

Effects of Individual Differences in Emotion Estimation Based on Music and EEG

Kentaro FUKAO^{*}, Taishi KOBAYASHI^{*}, Keiko ONO^{**} and Erina MAKIHARA^{***}

(Received June 28, 2022)

Emotion recognition using electroencephalogram (EEG) has been paid much attention based on various machine learning techniques. Recognizing emotion is a difficult task due to the non-linear nature of the EEG signal. In this study, we propose a regression model for emotion estimation using Convolutional Neural Networks (CNN). Using EEG open data set, DEAP, we verify that the proposed method slightly improved the performance by incorporating regularization terms. We also observed that the performance is different in each individual. The proposed method is a general model but would enhance the performance by incorporating individual differences.

Key words : emotion estimation, brain waves, music

キーワード : 感情推定, 脳波, 音楽

音楽と脳波に基づく感情推定における個体差の影響

深尾 賢太郎・小林 汰司・小野 景子・槇原 絵里奈

1. はじめに

信号の計測および処理技術の発展により、生体信号が様々な分野で注目されている。ウェアラブル端末では、生体信号を用いた健康管理システムが導入され、リアルタイムで自身の健康状態を知ることができる¹⁾。また、生体信号を用いて運転者の表情や視線運動を把握し、異常を検出した場合に警告する事故防止システムも提案されている²⁾。生体信号の活用は、私たちの生活を支援する一助を担っている。

生体信号のうち脳波は、精神や意識を支配する脳に

発現する生体信号であり、意識的に偽ることは容易ではない。ゆえに、脳波は人間の感情を反映した情報であると考えられているため、様々な感情推定手法に用いられている。鈴木らは、音楽聴取時の脳波を多チャンネル脳波計で計測し、機械学習による感情推定モデルの構築に用いた³⁾。また池田らは、ジャグリングステージ鑑賞時の脳波と脈拍を用いて感情推定手法の性能を評価した⁴⁾。しかし、いずれの研究も脳波計測時に聴取した音楽の情報を感情推定に用いていない。音楽は、脳波の変化に影響を与えていると考えられるため、感情推定において重要な情報を有する可能性

^{*} Graduate School of Science and Engineering, Doshisha University, Kyoto
Telephone:+81-774-65-6930, Fax:+81-774-65-6716, E-mail:fukao.@mikilab.doshisha.ac.jp
^{**} Faculty of Science and Engineering, Doshisha University, Kyoto
Telephone:+81-774-65-6930, Fax:+81-774-65-6716, E-mail:kono@mail.doshisha.ac.jp
^{***} Faculty of Science and Engineering, Doshisha University, Kyoto
Telephone:+81-774-65-6930, Fax:+81-774-65-6716, E-mail:emakihar@mail.doshisha.ac.jp

がある．そこで，本研究では，Convolutional Neural Networks（以下，CNN）を用いて，感情推定を行う回帰モデルを提案する．加えて，感情推定に対して人の個体差が及ぼす影響に関して，提案した回帰モデルの性能評価とパラメータ分析の観点から考察する．

2. 生体信号による感情推定の先行研究

2.1 脳波指標のみを用いた感情推定

音楽聴取時の感情推定に関する先行研究において，脳波のみを用いた感情分類モデルが提唱されている．機械学習による感情分類において，単チャンネル脳波計から得られた脳波指標のみを利用した感情分類は容易でない．そこで，鈴木らは，単チャンネル脳波計ではなく，多チャンネル脳波計から取得した脳波指標のみを用いて音楽を聴取した際の感情分類を行い，約72%の精度を実現している³⁾．

2.2 脳波と脈拍を組み合わせた感情推定

感情推定に関する先行研究において，表情の映像を用いた感情推定の精度に比べ，生体信号を利用した感情推定の精度の方が高いことが報告されている⁵⁾．池田らは，生体信号として脳波と脈拍を利用し，感情推定の精度を評価した．脈拍センサから得られた，pNN50と呼ばれる数値指標と，脳波センサから得られた，センサ独自のアルゴリズムに基づいた数値指標を利用している．得られた被験者の数値データを，感情円環モデルであるRussellの二次元座標上⁶⁾にプロットすることで感情を推定し，表情の映像を用いた場合よりも高い精度を実現している．

また，ジャグリングステージ鑑賞時の感情推定に関する先行研究において，脳波と脈拍を利用した感情推定の有効性が報告されている⁴⁾．脳波は人の精神や意識を反映した生体信号であるため，脳波を利用することで，人が意識的に行う表現や主観を除去した形で感情推定を実現している．また，脈拍も身体反応から直接得られる生体信号であり，客観性の高い推定を可能にする．結果として，生体信号を用いた感情の推定結果と被験者による主観評価の相関係数が高い値で得られたことから，生体信号が感情推定において有効であるといえる．

