

時系列データを用いた動画視聴時の 集中度推定に関する研究

小川 泰毅^{1,a)} 福元 伸也^{1,b)} 鹿嶋 雅之^{1,c)} 渡邊 睦^{1,d)} 西 正満^{†1,e)}

概要: 感染症の拡大により、遠隔やオンデマンドでの授業・会議が増加し、視聴者全員の集中度を把握することが困難になっている。従来の対面形式では、顔表情や振る舞いから集中度を推定できたが、遠隔環境では視聴者の様子を把握しにくい。本研究では、動画視聴者の顔特徴を抽出し、集中度を推定する手法を開発する。具体的には、顔特徴、傾き動作、視線・視点データを収集し、学習モデルを構築する。また、鼻尖部の温度と集中度の関連性を考慮し、サーモグラフィーを活用する。さらに、テストの告知による集中度の変化を検証し、鼻尖温度や視線検出の有効性を評価する。時系列情報を活用し、深層学習を用いたモデルを構築することで、高精度な集中度推定を目指す。

A study on the estimation of concentration during video viewing using time series data

1. はじめに

近年、新型コロナウイルスの影響で遠隔授業やビデオ会議が急速に普及している。対面では話者が聴講者の表情や行動を直接観察し、非言語コミュニケーションを通じて集中度を推定できるが、遠隔環境では難しい。

Mehrabian によると、コミュニケーションの情報割合は、言語情報 7%、聴覚情報 38%、視覚情報 55% とされ [1]、特に視覚情報が重要である。しかし、ビデオ会議では聴講者の様子を把握しにくく、新たな視覚情報を活用した手法が求められている。

本研究では、遠隔環境での聴講者の集中度を顔表情から自動推定する手法を開発する。顔表情の変化、視線、頭部動作、鼻尖温度を解析し、集中度を数値化するアルゴリズムを構築することで、遠隔授業やビデオ会議における非言語的

フィードバックを強化し、より効果的なコミュニケーション環境の実現を目指す。

2. 関連研究

2.1 生理・行動センサを用いた集中度推定

ここでは、身体にセンサを装着して生理指標（心拍、皮膚電気活動、呼吸数など）や動作を計測する研究について述べる。

2.1.1 種々のセンサを併用した集中度センシング法の検討

スマートフォン内蔵センサやウェアラブルセンサから得られる加速度、心拍数、皮膚導電率などの各種センシングデータを入力として、ユーザの集中度を定量的に評価するシステムが提案されている [2]。呼吸数はマイクによる呼気音解析、瞬目数は動画解析によってそれぞれ計測した。呼吸数と瞬目数は 1 分間の回数、心拍数と皮膚導電率は 1 分間の平均値で示している。

実験の結果、精神的発汗が発生したタイミングと心拍数が上昇し始めたタイミングが一致しており、これらの指標から精神的負荷の上昇が確認された。また、集中し始めると体の動きが減少し、集中が切れると椅子にもたれて楽な姿勢になることが観察された。このように、体を動かすタイミングと集中度の変化が同期している可能性が示唆された。

したがって、加速度センサを用いることでユーザの集中

¹ 鹿児島大学大学院理工学研究科工学専攻情報生体工学プログラム
Presently with Department of Engineering, Graduate School of Science
and Technology, Kagoshima University

^{†1} 現在、鹿児島大学大学院理工学系研究科技術部
Presently with Presently with Department of Technology, Graduate
School of Science and Engineering, Kagoshima University

a) k9265042@kadai.jp

b) fukumoto@ibe.kagoshima-u.ac.jp

c) kashima@ibe.kagoshima-u.ac.jp

d) mutuyas256@gmail.com

e) nishi@eng.kagoshima-u.ac.jp

度を推定できる可能性が十分にあると考えられた。

2.1.2 同期型オンライン授業における集中度推定システムの検討

各受講者の顔画像から学習状態に関連すると考えられる「瞬き」、「心拍数」、「頭の動き」、「腕の動き」、「画面との距離」、「顔表情」の6つの特徴量を抽出し、学習状態の判定を行った。これらの特徴量と集中度の相関を調べるために、Focus Calm と呼ばれる脳波解析デバイスを用いて実験を行い、各特徴量と集中度の関連性を検証した [3]。

その結果、「腕の動き」と「頭の動き」に集中度との相関がある可能性が示唆された。具体的には、腕の動きがある場合には集中力に関連する脳波が低下し、顔の動きがない場合には集中力に関連する脳波が上昇する傾向が確認された。また、疲労度と集中度の関連性を検証するために PERCLOS 値を用いた集中度の推定を行ったが、疲労度と集中度の間には明確な相関は見られなかった。

