

## 脳波と深層学習を活用した記憶の想起に伴う快不快感情の認識

メタデータ	言語: ja 出版者: 技術情報協会 公開日: 2020-08-20 キーワード (Ja): キーワード (En): 作成者: 森田, 純哉, 佐久間, 一輝, 野村, 太輝, 平山, 高嗣, 榎堀, 優, 間瀬, 健二 メールアドレス: 所属:
URL	<a href="http://hdl.handle.net/10297/00027608">http://hdl.handle.net/10297/00027608</a>

タイトル:脳波と深層学習を活用した記憶の想起に伴う快不快感情の認識

著者:森田 純哉,佐久間 一輝,野村 太輝,平山 高嗣,榎堀 優,間瀬 健二

## はじめに

現代に生きる人々にとって、感情の制御は切実な問題である。そもそも、感情あるいは情動は、進化の過程において、生存のために獲得された意思決定機構である<sup>1)</sup>。理性に基づく意思決定に対して、感情による意思決定は、素早く効果的な環境(獲物や外敵)への対処を可能にする。しかし、現代の多くの社会的場面で、感情による決定は非合理的な行動を導く。たとえば、会社や学校を含む組織では、様々なハラスメントが問題となっている。これらを防止するために管理職や教員、養育者には、感情の制御、たとえばアンガーマネージメントなどが求められるようになった。

感情の制御は対人場面だけでなく、個人の生活にも影響を及ぼす。ストレスなどを原因とした抑うつは典型的な感情の制御に関わる問題である。抑うつへ至る行動として、ネガティブな事柄を繰り返し想起する反芻思考が知られている。反芻思考におちいることで、気分は落ち込み、動機づけが減退する。結果として、正常な社会生活を営むことが困難になる。このような抑うつは年代を問わず、多くの現代人にとって深刻な問題となっている。勤労世代においては生産性の低下、高齢者においては生きる活力の低下に結びつく。

上記の不合理的を回避し、個人および社会の生産性を向上させるためには、人間の内側で生じる精神的なプロセスの客観的な把握が必要である。近年、人間内部の主観的な精神状態を客観化する手法が、種々検討されている。その一つとして注目されるマインドフルネス瞑想においては、人間が自分自身の身体状態(たとえば呼吸など)を知覚することによる自己の客観化が行われる<sup>2)</sup>。それに対して、アフェクティブコンピューティング<sup>3)</sup>においては、人間による主観を排した計算機を利用した感情の認識が目指される。この分野における主要な手法が、生体信号から抽出される特徴量を用いた機械学習である。そもそも、感情は、認知神経科学において、情動と呼ばれる生体的な反応の知覚と捉えられる<sup>4)</sup>。つまり、アフェクティブコンピューティングが行うことは、生体内で生じる情動に対する知覚を、人間による認知機能の代替として計算機に担わせるものである。

以上の背景から、著者らは計算機による感情の認識機構を組み入れた感情制御の研究を進めてきた。その目指すところは、個人の感情への適切な介入による生きる気力の維持と向上である。本稿では、著者らの取り組みの大枠を示し、その要素技術として検討を進めている脳波からの感情認識の試みを示す。

## 1. モデルベース回想法の構想

計算機を利用した感情の認識と制御については多様な流れが存在する。先述のアフェクティブコンピューティングに加え、ポジティブコンピューティング<sup>5)</sup>、計算論的精神医学<sup>6)</sup>など、近年においても新たな関連領域が続々と生まれている。こういった流れの中で、著者らが着目するのは記憶の想起を利用した感

情の制御である。人間の脳内において、記憶と感情を司る部位は密接に関連し、個々の記憶には様々な感情が付与される。よって、特定の記憶の想起を促すことで、対象者の感情を意図した方向に誘導できると考える。このような考えにもとづく実践は、従来、臨床心理学などの分野において、回想法と呼ばれてきた。

臨床の場で行われる回想法に対しては複数の問題が指摘されている<sup>7)</sup>。回想法においては、懐かしさを感じさせるであろう写真や音楽が、記憶の想起を促す刺激として利用される。しかし、そのような刺激に対する実際の反応は個人によって異なる。ある人にとっては親しみのある刺激が、他の人にとっては全く覚えのない刺激である可能性がある。また、たとえ知っている刺激であっても、それに付随する感情は人によって異なる。ポジティブな感情を喚起させることを意図した刺激が、ネガティブな感情を喚起させることもある。こういった問題を抑制するために、回想法の実施者には注意深い刺激の統制や対象者へのケアが求められる。回想法の根本的な問題は、そのような対象の個別性への対応が体系化されていないこととされる。

これらを踏まえ、著者らは計算機上にユーザのモデルを実装することにより、回想法に関わる問題を解決する「モデルベース回想法」の実現を目指している<sup>8)</sup>。この手法では、電子的に蓄積されたユーザの個人的な経験(ライフログ)を計算機に搭載することで、ユーザの記憶のモデルを構築する。ライフログとしては、ユーザが人生において蓄積してきた写真を想定する。これをユーザが保持する記憶の項目とみなし、モデルに基づいてユーザの感情に働きかける写真提示を行う。人間の記憶に関するモデルとして、認知科学の分野で長く検討が進められてきた ACT-R (Adaptive Control of Thought-Rational<sup>9)</sup>) と呼ばれる枠組みを想定する。ACT-R には人間の記憶の思い出されやすさに関わる様々なパラメータが含まれる。例えば、学習や忘却の速さ、文脈への影響されやすさ、類似した記憶の混同のしやすさなど、個人や状況の違いによる傾向を、可変の数値として表現する。そして、これらを調整することで、発散的な記憶の想起、特定の対象に焦点を当てた記憶の想起、ネガティブな記憶の反芻など多様な記憶の想起パターンを構築できる。モデルベース回想法においては、ACT-R における写真の提示パターンと個々のユーザが有する記憶想起の傾向を適度に同調させつつ、ネガティブな記憶の反芻をよりポジティブな記憶の想起に導くことを意図する。

認知科学の分野において ACT-R を利用する研究者は多く、多岐にわたるモデルが構築されてきた。そのなかには、感情をともなう記憶の想起に関するモデルも含まれる<sup>10-11)</sup>。そして、それらのモデルにおいて、ACT-R におけるパラメータの変動は、神経伝達物質の放出や自律神経系の働きを反映した生理反応と対応すると説明される<sup>12)</sup>。よって、ユーザから得られる生理反応を、実行中のモデルのパラメータに直接的に反映させることで、ACT-R はユーザの状態をリアルタイムに反映するモデルとなる。つまり、ユーザにモデルベース回想法の出力としての写真を観察させ、得られた生理反応によってモデルのパラメータを調整する。これによりその場のユーザに合わせた写真提示をおこなうことができると考えている。

