

NIRS を用いたサムネイル画像が視聴者へ与える印象分類による スナップショット抽出に関する研究

竹内 広樹^{†1} 金井 秀明^{†2}

概要: NIRS は近赤外光により外側皮質表面上のヘモグロビンの濃度の変化を測定することができる装置である。この装置により、従来の感情評価方法であるアンケートには表れない深層心理や咆哮性などを評価することができる。本研究は映画の予告動画のサムネイル画像を対象とし、NIRS によって視聴者に与える印象を分類するとともに、予告動画内からサムネイル画像に適したスナップショットを抽出する。

キーワード: NIRS, サムネイル画像, スナップショット, ニューラルネットワーク

Study on extraction of snapshot by impression classification given to viewers by thumbnail images using NIRS

HIROKI TAKEUCHI^{†1} HIDEAKI KANAI^{†2}

Abstract: NIRS is a device that can measure changes in the concentration of hemoglobin on the outer cortical surface by near infrared light. With this device, it is expected that it can evaluate deep psychology and roaring property etc. which does not appear in questionnaire which is a conventional emotion evaluation method. In this research, thumbnail images of preview movies of movies are targeted, classified impressions given to viewers by NIRS, and snapshots suitable for thumbnail images are extracted from the advance notice movie.

Keywords: NIRS, Thumbnail Image, Snapshot, Neural Network

1. はじめに

今日、市場分析は様々な分野で推進されている。市場分析の重要な課題は、消費者の満足度と商品のニーズを高めるために、消費者の心理および行動を理解することである。

従来の市場分析は主にアンケートやインタビューによって行われてきた。これらの方法はデータを簡単に収集できるメリットがある。一方で、消費者は質問を不正確に解釈し、意図的に回答している可能性がある。

近年、ニューロマーケティングが注目されている。ニューロマーケティングとは、脳の活動や状態を特殊な機械により測定することで、アンケートやインタビューなどの主観評価では現れない深層心理や咆哮性などを評価するマーケティング手法である。ニューロマーケティングにより、企業側は、消費者による意図的な操作が行えない情報を得ることができる。脳の活動や状態を測定する代表的な装置として fMRI (functional magnetic resonance imaging: 磁気共鳴機能画像法), NIRS (near - infrared spectroscopy: 近赤外線分光法), EEG (Electroencephalogram: 脳波) などが挙げられる。

fMRI はニューロイメージング技術であり、脳全体の活動を高い解像度でスキャンすることが可能である。しかし、

測定中に移動することができないという欠点があり、限られた条件下でしか使用することができない。このことから fMRI は実際の環境では有用ではない。したがって、マーケティングにおいては、NIRS および EEG が注目を集めている。ニューロマーケティングをテーマとした研究例として、Negishi ら[1]は受賞歴のある TVCM と受賞歴のない TVCM を被験者に見せ、その時の脳活動の違いを EEG によって測定した。その結果、受賞歴のある TVCM を見たときは集中とストレス感情に相関があることを見出した。Yanagisawa ら[2]は国際感情画像システムから採取した「快画像」と「不快画像」を被験者に提示し、そのときの脳活動の違いを NIRS によって測定した。その結果、「快画像」と「不快画像」の間には前頭葉の中央付近で有意な差があることを見出した。このように脳活動の状態を測定することでアンケートやインタビューといった主観評価では得られない感情や咆哮性を見出すことに成功している。

しかし、これらの研究では受賞歴のある CM とない CM、快画像と不快画像といった、提示した対象物に明らかな差があるものを比べている。しかし、実際の市場を考えると顧客を不快にまでする製品が多く流通しているとは考えられない。CM に関しても賞を取ったことのないものがほとんどであろう。このような点からマーケティングという観

^{†1} 北陸先端科学技術大学院大学
Japan Advanced Institute of Science and Technology

^{†2} 北陸先端科学技術大学院大学
Japan Advanced Institute of Science and Technology

点からすると、上記のような結果が有用であるかは疑問である。そこで、本研究は映画の予告動画のサムネイル画像を対象とし、NIRS によって視聴者に与える印象を分類するとともに、予告動画内からサムネイル画像に適したスナップショットを抽出する。映画予告のサムネイル画像という一見、画像間に差のないもので実験を行うことで、マーケティングのためのより有用なデータの獲得を目指す。