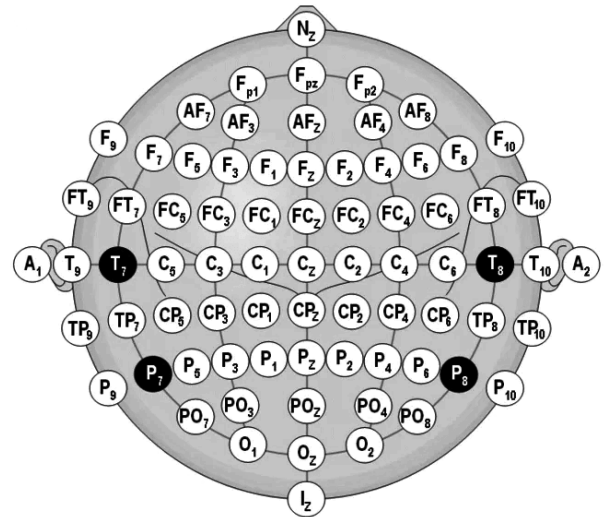


Fig. 1. EEG electrode layout.

3. DEAP

A Database for Emotion Analysis using Physiological Signals（以下，DEAP）は，ミュージックビデオ（以下，MV）視聴時に計測した脳波と視聴後の人間の感情を紐付けたデータセットである⁷⁾．本データセットは，脳波による人間の感情推定に関する研究に用いられている．本研究ではDEAPを利用し，感情推定を行う回帰モデルの構築を目指す．

DEAPは，32名の被験者を対象に実施したMV視聴実験の結果を含んでいる．実験において被験者は，40本のMVを視聴しており，視聴時の脳波データが計測されている．視聴するMVは，感情を刺激しやすい60秒のハイライト部分を用いている．また，脳波計測時の電極配置をFig. 1に示す⁸⁾．脳波データは，国際基準による電極配置法に従って配置された32個の電極から取得している．

4. 音楽と脳波の関係性調査

4.1 検証概要

本検証では，感情推定モデルの学習前の検証として，脳波による人間の感情推定に関する研究に用いられているDEAPを用いて音楽と音楽聴取時の脳波の関係性を調査する．脳波に影響を与えると考えられる音楽を学習に用いることで，感情推定の精度向上が期待で

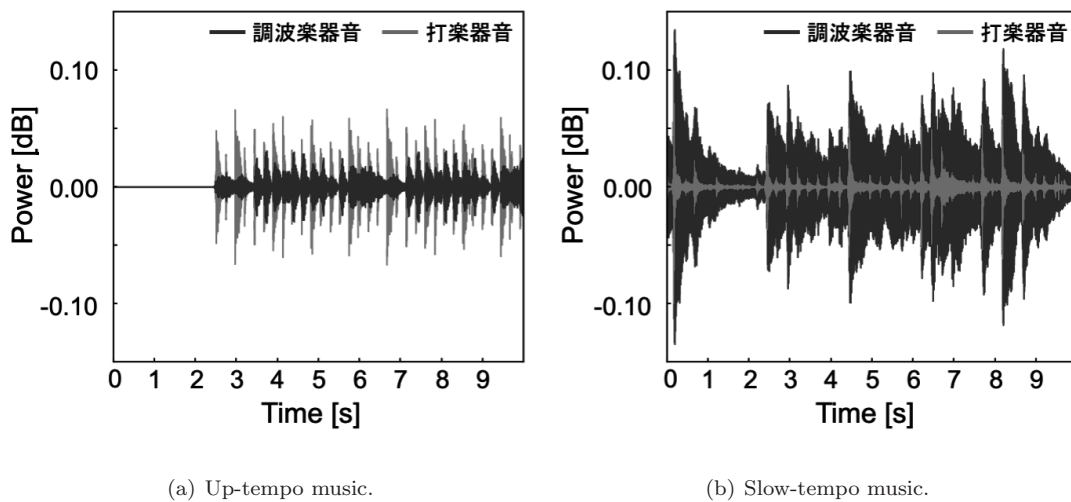


Fig. 2. Harmonic/percussive sound separation.

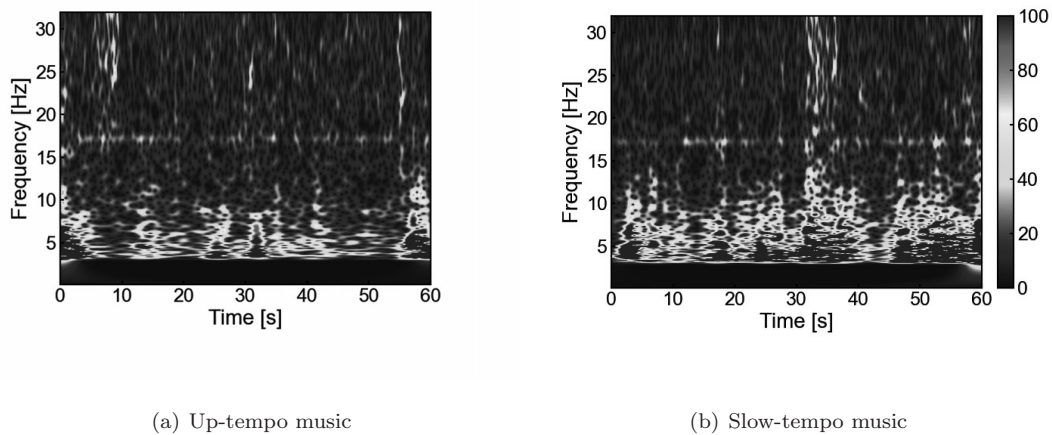


Fig. 3. Comparison of EEG.