生理反応に基づく感情の認識には様々な手法を用いることができる。著者らは、モデルベース回想法の実現に向けた実験を複数試行してきた<sup>13-16)</sup>。以下では、そのうちの簡易脳波計を用いることによる快不快感の認識に関わる研究を示す。脳波とは、脳内で生じる様々な神経活動の集積を、頭皮上の電位変化として観測するものである。生理学と心理学の境界領域としての生理心理学においては、脳波を測定部位や周波数帯域によって区別し、感情を含む多様な精神的機能と関連づける研究がなされている<sup>17-18)</sup>。一方で、複雑な脳波の情報から、精神状態を認識するために、機械学習を利用することも可能である<sup>3)</sup>。機械学習の一手法である深層学習は、人間の神経系に着想を得た技術であるが、現代では多様な対象に対して、高性能の認識を実現している。その最大の特徴はトップダウンな特徴の設計なしに認識対象となる現象の構造を学習することにある。以下に示す研究<sup>15-16)</sup>においては、生理心理学における知見と対比させつつ、深層学習ベースでの感情のモデル化の可能性を検討している。

## 2. 記憶に伴う感情と脳波のデータセット

モデルベース回想法の実現に向け、脳波からの感情認識に関わるデータセットを収集する実験を実施した。実験では、記憶想起に伴う感情の主観評価、および記憶想起時の脳波 (Electroencephalogram: EEG) をデータとして収集した。以下、実験の手続きとそこで得られた脳波の簡単な解析結果を示す。

### 2.1 方法

#### 2.1.1 実験参加者

男性 12 人 (平均年齢 32.5 歳)、女性 20 人 (平均年齢 32.2 歳) の合計 32 人 (平均年齢 32.3 歳) が実験に参加した。

#### 2.1.2 手続き

図 1 に実験全体の流れを示す。実験は「写真撮影実験」と「写真閲覧実験」に分かれた。写真撮影実験において、実験参加者は 10 日間連続して日常生活の風景を撮影することを求められた。写真撮影に利用する機材は、実験者から貸与された。また、写真撮影にあたって実験参加者は以下の指示を受けた。

1. 1 日あたり 10 枚以上の写真を撮影すること。
2. 10 時から 22 時までの間の 2 時間ごとに撮影を行うこと。
3. 撮影期間中に写真の見返しを行わないこと。

撮影期間終了後、平均 10 日 (範囲 8 - 10 日) の期間を空けて写真閲覧実験を実施した。写真閲覧実験では、写真収集実験で撮影した写真を参加者に提示し、写真撮影時の記憶を想起させた。なお、著者らが最終的に目指すモデルベース回想法において、写真の選択は個人化された記憶のモデルによりなされる。今回の実験においては、脳波からの感情の認識に焦点をあてるため、モデルを用いずランダムな写真の提示順とした。また、記憶に基づく感情と脳波の対応を検討するために実験参加者本人が撮影し

た 90 枚に、他の実験参加者が撮影した 90 枚を加えた 180 枚を実験刺激とした。ただし、この条件の差異については分析の対象としなかった。

<図 1 挿入箇所

<図 2 挿入箇所

写真閲覧実験では、4 秒間の注視点画像の提示直後に撮影した写真を 4 秒間提示した。参加者は、各写真の閲覧直後に、写真を閲覧した際に生じた感情を主観評定 (5 段階) によって答えた。感情の評定項目は快不快と覚醒度であった。これらの項目は、感情の状態を二軸で表現するラッセルの円環モデル<sup>19)</sup>に従うものである。

写真閲覧中の脳波の収集には、Emotiv 社製の脳波計測装置である Emotiv Epoc を用いた。Emotiv Epoc は、左右に 7 ずつ 14 個の電極を有しており、それぞれの電極からサンプリング周波数 128Hz の脳波が計測される。通常の脳波計の利用では、粘着性のペーストを利用することで電極を一つ一つ頭皮に貼り付ける。それに対して、Emotiv Epoc は、電極が事前に配置されたヘッドギアを頭部に装着するだけで脳波を計測できる。そのため脳波計測の手続きが簡素化され、モデルベース回想法の実利用に向けた有効な装置と考えた。

写真の提示には、15.6 型 (1920 x 1080 pixels) のノート PC のディスプレイを使用した。実験参加者とディスプレイの距離は約 800mm であった。上記の実験の風景を示す写真を図 1b に示す。

## 2.2 データセットの概要

実験に参加した全実験参加者の全写真閲覧を通して、4998 の各感情評定値が得られた。図 2 の「評定値全体」と記されたバーは、快不快と覚醒度の評定値の分布を示している。快不快の評定が概ね左右対称の分布となっているのに対して、覚醒度の分布は左に偏っていることがわかる。このような分布の特性も鑑み、本稿では快不快に焦点を当て、写真閲覧時の感情を認識する深層学習モデルを設計し、その性能を検証する。

深層学習モデルの構築に先立ち、体動や眼球運動などによるノイズの影響の低減を目的とし、前処理を行った。初めに各参加者の実験全体に渡って収集された脳波に対して [1-50Hz] の FIR フィルタを適用した。次に、各参加者の脳波データ中に含まれる写真閲覧区間をトライアルとし、写真閲覧開始前 3 秒から閲覧開始後 4 秒まで区間を抽出した。その後、抽出された全トライアルに対して外れ値除去を行った。ここでは、[-500  $\mu$ V, 500  $\mu$ V] の範囲を超える電位を含むトライアルを除外した。図 2 の「前処理後」と記されたバーは、トライアル除外後の感情評定値の分布を示している。これらのノイズ処理ののちに、EEGLAB<sup>20)</sup> の 3 周期のマザーウェーブレットによるウェーブレット変換を利用し、シータ波 (4-7Hz)、アルファ波 (8-13Hz)、ベータ波 (14-30Hz) 及びガンマ波 (31-50Hz) の各周波数帯域成分を算出した。

## 2.3 快不快での平均的な脳波の比較

本研究のデータセットに感情に伴う脳波の変化が含まれていることを確かめるため、快不快の感情別での脳波の平均値を比較した。生理心理学における多くの先行研究で、快感情では前頭部シータ波とアルファ波が活性化することが指摘されている<sup>17-18)</sup>。特に、快感情が喚起された時に前頭部左部のシータ波とアルファ波が活性化するのに対し、不快感情が喚起されたときには前頭部右部のアルファ波とシータ波が活性化するなどの知見が報告されている。また、生理心理学では何らかのイベントの発生に伴う脳波の反応傾向を見る際に、事象関連電位と呼ばれる時系列的な脳波の推移を検討することが多い。この推移を比較可能な形で検討するために、図3に快不快それぞれの脳波の時系列推移を部位別、周波数帯域別に示した。実線によって快不快のそれぞれの平均値を示し、その上下の領域に信頼区間(信頼度 95%)を示している。