2. NIRS

本研究ではサムネイル画像及びスナップショットの印象を計測する装置として NIRS（日立ハイテクノロジーズ WOT-220）を用いる。NIRS は、光ファイバーを通したイルミネーターから検出器までの近赤外光の変化を検出することによって、酸素ヘモグロビン及び、脱酸素ヘモグロビンの濃度を測定することができる[3]。一般的に脳活動が活発になると局所的な血流が増加し、酸素が供給されるため酸素ヘモグロビンの値が上昇し、脳活動が落ち着いているときは脱酸素ヘモグロビンの値が上昇するとされている。



図 1 日立ハイテクノロジーズ WOT-220

3. 実験

3.1 サムネイル画像

本研究ではサムネイル画像を対象とし、実験を行った。サムネイル画像とは、動画の視認性を高めるための縮小された画像のことである。近年 YouTube やニコニコ動画などの動画サイトが普及し、再生数を伸ばすため様々な工夫がされている。再生数に影響を与える一つの要因としてサムネイル画像が挙げられているが、どのようなサムネイル画像が視聴者に好印象を与えているかは明らかにされていない。このような背景から、本研究ではサムネイル画像に焦点を当てた。

実験で用いるサムネイル画像は動画サイト YouTube[4]の東宝 MOVIE チャンネルから同じ映画の予告動画で同日にアップロードされており、かつ再生数に差のある動画のサムネイル画像（2枚1セット）を10セット（計20枚）取得した。なお、今回実験に用いたサムネイル画像は広告用に作成されたオリジナルの画像でなく、動画内のワンシーンをキャプチャした画像とした。

3.2 実験条件

同じ内容で再生数の異なった動画のサムネイル画像を見た際の脳活動の違いを測定するために実験を行った。

実験は次の（1）～（6）の手順で行われる。

- （1） NIRS 装置を装着してもらい、十分に落ち着いた精神状態を作ってもらう
- （2） 平常状態の脳活動状態を計測するためにディスプレイの中央に黒い点を映し、これを5秒間見ってもらう
- （3） 画像Aを15秒間提示する
- （4） 再び5秒間平常状態を計測する
- （5） 画像Bを15秒間提示する
- （6） 主観評価としてAとBのどちらのサムネイル画像に興味を引かれたか選択してもらう

上記の手順を1セットとし、これを10セット繰り返す。実験には20代の男女12名（男性：9名、女性：3名）に協力してもらった。

実験手順とディスプレイに表示されるイメージを図2に示す。被験者には提示するサムネイル画像の動画の再生数を知らせず比較および選択してもらった。また、サムネイル画像の提示順番はランダムとした。