きる。また、検証においてはMVの映像と音楽の情報のうち、音楽の情報のみに着目した。

4.2 検証内容

音楽に複数の楽器の音やアーティストの声が含まれているため、構成要素の抽出を行った。音楽の構成要素の抽出には、調波打楽器音分離を用いた。調波打楽器音分離とは、スペクトログラムに現れる特徴を用いて、音楽をメロディ部を構成する調波楽器音と、リズム部を構成する打楽器音に分離する手法である。調波打楽器音分離には、liblosa と呼ばれる音楽や音声解析を行うためのPythonパッケージ⁹⁾を利用している。また、脳波は一般に、0 Hz から 30 Hz の周波数成分から構成され、感情の状態に応じて特定周波数帯に

変化が現れることが知られている。そのため、ウェーブレット解析を用いて脳波の周波数成分を解析した。ウェーブレット解析では、低周波領域で周波数の分解能を優先し、高周波領域で時間の分解能を優先して変換を行うことで不確定性原理を解消し、時間領域の情報を残した形での周波数解析を実現している。ウェーブレット解析には、swan と呼ばれるウェーブレット解析ツールを利用している。

4.3 検証結果および考察

音楽の構成要素を調波打楽器音分離で抽出した結果を Fig. 2 に示す。Fig. 2 より、視覚的に音楽が構成要素毎に分離されていることが確認できる。加えて、各要素を聴取すると、音楽がメロディ部とリズム部に

分離された様子が確認できた。

また、アップテンポな音楽とスローテンポな音楽の構成要素について比較すると、打楽器音成分と調波楽器音成分の相対的な大きさと構成に差異がある。アップテンポな曲は、調波楽器音に対して打楽器音成分が占める割合が大きく、打楽器音は短い間隔で発生している。一方で、スローテンポな曲は打楽器音に対して調波楽器音成分が占める割合が大きく、打楽器音は長い間隔で発生している。以上より音楽の構成要素の差異は、音楽の特徴として学習に用いることが可能であると考えられる。

ウェーブレット解析により電極 PO4 における脳波の周波数成分を解析した結果を Fig. 3 に示す。Fig. 3 より、アップテンポな音楽とスローテンポな音楽について比較すると、脳波の周波数成分に着目した際、 α 波周波数帯域の発現の様子が異なることが確認できる。アップテンポな音楽では、 α 波の発現は一時的だが、スローテンポな音楽では、 α 波が連続的に強く発現している。 α 波とは、8~13 Hz の周波数成分であり、安静、覚醒、閉眼状態で正常成人の頭頂部と後頭部で最も著明に発現する¹⁰⁾。反対に、目を開いたり精神活動を行ったりすると減衰する。すなわち、 α 波は人の感情状態を反映している情報であり、脳波の特徴として学習に用いることが可能であると考えられる。8 Hz より低い周波数帯域は、 δ 波、 θ 波に分かれているが、種々の脳機能障害の判別に用いられるため、本研究においては α 波周波数帯域に注目した¹⁰⁾。電極 PO4 は Fig. 1 に示す通り後頭部の位置に対応しており、 α 波の特徴の学習に適している電極といえる。また、頭頂部に対応する電極 Cz も同様に学習に適していると考えられる。実際に、 α 波に注目すると、Valence の尺度と電極 Cz、Arousal の尺度と電極 PO4 について、それぞれ関係性がみられる可能性があることが報告されている⁷⁾。

5. 提案する感情推定モデル

5.1 モデルの概要

本研究では、音楽における調波楽器音分離の結果画像、および脳波におけるウェーブレット解析の結果画像

を CNN を用いて学習することで、脳波測定時に用いた音楽を考慮した感情推定を行う回帰モデルを提案する。入力画像とし、出力は Valence と Arousal の推定値とする。また、感情推定には Russell の円環モデルを利用する。Russell の円環モデルとは、Valence と呼ばれる快・不快の度合いと、Arousal と呼ばれる感情の強さの、2つの指標を二次元座標上に対応させることで感情を表現するモデルである¹¹⁾。

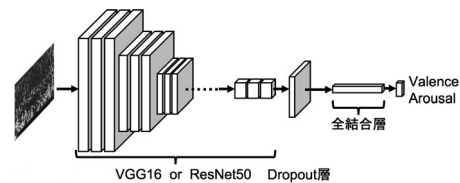


Fig. 4. Network architecture.