<図3 挿入箇所

結果、シータ波やアルファ波といった低周波帯域において、快と不快の間での推移の差異が認められた。特に、シータ波では AF3, F7, F4, FC5, P8, O2, アルファ波では AF3, FC5, P7 の電極位置で快の場合と不快の場合での推移が大きく異なった。また、AF3-AF4, F7-FC8, FC5-FC6 といった左右対称の電極位置に注目すると、前頭部左部では前頭部右部に比べて快不快で推移の差が強く出ている。以上の結果から、低周波成分における前頭部左部の活性、あるいは前頭部での電位変化の非対称性が、快不快の感情状態と関連するという過去の研究<sup>17-18)</sup>と一貫した結果が得られた。さらに、電位の時間変化に着目しても、シータ波とアルファ波で共通する傾向を確認できた。不快の場合には刺激の提示に伴う電位(事象関連電位)のピーク後に電位が大きく減少したのに対し、快の場合には電位は大きく減少せず、一定の電位を維持した。この結果は、感情と関連した刺激によって誘発される事象関連電位の特徴(Late Positive Potential)<sup>21)</sup>と一貫している。

## 3. 深層学習による快不快感情の認識

上記の平均値の比較から、提示される写真に対して喚起される感情が、脳波という生理的な信号に反映されることが示される。以下では、ここまで得られた感情評定と脳波を利用し、感情認識の機械学習モデルを構築する。本稿では特に深層学習を利用した感情認識の手法を提案し、その性能を既存の手法であるサポートベクトルマシン(Support Vector Machine: SVM)と比較する。

### 3.1 ネットワーク構成

図3の分析に示されるように、生理心理学分野の解析では、人間の精神活動との関連を探る際に、脳波の周波数帯域と電極位置(部位)を区別しつつ、脳波の時間的な変化を検討する。よって、脳波による感情認識を行う際に、空間軸、周波数軸、時間軸に関して独立に特徴を抽出することは理に適っている。こ

の考えのもと、本研究では、空間軸、周波数軸、時間軸の各軸に対応した畳み込み層を有する Convolutional Neural Network (CNN) を設計した。図 4 にネットワーク構成の概要とネットワークへの入力を示す。

<図 4 挿入箇所

入力特徴は、*Depth* を 1, *Width* を時間軸のサンプル数 384, *Height* を電極数  $N_c$  と周波数帯域の数  $N_f$  の積  $N_f * N_c$  とする 2 次元の画像として表現する。この入力に対して、時間軸・空間軸・周波数帯域軸の順に畳み込む T-S-F-Net, 空間軸・周波数帯域軸・時間軸の順に畳み込む S-F-T-Net, 時間軸・周波数帯域軸・空間軸の順に畳み込む T-F-S-Net 及び周波数帯域軸・空間軸・時間軸の順に畳み込む F-S-T-Net の軸間で畳み込み順が異なる 4 種類の CNN を設計した。各 CNN の設計に用いた具体的なパラメータを表 1 から 4 に示す。

<表 1 挿入箇所

<表 2 挿入箇所

<表 3 挿入箇所

<表 4 挿入箇所

畳み込み順の操作に加えて、2.3 にて述べた生理心理学的知見を利用し、周波数帯域成分と電極位置に関わる入力特徴を抽出することを試みた。全電極から得られた全ての周波数帯域成分を入力とする ALL × ALL, 全電極から得られた低周波数成分のみを入力とする  $\theta \alpha \times$  ALL, 前頭部から得られた低周波数成分のみを入力とする  $\theta \alpha \times$  FRONTAL, 図 3 において快不快の差が大きかった部位での低周波数成分のみを入力とする  $\theta \alpha \times$  EFFECT をそれぞれ設計した。各特徴の概要について表 5 に示す。構成した畳み込み順の異なる 4 種類の CNN に対して、4 種類の入力特徴をそれぞれ用いた際の識別性能について比較する。更に、深層学習ベースの感情認識の有効性を検証する上でのベースラインとして、機械学習分野の古典的な手法である SVM と比較する。SVM の入力には、4 種類の入力特徴の時間軸方向のみの平均(それぞれ 56, 28, 16 及び 16 次元)を用いた。

<表 5 挿入箇所

### 3.2 感情認識の方法

図 2 の快不快感情の評定から、不快・快・平常の 3 クラスの識別問題を設定する。各クラス 609 個のデータを評価に用い、その内の 4/5 を学習用データ (内 3/4 を学習データ, 1/4 を検証データ), 1/5 をテストデータとした 5 分割交差検証を行った。また、入力特徴については事前に全成分の平均と標準偏差を用いた Z スコアにより標準化した。さらに、平滑化, ノイズ付加, および抽出する時間をずらすタイムクロッピングを適用し、元の学習データを 45 倍に水増した。各層のパラメータの初期値は、各層の入力の大きさ  $N_i$  (CNN の場合  $N_i = Width * Height * Depth$ ) により、 $\mu = 0, \sigma = \sqrt{2/N_i}$  の正規分布に従う乱数で初期化し、学習

率  $\alpha = 5.0 * 10^{-6}$  の Adam を用いて学習させた。一方で、SVM ではグリッドサーチによるパラメータ探索により、最適なハイパーパラメータを導出し、各交差検証で最適パラメータによる学習後の SVM をテストデータによる評価に用いた。

### 3.3 感情認識の結果

各入力各手法における識別性能 (accuracy) の比較結果を図 5 に示す。快・平常・不快の 3 クラスを分類する偶然の確率 (チャンスレート) は 33.33% である。全ての深層学習モデルにおいて識別性能は、この確率を超えた。このことから、脳波からの深層学習によって快不快感情を偶然ではない確率で認識できることが示された。一方で、SVM の識別性能の平均は 49.73% となっており、どの深層学習モデルも 6% 近く識別性能で劣った。このことから今回の CNN による感情認識は、既存手法に比べ有効とは言えない結果となった。

また、各深層学習モデル間の識別性能に有意差はなく畳み込み順による識別性能への影響はなかった。制約を設けない入力特徴である ALL × ALL は制約を設けた入力特徴よりも有意に優れた識別性能を示した。このことから深層学習による特徴量の選択において、トップダウン的な絞り込むこみを行うことは、脳波が元々保持する識別に有用な情報の損失を招きうることを示された。

