バッチ正規化層, ReLU 層, ドロップアウト層 (ドロップアウト率: 0.5), 全結合層 (出力次元: 8), バッチ正規化層, ReLU 層, ドロップアウト層 (ドロップアウト率: 0.2), 全結合層 (出力次元: 2), 回帰出力層である。始めに, DEAP データセットを用いてモデルを学習した。その後, 我々のデータセットを用いて参加者ごとに CNN の最後の 2 層において再学習を行なった。モデルの事前学習と再学習において, 最適化手法は Adam を適用した。モデルの事前学習時の学習率は 0.001, バッチサイズは 128, 学習エポックは 5 である。モデルの再学習時の学習率は 0.0005, バッチサイズは 64, 学習エポックは 100 である。

3.2.4 CNN と音楽生成器の入力を用いたニューラルネットワーク

脳波から予測された感情が参加者の感じた感情と大幅に異なる場合, 本研究で提案した感情誘導システムは感情誘導に適さない音楽を生成する可能性がある。そこで, 脳波以外の情報を加えた感情予測を提案する。先行研究では, 脳波と音声情報を用いた音声の品質予測 [23] や, 脳波と皮膚電気反応を用いた感情予測 [24] が行われてきた。本研究では迅速な感情予測が求められるため, 最小限の情報やモデルから感情を予測する必要がある。そこで我々は, 音楽生成器に入力した valence と arousal を学習に利用することを提案する。音楽生成器で作られる音楽は, 音楽生成器に入力した感情に近い感情を誘発することが我々の先行研究で示された [13]。このことから音楽生成器への入力, 誘発される感情の予測値であると捉えることができる。

脳波から予測した感情と音楽生成器の入力を利用することで, 感情予測の精度向上を目指した。

ニューラルネットワークの構造を図 3 (c) に示した。まず 3.2.2 CNN で学習したモデルを用いて脳波から valence と arousal を予測した。次に, 脳波から予測した valence と arousal, 音楽生成器に入力した valence と arousal の 4 つの入力からニューラルネットワークを用いて最終的な valence と arousal を予測した。この 4 つの入力から感情を予測するモデルの構造は全結合層 (出力次元: 8), ReLU 層, ドロップアウト層 (ドロップアウト率: 0.2), 全結合層 (出力次元: 2), 回帰出力層である。最適化手法は Adam を適用した。学習率は 0.001, バッチサイズは 64, 学習エポックは 100 である。

3.3 結果

4 種類のモデルは参加者ごとに学習された。また, テストデータを 1 曲分のサンプル, 学習データを残った曲のサンプルとした leave-one-music-out 交差検証を適用した。モデルの感情の予測値と参加者が評価した感情の実測値の Root Mean Squared Error (RMSE) を計算した結果, 表 1 のようになった。表 1 では, 線形回帰, CNN, 転移学習ありの CNN, CNN と音楽生成器の入力を用いたニューラルネットワークの RMSE を示している。また, ニューラルネットワークの学習で使用した音楽生成器の入力の影響を調査するため, 音楽生成器に入力した感情と参加者が評価した感情の実測値の RMSE も示している。これらの RMSE を比較すると, 最も参加者の RMSE の平均が低くなったのは CNN と音楽生成器の入力を用いたニューラルネットワークであった。複数の情報を組み合わせたニューラルネッ

トワークが最も RMSE が低くなることがわかった。また、ウィルコクソンの符号順位検定により、このニューラルネットワークと Ehrlich らの類似モデルでありベースラインモデルの線形回帰の間に有意差があることがわかった ($p < 0.05$)。さらに CNN と線形回帰や、転移学習ありの CNN と線形回帰の間にも有意差があった ($p < 0.05$)。以上のことから提案した CNN を利用した 3 種類のモデルは線形回帰より有意に RMSE が低くなることがわかった。

4. 結 論

本研究の目的は、音楽聴取時の感情の個人差を考慮した音楽生成により、個人に適応した感情誘導を行うことであった。この目的のために、本稿では脳波を用いた感情予測について取り組み、CNN を利用したモデルを 3 種類提案した。

学習したモデルの予測値と実測値の RMSE を計算した結果、CNN を利用したモデルがベースラインモデルである線形回帰より RMSE が有意に低くなった。このことから、脳波の電極の位置関係を考慮できる CNN による学習が感情予測に効果があることがわかった。また、比較した 4 種類のモデルの中で CNN と音楽生成器の入力を用いたニューラルネットワークが最も RMSE が低くなった。CNN から予測した感情または音楽生成器に入力した感情が実際に誘発された感情と大きく異なっていたとしても、音楽生成器に入力した感情または CNN から予測した感情が修正することで実際の感情と近い感情を予測することができたと考えられる。

今後は、音楽生成器と本稿で提案した脳波を用いた感情予測を組み合わせることで感情誘導システムを構築する。また、音楽生成器の入力のコントロール方法について検討し、システムで作られた個人に適応した音楽が感情誘導に有効であるか調査する。