5.2 モデルの構成と性能評価

提案モデルのネットワークの構成を Fig. 4 に示す。本研究では、特徴抽出部に Visual Geometry Group (以下、VGG) 16 または Residual Network (以下、ResNet) 50 を用いたモデルを提案する。提案モデルでは、特徴抽出部の最終層の全結合層を除いたのち、確率 0.5 の Dropout 層、活性化関数に ReLU を用いた 256 次元の全結合層、Valence と Arousal の 2 値を出力する出力層を結合する。

次に、学習条件について述べる。損失関数には、以下の式 (1) で表される、平均二乗誤差 (MSE: Mean Squared Error) を用いる。

$$f_{MSE} = \frac{1}{n} \sum_{i=1}^n ((y_i - \hat{y}_i)^2) \quad (1)$$

ここで、 n はデータ数、 y は真値、 \hat{y} は推定値を表す。各データに対して真値と推定値の差の二乗値を計算し、その総和をデータ数で割った値を出力する関数である。最適化アルゴリズムには Adam を使用し、学習率 η を 0.00001 としている。学習は、バッチサイズを 16 としたミニバッチ学習を行う。全ての層のパラメータを更新可能な状態で fine-tuning を行っている。fine-tuning とは、学習済みモデルのパラメータをコピーして、別のニューラルネットワークのパラメータの初期値として利用する手法である。

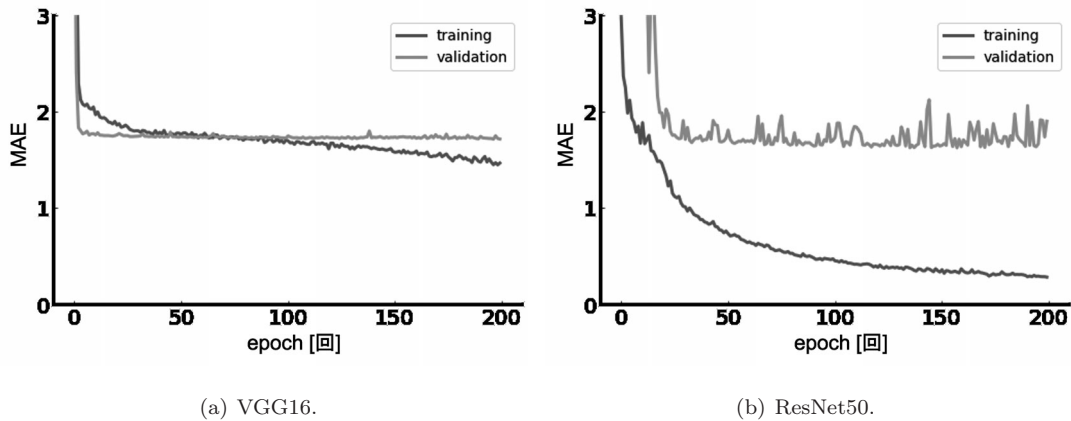


Fig. 5. Learning curve at PO4.

次に、評価方法について述べる。評価関数には、以下の式 (2) で表される、平均絶対誤差 (MAE: Mean Absolute Error) を用いる。

$$f_{MAE} = \frac{1}{n} \sum_{i=1}^n \|y_i - \hat{y}_i\| \quad (2)$$

先ほどと同様に、 n はデータ数、 y は真値、 \hat{y} は推定値を表す。各データに対して真値と推定値の差の絶対値を計算し、その総和をデータ数で割った値を出力する関数である。本研究では、Valence と Arousal それぞれに対して一度 MAE の値を求め、2 つの MAE の平均値を最終的な MAE とする。性能の評価には、ホールドアウト法を用いている。ホールドアウト法とは、データセットを任意の割合で訓練データ、検証データ、テストデータに分割して処理を行うことで、学習済みモデルの汎化性能を評価する方法である。

6. 脳波データに対するモデル間の性能評価

6.1 実験概要

本実験の目的は、脳波データを学習に用いた提案モデルの性能を評価することである。音楽の情報を考慮する前の段階で、脳波データのみを用いた感情推定モデルの性能を比較検証する。性能の比較検証は、特徴抽出部に VGG16 を用いたモデルと、ResNet50 を用いたモデル間で行う。学習に用いるデータは、電極 PO4 と電極 Cz のウェーブレット解析の結果画像であり、640x480 のカラー画像である。学習は、各電極ごとに分けて行う。各電極につき 1280 枚の画像があり、訓練データ 896 枚、検証データ 256 枚、テストデータ

128 枚とする。

6.2 実験結果および考察

はじめに、電極 PO4 に関する結果について述べる。特徴抽出部の違いによって、MAE の推移を比較した結果を Fig. 5 に示す。いずれのモデルにおいても検証データの学習が停滞しており、過学習が発生していると考えられる。訓練データに対しては、ResNet50 を用いたモデルでは MAE が約 0.3 まで下がっていることから、特定の学習データに適したモデルが実現できているといえる。