<図 5 挿入箇所

## 4 深層学習による感情を伴う脳部位の可視化

前項の検討からは、他の手法に比べた深層学習の利点を見出すことはできなかった。しかし、深層学習のメリットは、ただの識別器としてだけでなく、学習結果の可視化を行える点にもある。最近では、学習済みモデルの中間層を可視化することによる説明可能な AI の可能性が盛んに検討されている。本項では、その流れを踏襲し、感情を伴う脳部位の可視化を行う。そのことにより、深層学習による感情認識において、妥当な脳部位への注目がなされていることを確かめる。

### 4.1 可視化の方法

Ramprasaath らは CNN の可視化手法として、Gradient-weighted Class Activation Mapping (Grad-CAM) と Guided Backpropagation (Guided BP) を組み合わせた Guided Grad-CAM (GGC) を示している<sup>22)</sup>。本稿では、この手法を前項の深層学習モデルに適用し、中間層の可視化と集約を行う。手法の概要を図 6 に示す。以下、入力と可視化と集約の各手順について述べる。

<図 6 挿入箇所

#### 4.1.1 入力

学習済みモデルによって正しく感情を認識できた脳波の信号 (2 次元の画像) を入力する。

#### 4.1.2 可視化

出力層における、可視化したいクラスの値を1, その他のクラスの値を0としパラメータを更新せずに誤差逆伝播 (Backpropagation) を実行する. Guided BP とは, この操作を行った後の入力層の勾配の値である. Grad-CAMはCNN内の全結合層の直前の畳み込み層が出力した特徴マップを用いて出力する. 各特徴マップの勾配の平均値と, 各特徴マップの重みを掛け合わせたものに ReLU 関数を適用する. 作成した Guided BP と Grad-CAM の積により GGC が構成される.

#### 4.1.3 集約

クラス毎に, 作成した GGC を平均した画像を作成する. これにより, 個別の状況をまたいだクラスの特徴が抽象化される.

<図7挿入箇所

### 4.2 可視化の結果

交差検証において最高の識別性能を示した深層学習モデル (ALL × ALL) に対し, GGC によって得られたヒートマップを図 7a に示す. これらのヒートマップは, 快, 平常, 不快のそれぞれのラベルが付与された脳波について, GGC によって特徴を抽象化したものである. 深層学習モデルへの入力(図4を参照)と対応し, 図の横軸が時間軸, 縦軸が空間軸と周波数帯域軸と対応している. 空間軸に関しては前頭部左部を始点として頭皮を周回する形に並べられている (AF3, F7, F3 ... F8, AF4). 図より, 快の場合にはヒートマップの下半分(右半球に相当するチャンネル), 不快の場合にはおよそ上半分(左半球に相当するチャンネル)が高い値になっていることがわかる. このことをより直感的に理解するために, 図3のヒートマップについて, 時間軸の値を平均することで, 頭皮上での注目強度を示すトポマップ<sup>23)</sup>を作成した(図7b). トポマップの作成において, 平均化する周波数帯域はアルファ波に限定した. 図に示されるように, 本研究において学習された深層学習モデルは快と不快にラベルづけされた脳波に対し, 左右で異なる部位に注目していることがわかる. このような快不快の感情状態を区別する脳部位の左右差は, 従来の生理心理学において指摘されている知見と対応するものといえる<sup>17-18)</sup>.

この結果が, 深層学習による学習の結果であることを確認するため, 未加工のアルファ波の脳波データを図3と同様の形式で平均化した画像(図7c), およびそのトポマップ(図7d)を作成した. これらより, 未加工の脳波データの単純な平均画像においては, 快と不快を区別する明確な特徴を観察することが困難であることが確認される.

#### まとめ

本稿では, 感情の把握と制御の必要を背景とし, モデルベース回想法の実現にむけた感情認識の研究を示した. まず, 日常的な記憶に対する感情評定, および記憶の想起にともなう脳波を収集する参加者実験を実施した. 得られたデータを利用することで, 複数の深層学習モデルを構築し, 快不快感情の

認識における性能を、既存の機械学習モデルである SVM と比較した。結果として、トップダウン的な特徴選択を行なった深層学習モデルは、そのような特徴選択を行わなかったモデルに比べ、優れた性能を示さなかった。また、どの深層学習モデルも SVM に対して劣る性能となった。その一方で、深層学習モデルを対象とした可視化の結果は、生理心理学的な知見と一致する構造(低周波成分における左右差)を学習していることを示した。

上記の結果によって、深層学習による脳波からの感情認識の特性が示される。深層学習による可視化は、図 3 で観察されたクラス間での平均値の差異を増強するものと見ることができ、人手によるデータの解釈を自動的に遂行するものとみることができる。一方で、平均値における差異の大きかった特徴を選択的に入力とすることは、深層学習の認識性能の低下を招いた。これらから、深層学習は現象を区別する構造的な特徴の発見には有効ではあるが、そのプロセスに人手による補助を組み合わせることは容易ではないと結論づけることができる。

また、認識性能に関して、深層学習モデルは十分な性能を示さなかった。しかし、このことは本研究におけるデータセット収集の特殊性による可能性がある。深層学習による脳波からの高い精度の感情認識を示す研究<sup>24)</sup>の多くにおいては、高価な装置を利用して厳密に計測された脳波が用いられる。また、本研究と同様に簡易脳波計を用いた研究においても、本研究のように参加者の自発的な感情の評定を予測するのではなく、一般的に快いとされる刺激(音もしくは画像)、あるいは不快とされる刺激(音もしくは画像)の提示に伴う脳波の分類が行われる。著者らの知る限り、記憶の想起に伴う自発的な感情の認識を、脳波からの深層学習によって検討した研究は存在しない。

上記のような状況を考えれば、今回の研究における認識性能の低さは、深層学習の問題というよりは、認識対象の難しさに由来する可能性もある。認知神経科学において、感情は情動の知覚であるとされる。情動が身体反応を直接的に表すのに対して、主観的な働きである感情は、特定の脳部位との対応をとることが困難であることも指摘されている<sup>26)</sup>。従来の生理心理学において指摘されている前頭部の左右差についても、主観的な感情評定との相関が指摘されているに過ぎず、根本的な感情の発生部位を示しているわけではない。