実験で用いる脳機能測定装置日立ハイテクノロジーズ WOT-220 は全22chで前頭皮質の活動を測定することができる。光ファイバーの配置を図3に示す。

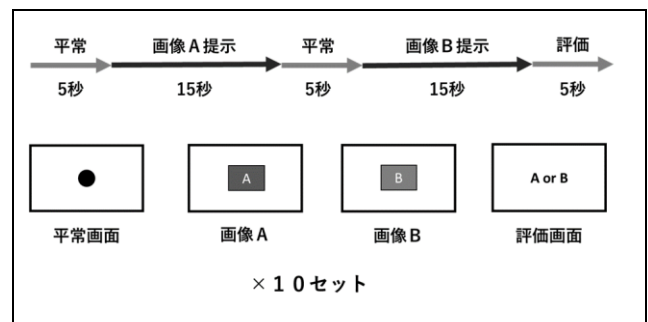


図 2 実験手順

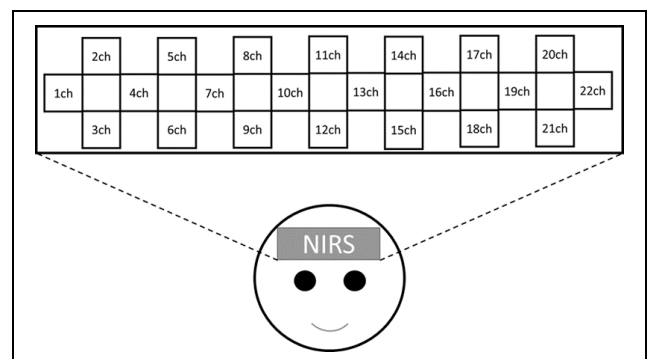


図 3 チャンネル配置

4. NIRS による印象分類結果

4.1 サムネイル画像と動画再生数の関係

サムネイル画像が動画再生数に影響を与えているのかを分析した。実験の際、被験者には2枚のうち好印象を与えられた1枚を選択してもらった。再生数の多い動画のサムネイル画像がもう一方のサムネイル画像より選択されやすいのかを明らかにするため、2つのサムネイル画像は同じ確率で選択されるという帰無仮説のもと二項検定を行った。

全120セットのうち再生数の多い動画のサムネイル画像が選択された数は74であった。検定の結果有意確率は0.013であり有意水準5%で帰無仮説が棄却された。

4.2 NIRS 値

NIRS の値から前頭前野における脳活動の傾向を調べるために、実験で得た数値をZスコアで平均化し、好印象を抱いた画像を見た時とそうでない画像を見た時の脳活動変化を比較した[6]。

図4は、ch21における好印象と判断した画像およびそうでないと判断した画像におけるNIRS信号の結果である。横軸は時間、縦軸はNIRS値を示している。山型のグラフは好印象のサムネイル画像を見た際の脳活動の様子で、谷型のグラフはもう一方のサムネイル画像を見た際の脳活動の様子である。好印象ではNIRS値が上昇し、もう一方ではNIRS値が減少していることが分かる。

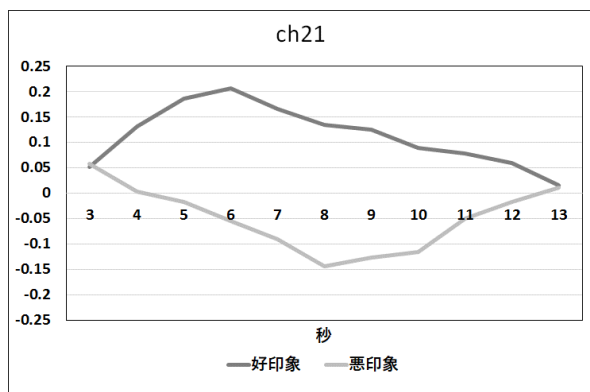


図4 好印象・悪印象のNIRS信号差

4.3 活性化部位ごとの脳活動差

NIRS のグラフから、好印象な画像とそうでない画像を見た時の脳活動の相違を確認することができた。したがって、次に前頭前野の活性化部位の間の脳活動の差異を統計的に試験した。

NIRS は特性上、脳活動が行われている大雑把な位置しか把握できない。よって、22チャンネルを5つの部位に分割し、それぞれの配列に対して平常時の波形から画像提示時の変化率を計算した。そして、各部位に対して以下の仮定

に基づき、ウィルコクソンの符号順位検定を行った。

- 帰無仮説: 好印象を抱いた画像を見た時のNIRSの値ともう一方の画像を見た時のNIRSの値には差がない
- 対立仮説: 好印象を抱いた画像を見た時のNIRSの値ともう一方の画像を見た時のNIRSの値には差がある

表1に結果を示す。グラフから19-22chの部位で優位な差 ($p < 0.05$) が生じていることが分かる。ここからさらにチャンネル域を限定し検定を行った。その結果20-22chで最も優位な差 ($p < 0.01$) が生じた。

これらの知見は、前頭前野皮質の左側に焦点を当てれば、印象を区別する可能性がより強いことを示している。

表1 各チャンネル域の有意値

1-4ch	5-9ch	10-13ch	14-18ch	19-22ch
0.251	0.937	0.315	0.572	0.013
				20-22ch
				0.003

4.4 ニューラルネットワーク

我々は、NIRSで測定した脳活動の値を用いて、好印象な画像とそうでない画像を見た時の脳活動の違いを見出すことができた。次に、測定されたNIRS信号をニューラルネットワークにより学習することで、好印象/悪印象を区別することが可能かどうかを調べた。ニューラルネットワークは、生物の脳の中の生物学的ニューラルネットワークをシミュレートするモデルである。これは入力レイヤー、隠れレイヤー、および出力レイヤーの3つのレイヤーで構成され、出力は入力レイヤーから出力レイヤーの方向に計算される。