次に、電極 Cz に関する結果について述べる。特徴抽出部の違いによって、MAE の推移を比較した結果を Fig. 6 に示す。電極 PO4 と同様に、過学習が発生していると考えられる。学習データに対しても、電極 PO4 の場合と同様の傾向が見られ、特定の学習データに適した感情推定モデルが実現できているといえる。したがって、提案モデルには汎化性はないが、特定のデータが学習可能であると考えられる。

各モデルの性能を比較するために、ホールドアウト法におけるテストデータに対する MAE を Table 1 に示す。いずれの電極においても、ResNet50 を用いたモデルの方が性能が高い。よって、脳波データのみを用いた感情推定モデルにおいては、ResNet50 を用いた場合の方が性能が高いことがわかる。しかし、音楽の情報を考慮する前に、過学習の解消が必要である。

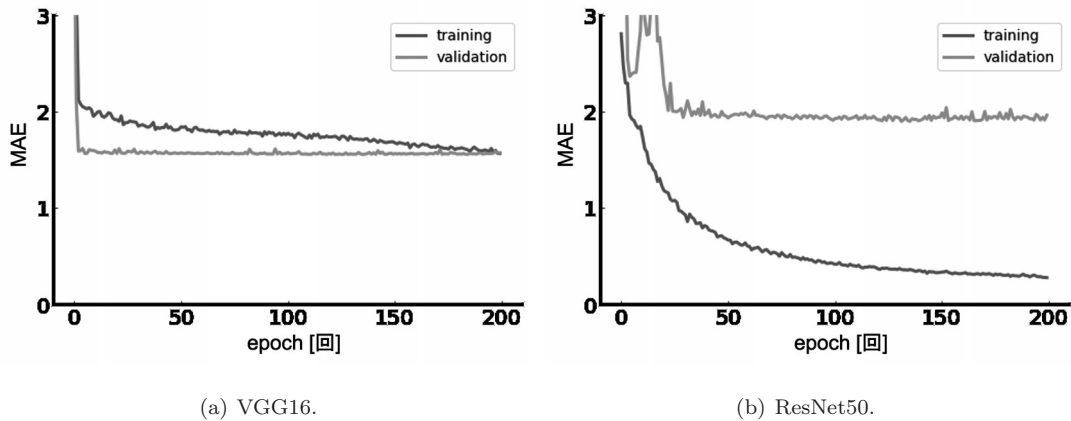


Fig. 6. Learning curve at Cz.

Table 1. MAE of each model

	VGG16/PO4	ResNet50/PO4	VGG16/Cz	ResNet50/Cz
MAE	1.947	1.691	1.762	1.732

Table 2. MAE when changing λ

	$\lambda = 0$	$\lambda = 0.0001$	$\lambda = 0.001$	$\lambda = 0.01$
MAE	1.511	1.507	1.512	1.572

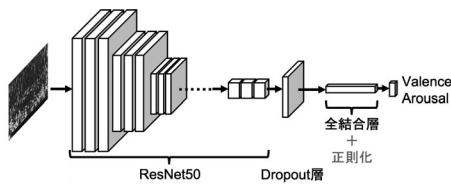


Fig. 7. Network architecture with regularization.

7. 脳波データに対する正則化パラメータの影響分析

7.1 実験概要

本実験の目的は過学習の解消および脳波データに対する正則化パラメータ λ の影響分析である。学習にL2正則化を用いる場合と、用いない場合の学習推移を比較検証する。学習に用いるデータは、脳波データに対するモデル間の性能評価において性能が高かった、電極 PO4 のウェーブレット解析の結果画像である。特徴抽出部には、脳波データに対するモデル間の性能評価において性能が高かった、ResNet50 を用いる。本実験で用いるモデルのネットワーク構成を Fig. 7 に示す。

7.2 実験結果および考察

実験結果を、Fig. 8 に示す。結果より、訓練データおよび検証データに対して、 λ の値による学習推移の

変化はみられなかった。原因として、一部被験者間で脳波のウェーブレット解析の結果画像に顕著な差異が見られることが挙げられる。また、テストデータに対する MAE を Table 2 に示す。 $\lambda = 0.0001$ の場合に、MAE が 1.507 で最も精度が高くなっている。Table 1 で示した電極 PO4 と ResNet50 を用いた際の MAE の値 1.691 と比較すると、正則化を用いることにより、僅かに性能が向上した。しかし、Russell の円環モデルにおいては、真値と推定値の差の絶対値が 1 より大きくなると、推定される感情の種類が変化する可能性があるため、感情推定の精度としては十分とはいえない。そこで、感情推定の精度向上の手段の一つとして、脳波測定時に用いた音楽の情報の活用が挙げられる。音楽は、脳波の変化に影響を与えていると考えられるため、感情推定において重要な情報を有する可能性がある。

8. 音楽データを追加したモデルの性能評価

8.1 実験概要

脳波データに加えて音楽データを追加した際の提案モデルに対する性能を評価する。学習に用いる脳波データは、脳波データに対するモデル間の性能評価において最も性能が高かった、電極 PO4 のウェーブ