機械学習に関する従来の研究では、通常は外的に定められたラベルに対する予測性能を向上させることが目指されてきた。しかし、感情認識のような絶対的な正解のない問題に対して、正確な予測を競うことの意味は明白ではない。そもそも著者らが目指すのは、計算機を利用した感情の制御である。この目的を達成するためには、特定の時点で人手によって付与されたラベルを予測するのではなく、個々の人間が求める **Well-being** に適応しつつ、人間による情動の認識(感情)と協同する方法を考えるべきである。そして、これを実現するためには、人間にとって納得のいく可視化を求めつつ、計算機技術の発展による人間自身の変化をも見通していく必要がある。

## 謝辞

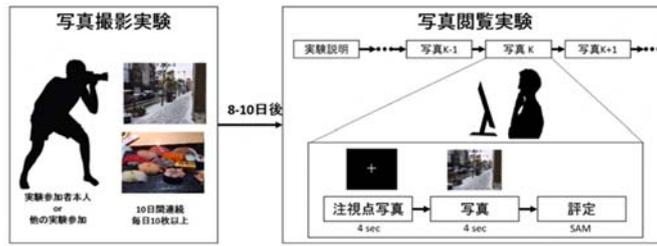
ここで示した研究は独立行政法人科学技術振興機構 (JST) の研究成果展開事業「センター・オブ・イノベーション (COI) プログラム」, JSPS 科研費 17H05859, 15H01615 などの助成を受けた。

## 参考文献

- 1) C. Darwin, *The Expression of the Emotions in Man and Animals*, New York: D. Appleton And Company (1899)
- 2) J. Kabat-Zinn, *Full catastrophe living: using the wisdom of your body and mind to face stress, pain, and illness*, Delta Trade Paperbacks, 1991
- 3) R. Picard, *Affective Computing*, MIT Technical Report #321 (1995)
- 4) A. Damasio, *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*, Mariner Books (2003)
- 5) R. Calvo and P. Dorsian, *Positive Computing: Technology for Well-being and Human Potential*, The MIT Press (2014)
- 6) P. R. Montague, R. J. Dolan, K. J. Friston, P. Dayan, *Computational psychiatry*, *Trends in Cognitive Sciences*, 16, pp. 72-80, 2012
- 7) 瀧川真也, なつかしさと記憶, 臨床的応用, 懐かしさの心理学-思い出と感情(楠見孝編), 誠信書房, pp.23-40 (2014).
- 8) 森田純哉, 平山高嗣, 間瀬健二, 山田和範, *メンタルタイムトラベルを誘導するモデルベース回想法*, 研究報告ユビキタスコンピューティングシステム (UBI), Vol. 2015, No. 16, pp. 1-6 (2015).
- 9) J. R. Anderson, *How can the human mind occur in the physical universe?* Oxford University Press (2009)
- 10) I. Juvina, O. Larue, A. Hough, *Modeling valuation and core affect in a cognitive architecture: The impact of valence and arousal on memory and decision-making*, *Cognitive Systems Research*, 48, 99. 4-24 (2018)
- 11) M. V. Vugt, M. van der Velde, and *ESM-MERGE Investigators*. *How does rumination impact cognition? A first mechanistic model*. *Topics in cognitive science*, 10, pp. 175-191 (2018)
- 12) C. L. Dancy, *ACT-RΦ: A cognitive architecture with physiology and affect*, *Biologically Inspired Cognitive Architectures*, 6, pp. 40-45 (2013)
- 13) 森田純哉, 平山高嗣, 間瀬健二, 山田和範, *ライフログ写真と認知アーキテクチャを利用したモデルベース回想法: ケーススタディの報告*, 日本認知科学会第33回大会発表論文集, pp. 443-447 (2016)
- 14) 野村太輝, 森田純哉, 平山高嗣, 榎堀優, 間瀬健二, *CNNによる感情認識における生理心理的制約の効果*, 第32回人工知能学会全国大会論文集 4D1-OS-14c-04 (2018)
- 15) 佐久間一輝, 森田純哉, 野村太輝, 平山高嗣, 榎堀優, 間瀬健二, *深層学習の可視化による神経科学的知見の抽出*, 第32回人工知能学会全国大会論文集, 3G3-OS-18a-04 (2019)

- 16) 板橋和希, 森田純哉, 平山高嗣, 間瀬健二, 山田和範, モデルベース回想法における個人差の検討, 日本認知科学会第36回大会発表論文集, pp. 1012-1018 (2019)
- 17) R. J. Davidson, Affective neuroscience and psychophysiology: Toward a synthesis, *Psychophysiology*, Vol. 40, No. 5, pp. 655–665 (2003)
- 18) M. Sarlo, G. Buodo, S. Poli, and D. Palomba, Changes in EEG alpha power to different disgust elicitors: the specificity of mutilations, *Neuroscience letters*, Vol. 382, No. 3, pp. 291–296 (2005)
- 19) J. A. Russell, A circumplex model of affect, *Journal of personality and social psychology* 39, (1980)
- 20) D. Arnaud and M. Scott, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134, pp. 9–21 (2004)
- 21) T. A. Dennis and G. Hajcak, The late positive potential: a neurophysiological marker for emotion regulation in children. *Journal of Child Psychology and Psychiatry*, 50(11):1373–1383 (2009)
- 22) R. Ramprasaath, D. Abhishek, V. Ramakrishna, C. Michael, P. Devi, and B. Dhruv, Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization, *CVPR 2016* (2016)
- 23) A. Gramfort, M. Luessi, E. D. E. Larson, C. Strohmeier, D. Brodbeck, and M. Parkkonen, H. L. and, MNE software for processing MEG and EEG data, *Neuroimage*, Vol. 86, No. 1, pp. 440–460 (2014)
- 25) S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset. In *Twenty-Ninth IAAI Conference* (2017)
- 24) R. M. Mehmood, D. Ruoyu, and J. L. Hyo, Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors, *IEEE Access* 5, pp. 14797-14806 (2017)
- 24) 大平英樹, 木村健太, 白井真理子, 藤原健, 感情の心理学的構成主義に見るこれからの感情研究, *エモーション・スタディーズ*, 3, pp.38-51 (2017)

a. 実験手続き



b. 写真閲覧実験の風景



図1. データセットを収集するための参加者実験.