トレーニング方法にはバックプロパゲーションを使用した。バックプロパゲーションは、トレーニング信号とニューラルネットワーク出力値との間の差を徐々に減少させるために、ニューロン間のシナプス荷重を調整するためのニューラルネットワーク学習方法の1つである。

4.5 ニューラルネットワークによる印象分類器

12人の参加者のNIRS信号を印象分類器のためのデータとして使用した。

前半5セット(120データ)の実験結果は学習データとして使用され、後半5セット(120データ)の実験結果はテストデータとして使用された。この実験で用いた階層的ニューラルネットワークの構成を図7に示す。

4.3節で述べたように20-22chのNIRS信号に有意差を見

出した。よって、入力データとして用いる NIRS 信号は 20-22ch を採用した。

入力には 20-22ch の各チャンネルの変化率を求め、その平均を用いた。15 秒の NIRS 信号のうち遅延と慣れを考慮し最初と最後の 2 秒 (計 4 秒) を除外した 11 秒を入力の対象とした。

出力レイヤーは、好印象レベル (最大 1 と最小 0) と悪印象レベル (最大値 1 と最小値 0) の 2 つのノードから構成されている。隠れ層は、同じ NIRS 信号を学習した後に得られた識別結果と比較して最も高い精度を有することができる 20 ノードで構成された。ここでは、学習数が 50 万回で、学習率が 0.01 とした。

第 1~5 セットの実験データが教師信号として使用され、同じデータが再び 2 つの感情を区別するために使用され、階層型ニューラルネットワークが正しく構成されたことが確認された。実験データの第 6~10 セットを用いて識別を行い、識別力を評価した。

4.6 学習結果

前半 5 セット (120 データ) で学習したニューラルネットワークを用い、後半 5 セット (120 データ) を好印象かそうでないか判定し、主観評価と比較した。

入力するデータは学習時と同じく 20-22ch の NIRS 信号を平均した 11 秒間 (遅延と慣れを考慮し最初と最後の 2 秒を除外) の NIRS 変化率とした。

各被験者の正解率と平均正解率を図 6 に示す。平均正解率は 75.8% で最も高い正解率は 90% であり、低い正解率は 60% であった。全 120 データのうち、誤った判定がされたデータは 29 であった。そのうち好印象を悪印象と判定した数は 20、その逆は 9 であった。

各試行について、好印象と評価したにも関わらず NIRS

信号の平均変化率が低くなっていた場合、悪印象と判定された。逆に悪印象と評価したにもかかわらず NIRS 信号の平均変化率が大きかった場合、好印象と判定された。本実験では被験者に 2 枚のサムネイル画像のうちどちらか一方を必ず好印象と評価させており、どちらも好印象、どちらも悪印象という選択をさせていない。よって、一方のサムネイル画像の判定が誤っているセットでは印象がかぶっている可能性が高い。

セット別に見ると第 7 セットのサムネイル画像の正解率が約 92% と最も高く、かつこのセットで用いられた画像の動画再生数の開きは 130 万回以上と今回用意したものの中で最大であったため、このセットの 2 枚の画像の間にはクオリティに大きな差があることが分かる。また、最も正解率の低かった第 10 セットは恋愛映画のサムネイル画像であったため被験者に男性の多い今回の実験ではこのジャンルに興味の薄い人が多かったことが正解率の低下を招いた可能性がある。