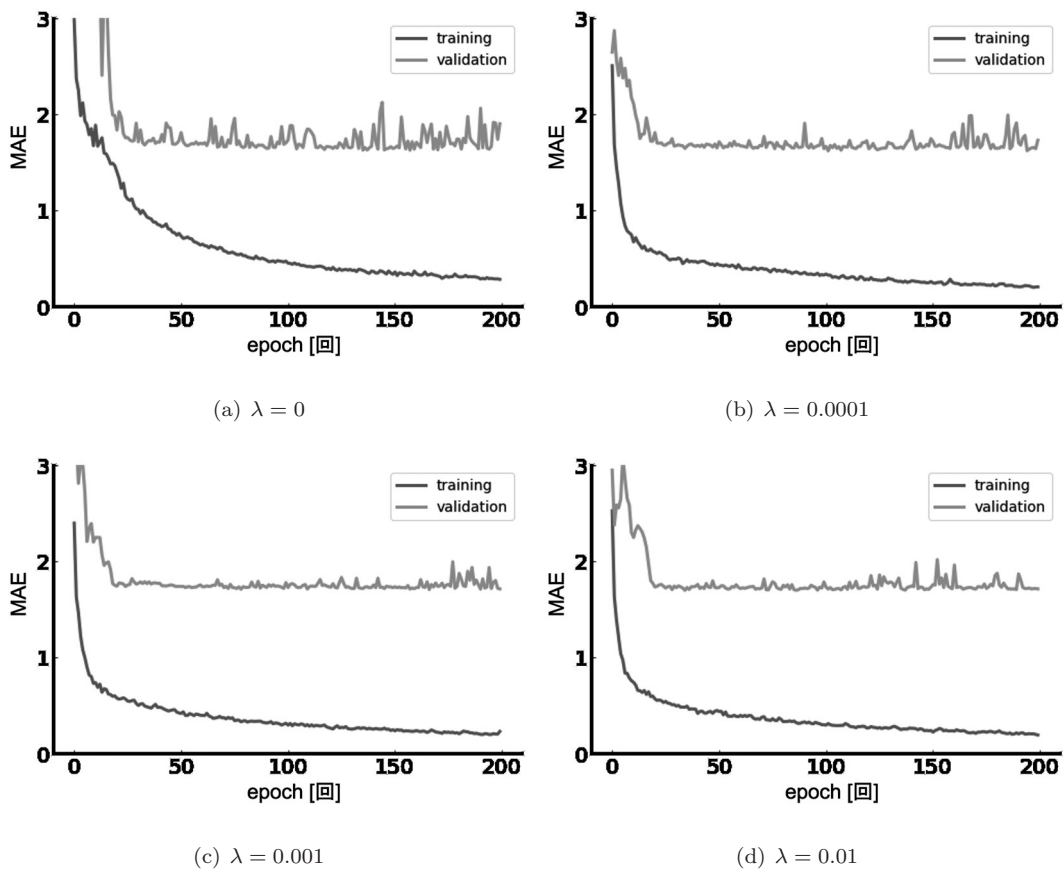
Fig. 8. Learning curve with changing λ .

Table 3. MAE when usign music data.

	$\lambda = 0$	$\lambda = 0.0001$	$\lambda = 0.001$	$\lambda = 0.01$
MAE	1.941	1.632	1.838	1.701

レット解析の結果画像である。音楽データについては、Fig. 2 に示した各 MV に調波打楽器音分離を行った結果画像を用いる。特徴抽出部には、ResNet50 を用いる。OpenCV を用いて、脳波データ (640x480) と音楽データ (640x480) を縦方向に結合したカラー画像 (640x960) を作成した。本実験で用いるモデルのネットワーク構成は、7章の実験で用いた Fig. 7 の構成と同様である。

電極 PO4 に対して 832 枚の画像があり、訓練データ 576 枚、検証データ 192 枚、テストデータ 64 枚とする。DEAP データセット内に、著作権の関係上閲覧できない MV が含まれていたため、脳波データのみを用いた実験時よりデータ数は少なくなっている。

8.2 実験結果および考察

実験結果を、Fig. 9 に示す。Fig. 8 および Fig. 9 より、訓練データおよび検証データに対して、音楽データの追加による学習推移の変化は認められなかった。音楽データを追加するだけでは、学習モデルの汎化性能をあげることが容易でない。原因としては、脳波データに対する正則化パラメータの影響分析と同様に、個体差による脳波データの差異が影響していると考えられる。同じ MV を視聴した際に、一部被験者間で脳波の発現の様子が異なっている一方で、Valance と Arousal の値が似通っている場合があるため、データの有する情報の乖離が大きくなり、訓練データと検証データに対する MAE の値が近づかないと考えられる。

また、テストデータに対する MAE を Table 3 に示す。Table 3 に示した脳波データのみを用いた場合と比較して、すべての λ 条件下で、学習の精度が悪化した。但し、音楽データ結合後も、 $\lambda = 0.0001$ の場合が最も精度が良く、提案モデルにおける最適なパラメー