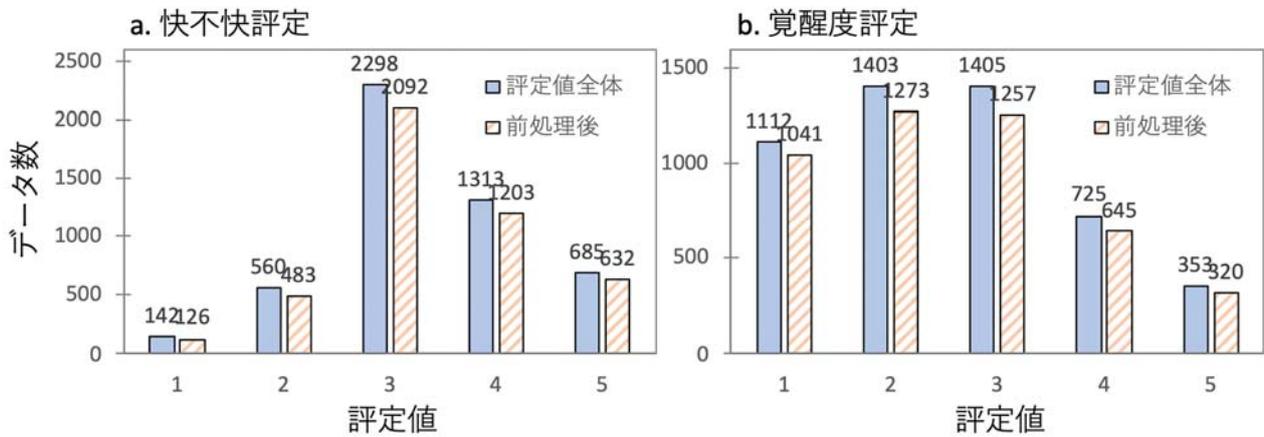


図2. 評定値の分布.

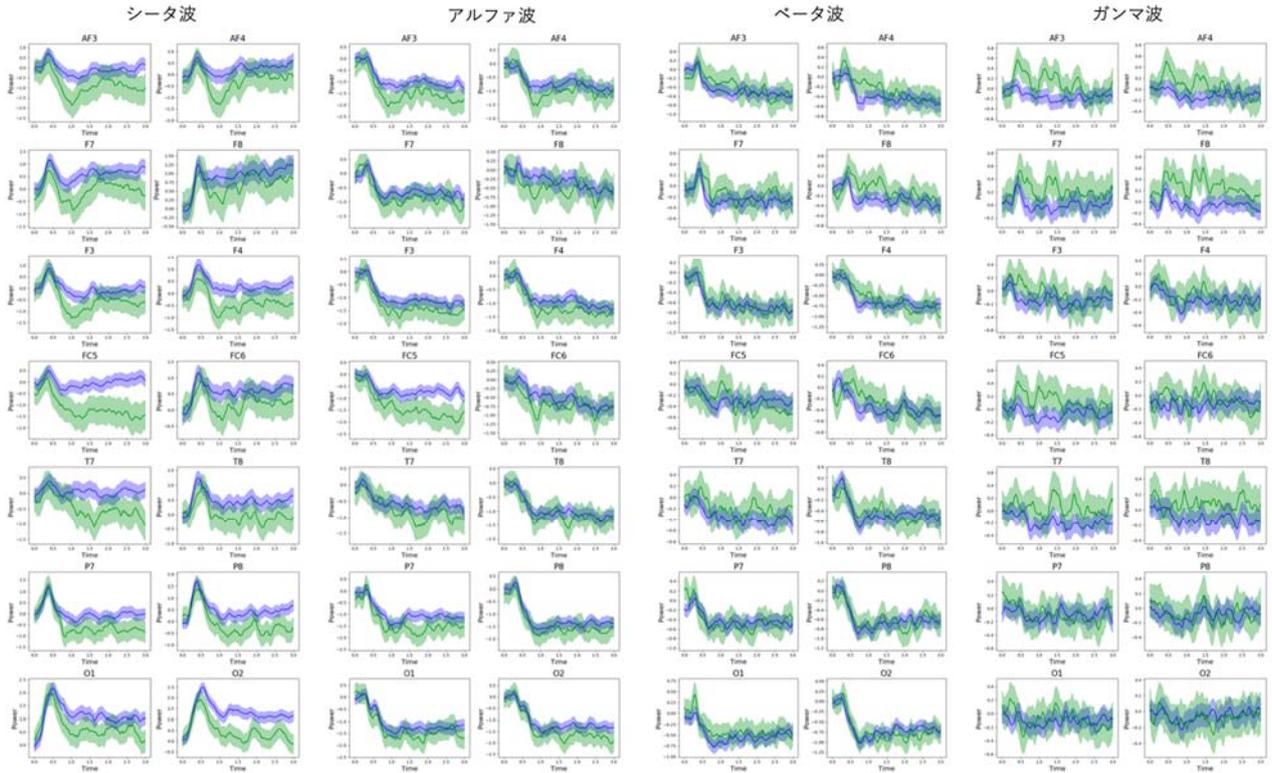


図 3. 各電極位置での  $\beta$  波の推移傾向. 青色が快 ( $n=609$ ), 緑色が不快 ( $n=1835$ ). 実線の周囲の領域は 95%信頼区間を表す.

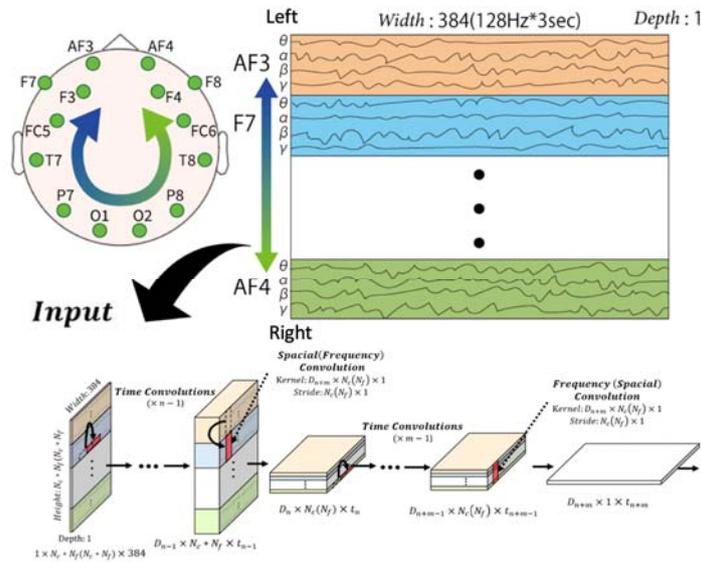


図 4. 各軸間で独立に特徴抽出する入力特徴と畳み込み層の設計. 空間的な畳み込み (Spatial), 時間的な畳み込み (Time), 周波数の畳み込み (Frequency) は相互に順序を入れ替えて配置可能. 14) より再掲.