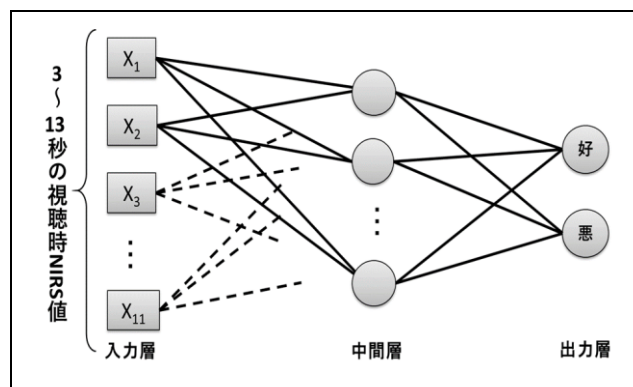


図 5 ニューラルネットワーク構成図

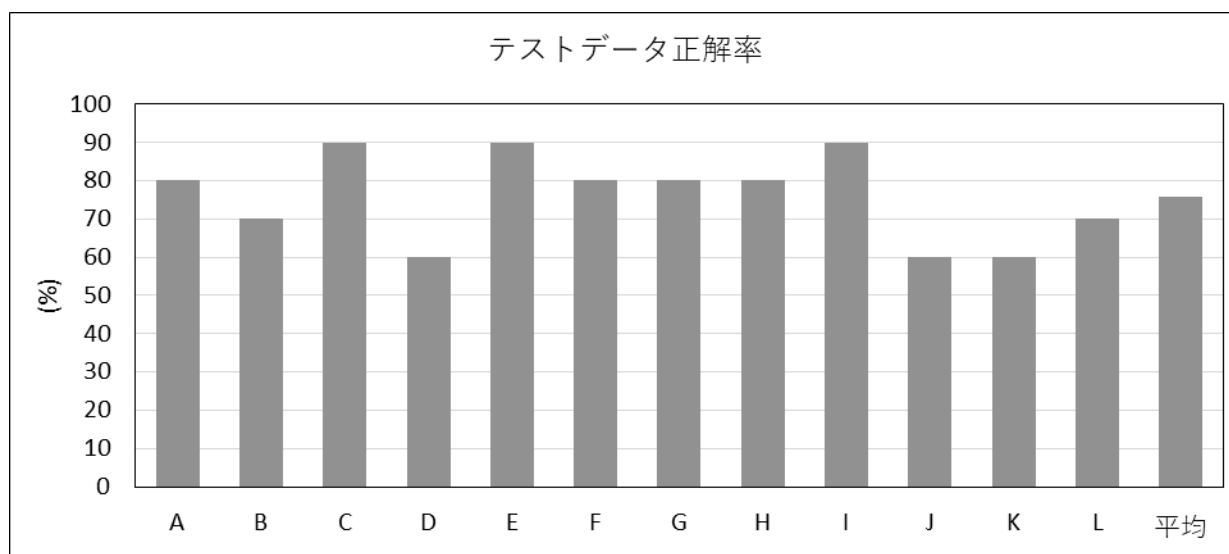


図 6 各被験者の正解率と平均

5. スナップショットの抽出

5.1 実験概要

ニューラルネットワークによりおよそ 75.8%の確率で好印象かそうでないかの分類に成功した。しかし、動画中のどのシーンを抽出すれば好印象を抱かれるサムネイル画像となりえるかは明らかにされていない。

我々は動画中から好印象を抱かれるシーンがどこなのかを特定するために、NIRS と学習済みのニューラルネットワークを用いて実験した。

被験者には映画予告の動画を視聴してもらい、その際の脳活動を NIRS によって測定する。測定された信号からいくつかのサムネイル画像候補を抽出し、その画像に対する印象を再度測定する。そして、4章で作成した分類器にかけることで、どのような脳活動をしているときのシーンがサムネイル画像として適しているかを検証した。

5.2 実験設定

動画中から好印象を抱かれるシーンがどこなのかを特定するために、以下2つの実験を行った。

実験 1

- (1) NIRS 装置を装着してもらい、十分に落ち着いた精神状態を作ってもらう
- (2) 平常状態の脳活動状態を計測するためにディスプレイの中央に黒い点を映し、これを5秒間見ってもらう
- (3) 予告動画を視聴してもらう
- (4) NIRS 信号の変化率からサムネイル画像の候補を抽出する

実験 2

- (1) 抽出した画像を被験者に見てもらい、その時の脳活動を測定する
- (2) 測定した NIRS 信号を学習済みのニューラルネットワークを用い、好印象かそうでないか分類する