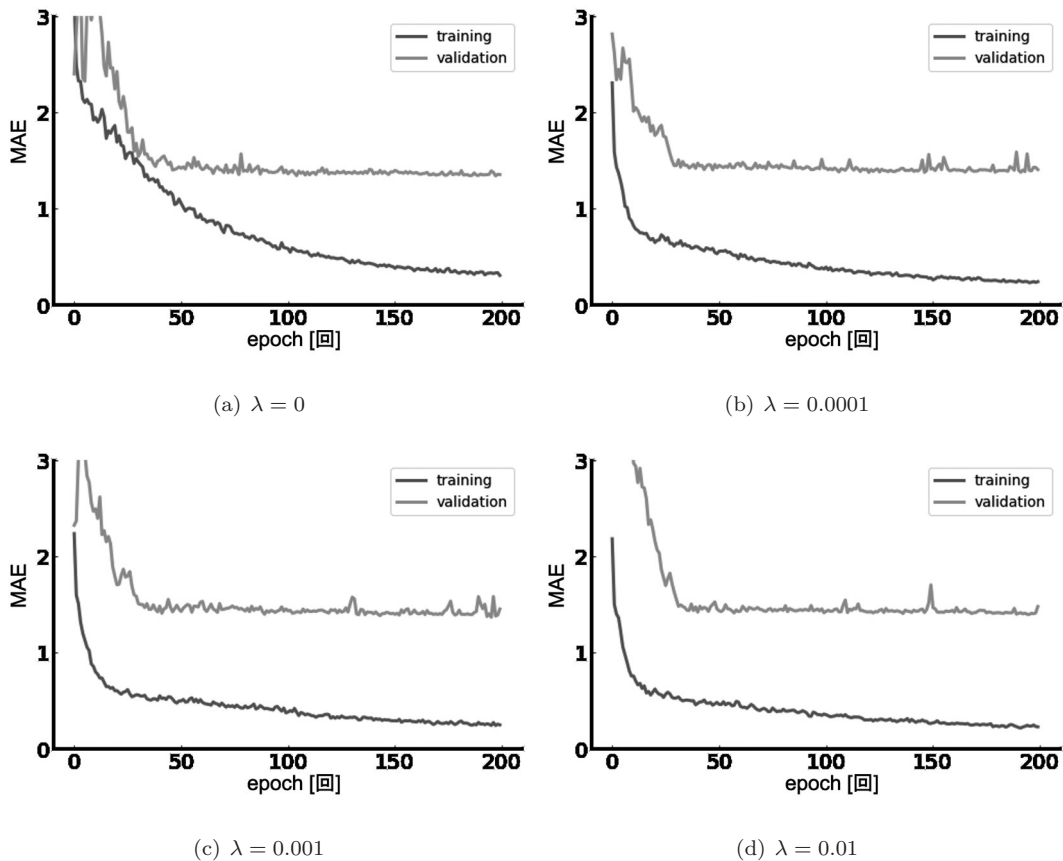


Fig. 9. Learning curve using music data.

タは、現時点で $\lambda = 0.0001$ であるといえる。

学習の精度が悪化した原因として、音楽データが一意に決まるデータであることが挙げられる。Valence と Arousal で評価された個人による受け止め方と、 α 波の発現の様子の乖離が大きい状態の学習データに対して、個体差に影響されず一意に決まる音楽データを追加したことで、学習が妨げられた可能性が考えられる。個体差の影響を調査するため、テストデータと推定結果の比較を Fig. 11 に示す。推定結果として示すのは、最も精度の高い $\lambda = 0.0001$ とした場合の結果である。比較すると、推定値が中央付近に偏って得られており、Valence と Arousal の極端な値は推定されていない。異なる被験者同士に注目した際、Valence と Arousal で評価された個人による受け止め方が類似している場合でも、 α 波の発現の様子が大きく異なるという個体差の影響を受けて学習が妨げられ、推定時に極端な値を得るのが困難になったと考えられる。個体差の影響により、各被験者からの推定値にばらつき

が見られ、結果的に推定結果が中央付近に集まっていると考えられる。

9. おわりに

本研究では、感情推定の精度向上のため脳波測定時に用いた音楽の情報に着目し、活用を検討した。提案した感情推定モデルに対して、脳波のウェーブレット解析の結果を用いて α 波周波数帯域のデータを学習に用いたところ、過学習が発生した。しかし、訓練データに対しては、MAE が約 0.3 まで下がっていることから、汎化性はないが特定のデータに適した感情推定モデルが実現できたといえる。また、過学習を防ぐ手段として正則化に着目し、脳波データに対する正則化パラメータの影響を分析した。結果として、 λ の値による学習推移の改善は見られたが大きな変化は認められなかった。原因として、生体信号としての脳波における被験者間の個体差が挙げられる。次に、脳波測定時に用いた音楽の情報に着目し、音楽データを追加した