表 1 TimeSpacialFreqNet のパラメータ

タイプ	構造	次元
入力	Depth $\times$ Height ( $N_f \times N_c$ ) $\times$ Width (Times)	$1 \times N_f \times N_c \times 384$
Time Conv 1	Kernel: $8 \times 1 \times 12$ , Stride: $1 \times 3$ , pad=0, elu	$8 \times N_f \times N_c \times 125$
Time Conv 2	Kernel: $8 \times 1 \times 12$ , Stride: $1 \times 3$ , pad=0, elu	$8 \times N_f \times N_c \times 38$
Time Conv 3	Kernel: $8 \times 1 \times 12$ , Stride: $1 \times 3$ , pad=0, elu	$8 \times N_f \times N_c \times 9$
Time Pool	Kernel: $1 \times 3$ , Stride: $1 \times 3$ , pad=0, Max	$8 \times N_f \times N_c \times 3$
Batch Norm 1	Dimensions: 8	$8 \times N_f \times N_c \times 3$
Dropout 1	Wight Decay: 0.5	$8 \times N_f \times N_c \times 3$
Spatial Conv	Kernel: $32 \times N_c \times 1$ , Stride: $N_c \times 1$ , pad=0, elu	$32 \times N_f \times 3$
Frequency Conv	Kernel: $32 \times N_f \times 1$ , Stride: $N_f \times 1$ , pad=0, elu	$32 \times 1 \times 3$
Batch Norm 2	Dimensions: 32	$32 \times 1 \times 3$
Dropout 2	Wight Decay: 0.5	$32 \times 1 \times 3$
FC 1	96, elu, Dropout: 0.5	$96 \times 1$
FC 2	Classes, Softmax	Classes $\times$ 1

表 2 SpacialFreqTimeNet のパラメータ

タイプ	構造	次元
入力	Depth $\times$ Height ( $N_f \times N_c$ ) $\times$ Width (Times)	$1 \times N_f \times N_c \times 384$
Spatial Conv	Kernel: $16 \times N_c \times 1$ , Stride: $N_c \times 1$ , pad=0, elu	$16 \times N_f \times 384$
Frequency Conv	Kernel: $16 \times N_f \times 1$ , Stride: $N_f \times 1$ , pad=0, elu	$16 \times 1 \times 384$
Batch Norm 1	Dimensions: 16	$16 \times 1 \times 384$
Dropout 1	Wight Decay: 0.5	$16 \times 1 \times 384$
Time Conv 1	Kernel: $16 \times 1 \times 12$ , Stride: $1 \times 3$ , pad=0, elu	$16 \times 1 \times 125$
Time Conv 2	Kernel: $16 \times 1 \times 12$ , Stride: $1 \times 3$ , pad=0, elu	$16 \times 1 \times 38$
Time Conv 3	Kernel: $32 \times 1 \times 12$ , Stride: $1 \times 3$ , pad=0, elu	$32 \times 1 \times 9$
Time Pool	Kernel: $1 \times 3$ , Stride: $1 \times 3$ , pad=0, Max	$32 \times 1 \times 3$
Batch Norm 2	Dimensions: 32	$32 \times 1 \times 3$
Dropout 2	Wight Decay: 0.5	$32 \times 1 \times 3$
FC 1	96, elu, Dropout: 0.5	$96 \times 1$
FC 2	Classes, Softmax	Classes $\times$ 1

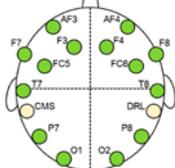
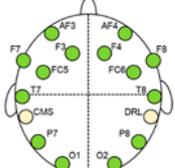
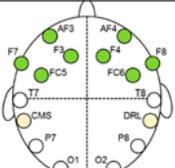
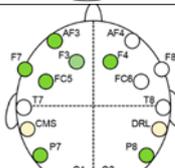
表 3 TimeFreqSpacialNet のパラメータ

タイプ	構造	次元
入力	Depth $\times$ Height ( $N_f \times N_c$ ) $\times$ Width (Times)	$1 \times N_f \times N_c \times 384$
Time Conv 1	Kernel: $8 \times 1 \times 12$ , Stride: $1 \times 3$ , pad=0, elu	$8 \times N_f \times N_c \times 125$
Time Conv 2	Kernel: $8 \times 1 \times 12$ , Stride: $1 \times 3$ , pad=0, elu	$8 \times N_f \times N_c \times 38$
Time Conv 3	Kernel: $8 \times 1 \times 12$ , Stride: $1 \times 3$ , pad=0, elu	$8 \times N_f \times N_c \times 9$
Time Pool	Kernel: $1 \times 3$ , Stride: $1 \times 3$ , pad=0, Max	$8 \times N_f \times N_c \times 3$
Batch Norm 1	Dimensions: 8	$8 \times N_f \times N_c \times 3$
Dropout 1	Wight Decay: 0.5	$8 \times N_f \times N_c \times 3$
Frequency Conv	Kernel: $32 \times N_f \times 1$ , Stride: $N_f \times 1$ , pad=0, elu	$32 \times N_c \times 3$
Spatial Conv	Kernel: $32 \times N_c \times 1$ , Stride: $N_c \times 1$ , pad=0, elu	$32 \times 1 \times 3$
Batch Norm 2	Dimensions: 32	$32 \times 1 \times 3$
Dropout 2	Wight Decay: 0.5	$32 \times 1 \times 3$
FC 1	96, elu, Dropout: 0.5	$96 \times 1$
FC 2	Classes, Softmax	Classes $\times$ 1

表 4 FreqSpacialTimeNet のパラメータ

タイプ	構造	次元
入力	Depth $\times$ Height ( $N_f \times N_c$ ) $\times$ Width (Times)	$1 \times N_f \times N_c \times 384$
Frequency Conv	Kernel: $16 \times N_f \times 1$ , Stride: $N_f \times 1$ , pad=0, elu	$16 \times N_c \times 384$
Spatial Conv	Kernel: $16 \times N_c \times 1$ , Stride: $N_c \times 1$ , pad=0, elu	$16 \times 1 \times 384$
Batch Norm 1	Dimensions: 16	$16 \times 1 \times 384$
Dropout 1	Wight Decay: 0.5	$16 \times 1 \times 384$
Time Conv 1	Kernel: $16 \times 1 \times 12$ , Stride: $1 \times 3$ , pad=0, elu	$16 \times 1 \times 125$
Time Conv 2	Kernel: $16 \times 1 \times 12$ , Stride: $1 \times 3$ , pad=0, elu	$16 \times 1 \times 38$
Time Conv 3	Kernel: $32 \times 1 \times 12$ , Stride: $1 \times 3$ , pad=0, elu	$32 \times 1 \times 9$
Time Pool	Kernel: $1 \times 3$ , Stride: $1 \times 3$ , pad=0, Max	$32 \times 1 \times 3$
Batch Norm 2	Dimensions: 32	$32 \times 1 \times 3$
Dropout 2	Wight Decay: 0.5	$32 \times 1 \times 3$
FC 1	96, elu, Dropout: 0.5	$96 \times 1$
FC 2	Classes, Softmax	Classes $\times$ 1