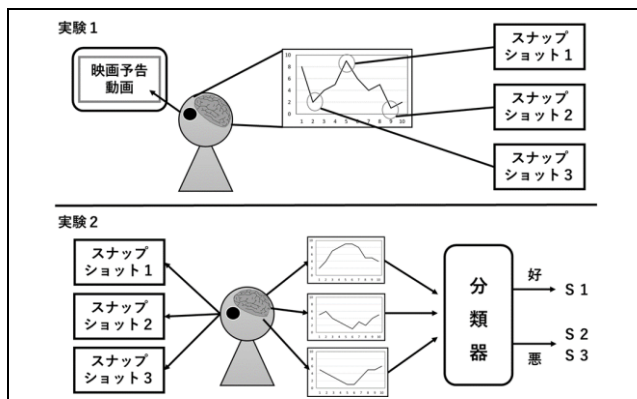


図 7 スナップショット抽出実験手順

実験 1 で用いた動画は動画サイト YouTube のからアクション、SF、ホラーの3ジャンルでそれぞれ3動画（計9動画）を被験者に視聴してもらった。

5.3 抽出するサムネイル画像候補

実験 1 の (4) にて抽出するサムネイル画像候補は以下の条件に従い抽出される。

- (1) NIRS の変化率が最大となるシーンの5秒前
- (2) NIRS の変化率が最大となるシーンの2秒前
- (3) NIRS の変化率が最小となるシーンの5秒前
- (4) NIRS の変化率が最小となるシーンの2秒前

多くの被験者の NIRS 信号は動画視聴開始時点で大きく上昇している。これは動画が始まったことに対する初期反応だと思われる。また、動画の最後には上映予告日などのテキストだけによるシーンがあり、このシーンはサムネイル画像として適切でないと判断した。よって、上記4つの条件に加え最初の10秒間と最後のテキスト予告の時間を除外した。抽出シーンの例を図8に示す。

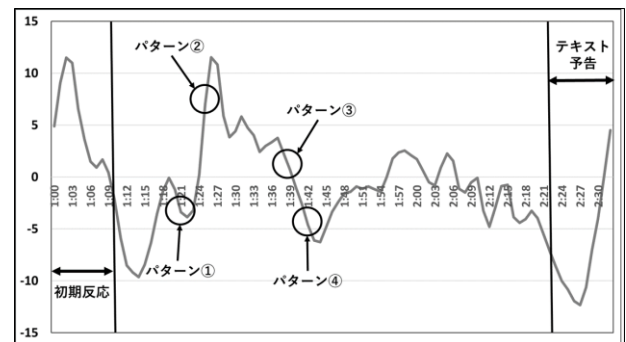


図 8 抽出シーン

5.4 分類器による判定結果

実験 1 で抽出したスナップショットを用い、実験 2 を行った。アウトプットされた NIRS 信号は学習済みのニューラルネットワークにより印象を分類された。各ジャンルの各パターンの結果を図9～図11に示す。

アクションジャンル、SFジャンルではパターン①および②がパターン③、④に比べ好印象と多く判定された。

ホラージャンルは4パターンの判定に大きな差はみられなかった。

3つのジャンルの各パターンの好印象判定数の差が統計的に有意な差であるかコクランの Q 検定を行った。その結果を以下に示す。

- アクションジャンルに対する検定の有意確率は 9.1E-05 であり、優位水準 1% でそれぞれのパターンに差があることを示した。

- SF に対する検定の有意確率は 0.012 であり、優位水準 5% でそれぞれのパターンに差があることを示した。
- ホラーに対する検定の有意確率は 0.707 であり、それぞれのパターンに有意差はなかった。

これらの結果は、アクションと SF に我々が作成した分類器が有効であることを示唆する。一方、ホラーに関してはこのジャンルの好き嫌いが大きく影響し、全てのパターンを好印象とする被験者と全てのパターンを悪印象と判断する被験者がみられ、今回のような結果になったと思われる。