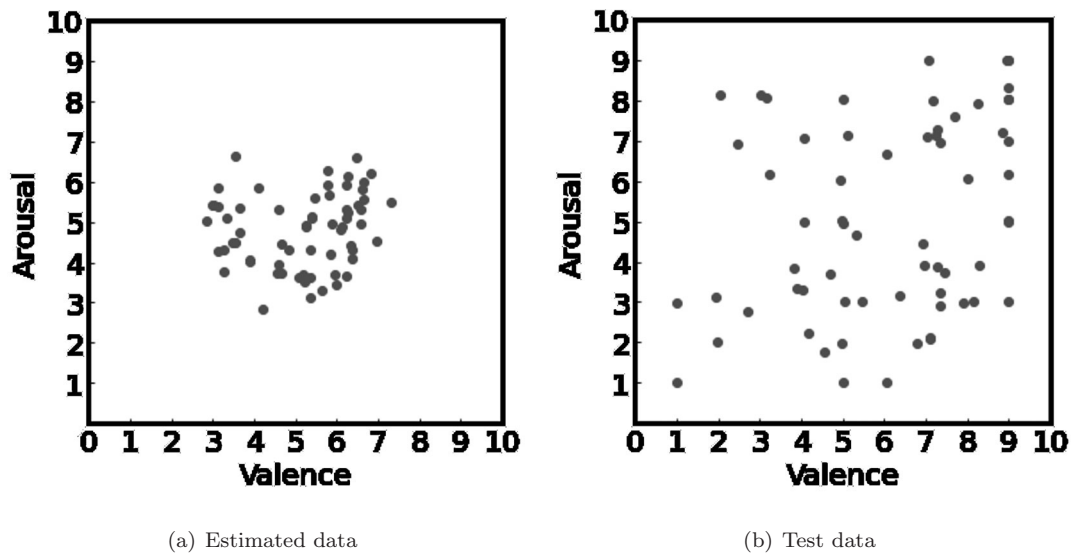


Fig. 10. Valance - Arousal.

モデルの性能評価を行った。音楽データを追加した場合、学習精度が悪化した。ValenceとArousalで評価された個人による受け止め方と、 α 波の発現の様子の乖離が大きい状態の学習データに対して、個体差に影響されず一意に決まる音楽データを追加したことで、学習が妨げられた可能性が考えられる。今後は、推定値とテストデータに付与されている真値との距離を調査することで、どのような音楽データに対して個体差の影響が大きくなっているのか検証したいと考えている。また、個体差を考慮した感情推定モデルの精度を調査し、有効性を検証したいと考えている。

本研究はJSPS科研費21K12097の助成を受けたものです。

参考文献

- 1) 鈴木琢治, 大内一成, 土井美和子, “LifeMinder ウェアラブル健康管理システム”, 電子情報通信学会技術研究報告, **101**[699], 33-38 (2002).
- 2) 後藤誠二, 峯岸由佳, 山本直樹, 森谷貴行, 中村創, 片山翔太, “カメラを用いたドライバー・モニタリングの開発”, マツダ技報, **36**, 130-135 (2019).
- 3) 金井健太郎, 鈴木圭, 菅谷みどり, “多チャンネル脳波計を用いた機械学習による4感情分類の検討”, JSKE 第23回日本感性工学会大会, 30-32 (2021).
- 4) 平松拓也, 池田悠平, 保科篤志, 馮晨, 高橋裕也, 菅谷みどり, “生体情報による感情推定手法とステージの観客反応による評価”, 分散協調とモバイルシンポジウム2017論文集, **2017**[699], 857-864 (2017).
- 5) Y. Ikeda, R. Horie and M. Sugaya, “Estimating Emotion with Biological Information for Robot Interaction”, *Procedia Computer Science*, **112**, 1589-1600 (2017).
- 6) L. Yu, L. Lee, S. Hao, J. Wang, Y. He, J. Hu, K. Lai and X. Zhang, “Building Chinese Affective Resources in Valence-Arousal Dimensions”, *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 540-545 (2016).
- 7) S. Koelstra, C. Mühl, M. Soleymani, J. 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*, **3**[1], 18-31 (2011).
- 8) B. Tang, X. Wei, G. Guo, F. Yu, M. Ji, H. Lang, and J. Liu, “The effect of odor exposure time on olfactory cognitive processing: An ERP study”, *Journal of integrative neuroscience*, **18**, 87-93 (2019).
- 9) B. McFee, C. Raffel, D. Liang, D. Ellis, M. Mcvicar, E. Battenberg and O. Nieto, “librosa: Audio and Music Signal Analysis in Python”, *Proceedings of the 14th Python in Science Conference*, 18-24 (2015).
- 10) 人見健文, 池田昭夫, “脳波の基礎知識”, 臨床神経生理学, **42**[6], 365-370 (2014).
- 11) J.A. Russell, “A Circumplex Model of Affect”, *Journal of Personality and Social Psychology*, **39**[6], 1161-1178 (1980).