表 5 設定した4種類の入力の特徴

特徴	周波数帯域	電極位置	概要
ALL×ALL	シータ波, アルファ波, ベータ波, ガンマ波		4種類全ての周波数帯域の全ての電極位置。(次元数:1×56×384)
θα×ALL	シータ波, アルファ波		θ波, α波の全ての電極位置。(次元数:1×28×384)
θα×FRONTAL	シータ波, アルファ波		快不快の主観的状态と特に関連が深いと言われるθ波とα波の前頭葉の8つの電極位置(AF3, F7, F3, FC5, FC6, F4, F8及びAF4)(次元数:1×16×384)
θα×EFFECT	シータ波, アルファ波		快不快の主観的状态と関連を持つと言われるθ波とα波のCo-henの効果量の大きかった8つの電極位置(AF3, F7, F3, FC5, P7, O2, P8及びF4)。(次元数:1×16×384)

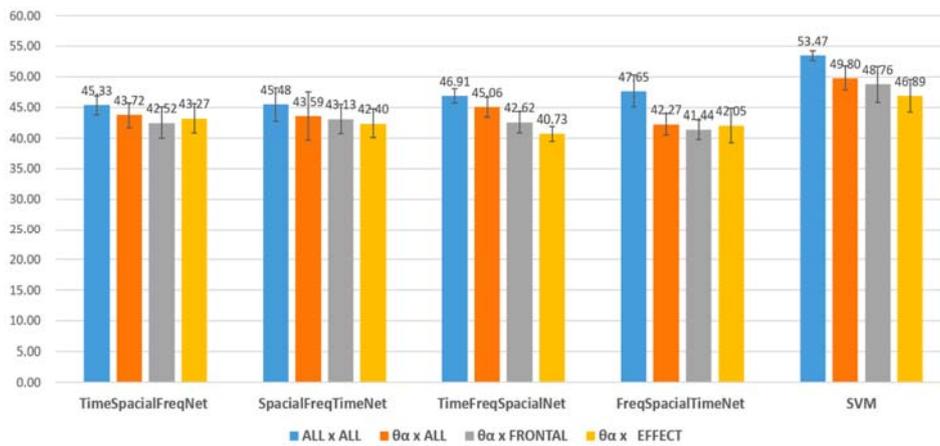


図 5. 認識性能 (accuracy) の比較. 14)より再掲.

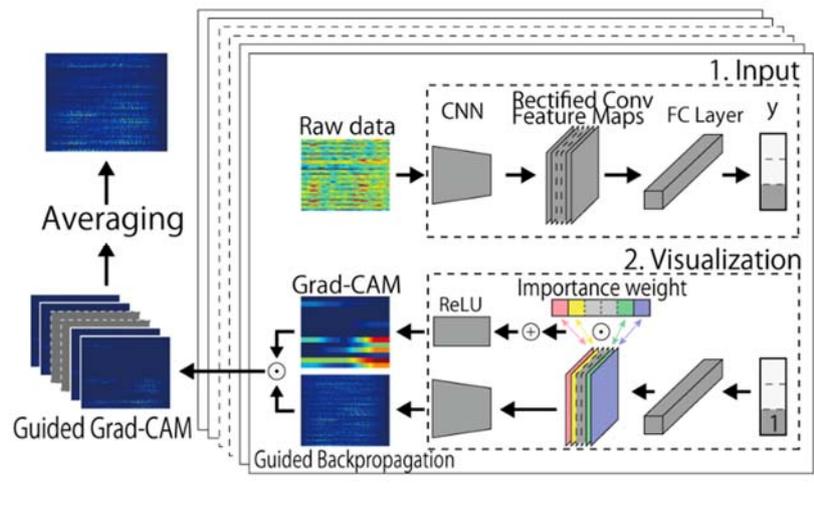


図 6. Guided Grad-CAM の手続き. 15)より再掲.

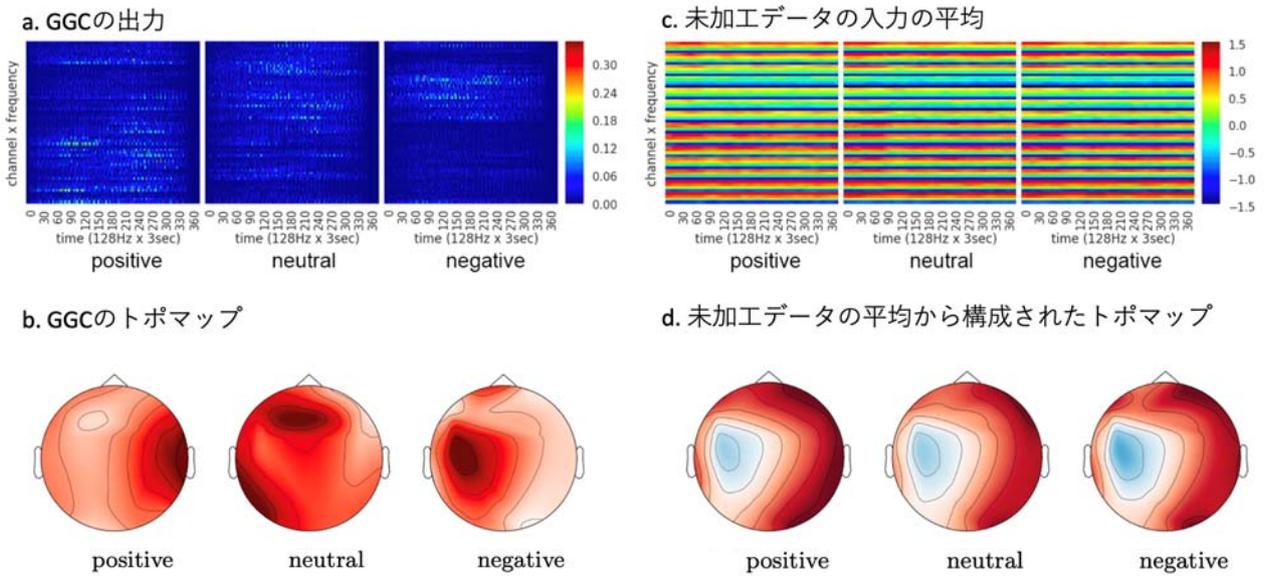


図 7. 可視化の結果. 15)より改変して再掲.