5. 謝 辞

本研究は JST, CREST, JPMJCR19A5 の支援を受けたものである。また、本研究は理研の大学院生リサーチ・アソシエイト制度の下での成果である。

文 献

- [1] V. Santos, F. Paes, V. Pereira, O. Arias-Carrión, A.C. Silva, M.G. Carta, A.E. Nardi, and S. Machado, "The role of positive emotion and contributions of positive psychology in depression treatment: systematic review," *Clinical practice and epidemiology in mental health*, pp.221–237, 2013.
- [2] J. Joermann and C.H. Stanton, "Examining emotion regulation in depression: A review and future directions," *Behaviour research and therapy*, vol.86, pp.35–49, 2016.
- [3] A. Compare, C. Zarbo, E. Shonin, W. Van Gordon, and C. Marconi, "Emotional regulation and depression: A potential mediator between heart and mind," *Cardiovascular psychiatry and neurology*, vol.2014, pp.1–10, 2014.
- [4] B.L. Fredrickson and C. Branigan, "Positive emotions broaden the scope of attention and thought-action repertoires," *Cognition & emotion*, vol.19, no.3, pp.313–332, 2005.
- [5] 山崎勝之, "ポジティブ感情の役割," *パーソナリティ研究*, vol.14, no.3, pp.305–321, 2006.
- [6] P. Gomez and B. Danuser, "Relationships between musical structure and psychophysiological measures of emotion.," *Emotion*, vol.7, no.2,

- p.377, 2007.
- [7] R.J. Larsen and T. Ketelaar, "Personality and susceptibility to positive and negative emotional states.," *Journal of personality and social psychology*, vol.61, no.1, p.132, 1991.
- [8] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of artifacts from eeg signals: a review," *Sensors*, vol.19, no.5, p.987, 2019.
- [9] A. Craik, Y. He, and J.L. Contreras-Vidal, "Deep learning for electroencephalogram (eeg) classification tasks: a review," *Journal of neural engineering*, vol.16, no.3, p.031001, 2019.
- [10] S.K. Ehrlich, K.R. Agres, C. Guan, and G. Cheng, "A closed-loop, music-based brain-computer interface for emotion mediation," *PLoS one*, vol.14, no.3, pp.1–24, 2019.
- [11] O. Sourina, Y. Liu, and M.K. Nguyen, "Real-time eeg-based emotion recognition for music therapy," *Journal on Multimodal User Interfaces*, vol.5, no.1-2, pp.27–35, 2012.
- [12] Y. Liu, O. Sourina, and M.K. Nguyen, "Real-time eeg-based emotion recognition and its applications," *Transactions on computational science XII*, pp.256–277, Springer, 2011.
- [13] K. Miyamoto, H. Tanaka, and S. Nakamura, "Music generation and emotion estimation from eeg signals for inducing affective states," *Companion Publication of the 2020 International Conference on Multimodal Interaction*, pp.487–491, 2020.
- [14] K. Miyamoto, H. Tanaka, and S. Nakamura, "Emotion estimation from eeg signals and expected subjective evaluation," *2021 9th International Winter Conference on Brain-Computer Interface (BCI)IEEE*, pp.1–6 2021.
- [15] M.M. Bradley and P.J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol.25, no.1, pp.49–59, 1994.
- [16] Y.-P. Lin and T.-P. Jung, "Improving eeg-based emotion classification using conditional transfer learning," *Frontiers in human neuroscience*, vol.11, p.334, 2017.
- [17] S. Siddharth, T.-P. Jung, and T.J. Sejnowski, "Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing," *IEEE Transactions on Affective Computing*, pp.1–1, 2019.
- [18] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from eeg with deep recurrent-convolutional neural networks," *arXiv preprint arXiv:1511.06448*, pp.–, 2015.
- [19] J. Li, Z. Zhang, and H. He, "Hierarchical convolutional neural networks for eeg-based emotion recognition," *Cognitive Computation*, vol.10, no.2, pp.368–380, 2018.
- [20] Y. Yang, Q. Wu, Y. Fu, and X. Chen, "Continuous convolutional neural network with 3d input for eeg-based emotion recognition," *International Conference on Neural Information Processing* Springer, pp.433–443 2018.
- [21] G. Xu, X. Shen, S. Chen, Y. Zong, C. Zhang, H. Yue, M. Liu, F. Chen, and W. Che, "A deep transfer convolutional neural network framework for eeg signal classification," *IEEE Access*, vol.7, pp.112767–112776, 2019.
- [22] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing*, vol.3, no.1, pp.18–31, 2011.
- [23] I.H. Parmonangan, H. Tanaka, S. Sakti, and S. Nakamura, "Combining audio and brain activity for predicting speech quality," *Proc. Interspeech 2020*, pp.2762–2766, 2020.
- [24] Y.-H. Kwon, S.-B. Shin, and S.-D. Kim, "Electroencephalography based fusion two-dimensional (2d)-convolution neural networks (cnn) model for emotion recognition system," *Sensors*, vol.18, no.5, p.1383, 2018.