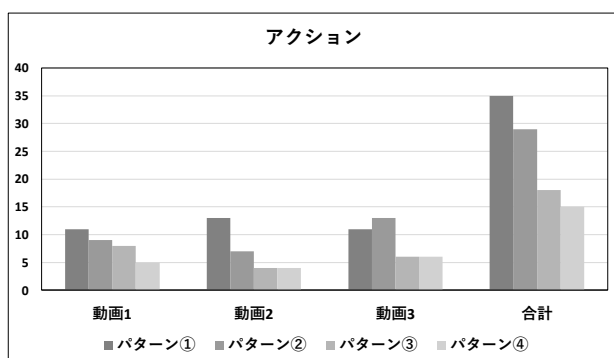


図 9 アクションの印象判定結果

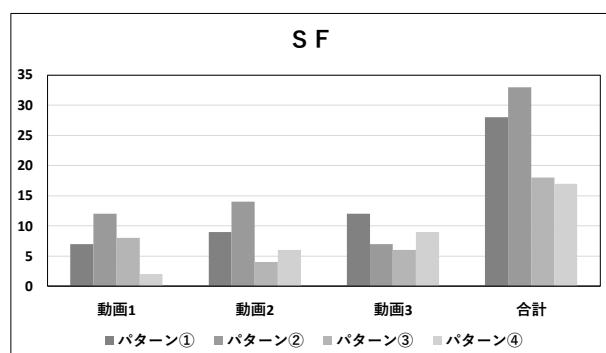


図 10 SF の印象判定結果

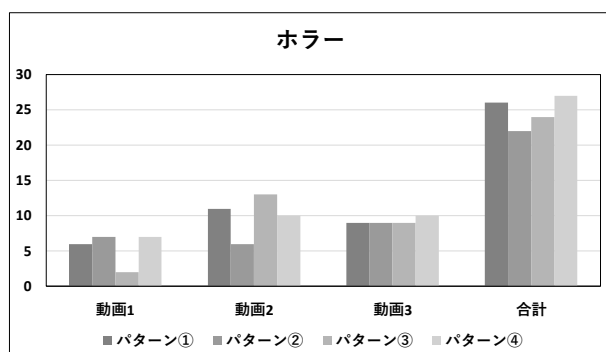


図 11 ホラーの印象判定結果

6. おわりに

本研究では、NIRS を用いてサムネイル画像が視聴者へ与える印象を分類するとともに、分類した結果をもとに動画中から好印象を与えると思われるスナップショットの抽出を試みた。

我々は、2 枚のサムネイル画像を提示し、その間の脳活動を計測した。検出された NIRS 信号を比較した結果、前頭皮質の左側に印象による差を見出した。これらの結果は、前頭前野皮質の左側に焦点を置くと、印象を区別できる可能性がより強いことを示す。

この結果をもとに好印象と悪印象を区別するため、ニューラルネットワークにより印象分類器を作成した。学習済みの分類器でテストを行った結果、平均成果率は 75.8% で最も高い正解率は 90% であり、低い正解率は 60% であった。

さらに学習済みのニューラルネットワークを用い、好印象なサムネイル画像が動画中のどのシーンなのか NIRS の値を元に検証した。アクション、SF、ホラーのジャンルで実験を行った結果、アクションでは動画視聴中の NIRS 信号が最大となる 5 秒前のスナップショットが最も好印象だと判断され、SF では NIRS 信号が最大となる 2 秒前のスナップショットが最も好印象だと判断された。この結果は統計的に見ても他のスナップショットより優位な差を示した。ホラーは動画視聴中の NIRS 信号が最小となる 2 秒前が最大の値となったが、有意差は見られなかった。

今後の展望としては被験者を増やすことでより高い精度の分類器を作成する、サムネイル画像以外の分野に適用するなどが挙げられる。

参考文献

- [1] Mana Negishi, Keio University, "Detection of KANSEI Value Trends Using the Electroencephalogram During Watching Memorable TV Commercials", IEEE International Symposium on Signal Processing and Information Technology, Dec. 2016.
- [2] Yanagisawa, Tsunashima, "Evaluation of pleasant and unpleasant emotions evoked by visual stimuli using NIRS", 15th International Conference on Control, Automation and Systems Oct. 2015.
- [3] Jobsis, F.F., "Non-invasive infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters," Science, Vol. 198, 1977, pp. 1264-1267.
- [4] YouTube, <https://www.youtube.com/>
- [5] Lang, P. J., Bradley, M. M., Cuthbert, B. N., "International affective picture system (IAPS) Instruction manual and affective ratings," The Center for Research in Psychophysiology, University of Florida, 1999.