2.2 顔や身体の動きを用いた集中度推定

ここでは、カメラや映像解析によって集中度を推定する研究について述べる。

2.2.1 顔の向き推定を用いた集中度の推定システム

フレーム間差分を用いた顔領域の追跡、色情報を用いた顔領域の追跡、および顔部品検出に基づき、顔の位置と向きを推定する手法を提案した [4]。前処理として、動画の最初のフレームで Viola-Jones 法を使用して正面顔を検出し、顔領域の座標、幅、高さを取得している。また、検出された顔領域の色情報から肌色基準値を求めている。顔が動いていないと判断されたフレームにおいては、顔領域の境界線上の画素値と肌色基準値を比較する。このとき、画素値と肌色基準値との差が閾値以上である画素が境界線上の 1/10 以上を占める場合、その境界線を逆方向に移動させる。移動は、差が閾値以上である画素が 1/10 以下になるか、移動距離が顔領域の幅または高さの 1/2 以上に達するまで継続する。時系列上で前の差分領域を始点とし、動領域を通過して次の差分領域を終点とする垂直および水平方向の線を引き、線の本数から顔領域の移動方向を判断する。線の本数が少ない場合、動きは発生していないと判断した。被験者に首を左右に振らせた動画を用いて顔追跡および向き推定を行った結果、顔の向き推定が集中度の推定に有効であることが確認された。

2.2.2 瞬目と表情による緊張推定

瞬目・表情・挙動の有効性を確認し、これらの3つの外部情報を統合し、緊張推定を試みた [5]。彼らは特に内部状態が表れる指標として自己親密行動に着目した。自己親密行動とは不安や緊張下におかれた際に、鼻を触る、頭を掻くなどの自身の体に触ることで緊張を抑制する無意識な行動である。この自己親密行動を自動検出するために、肌色検出とラベリングによる身体と手の接触判定を構築した。

2.3 時系列データのための深層学習モデル

2.3.1 TFT

時系列データの予測において、リカ連とニューラルネットワーク (RNN) や Long Short-Term Memory (LSTM) ネットワークは広く利用されてきたが、長期的な依存関係の学習や特徴量の解釈性に課題があった。

これらの問題を解決するために、Temporal Fusion Transformer (TFT) を提案された。

TFT は、長期および短期の時系列情報を統合するために、LSTM と Transformer の利点を組み合わせたモデルである。本モデルは、ゲーティングメカニズムを導入することで、不要な特徴の影響を抑えつつ、重要な特徴のみを動的に選択できる。さらに、静的特徴と動的特徴を統合して学習可能であり、これにより、従来の時系列予測手法と比較して高い柔軟性を持つ。

また、TFT は予測の解釈性を重視した設計となっており、特徴量の寄与度や注意機構 (Attention Mechanism) を活用することで、「どの特徴が予測に大きく影響を与えたか」を可視化できる点が特徴である。需要予測、金融データ分析、医療データ解析などの多様な時系列タスクにおいて、LSTM や N-BEATS などの従来手法よりも高精度な予測を達成している。

このように、TFT は長期的な時系列データのパターンを効果的に捉え、予測の透明性を向上させる手法として有望である。そのため、近年では LSTM に代わる手法として、様々な応用分野で活用が進んでいる。

2.4 従来研究の課題

本節では、集中度推定や感情推定に関する先行研究を紹介した。生理・行動センサを用いる接触型手法に関する研究では、センサやシートなどのデバイスを使用して不観測情報を取得し、推定が行われていた。しかし、これらの手法では、講義や会議といった実際のシーンを再現できず、被験者が機械の使用によりストレスを感じるという問題があった。その結果、集中度だけでなくストレスの影響が推定に混入する可能性が指摘されている。

映像解析を用いる非接触型手法に関する研究では、特定のタスクに対して独自のモデルを設計し、検証が行われてきた。しかし、これらの研究には、推定モデルが対象タスクを十分に表現できていないという課題が存在した。これは、対象が人間であり、可観測な情報から不観測な内部状態を推定するという高難度の問題設定に起因している。この課題を克服するためには、可観測情報から大量のデータを収集・解析し、一定のパターンを見出す強力な学習器が必要である。高い認識精度と特徴表現力を持つ深層学習を活用することで、集中度推定を実現することを目指している。また、内部状態は時系列的に変化するため、この動的な性質を捉えることが重要である。しかし、先行研究では画像認識に

とどまり、映像認識のような時系列情報を十分に活用できていなかった。これらの課題に対処するため、可観測情報に加え、サーモグラフィーを活用して不観測情報を学習に取り入れることで、より高い認識精度を実現する集中度推定手法を提案している。

3. 時系列データを用いた集中度推定手法

3.1 データ収集

客観的な集中度を推定するため、遠隔授業やビデオ会議を再現する必要がある。そのため、本研究では、様々な長さの動画と動画内容に関連するテストを用い、データ収集用のミニ講義システムを構築した。図1にこのシステムの動作画面を示す。



(a) 講義画面

太陽から各惑星までの天文距離が1.6である4番目の惑星は？

- a) 金星
- b) 木星
- c) ケレス
- d) 火星

ボーデの法則とはどのようなものか？

- a) 太陽からどのような惑星があるのか
- b) どの惑星が大きいのか、大きさの計算方法の法則
- c) 太陽から各惑星までの距離が決まっている
- d) 天文距離はどのぐらいなのかの計算方法

天王星が発見されたことにより人々にどのような影響を与えたか？

- a) ボーデの法則の信憑性をさらに高めた
- b) 他にも惑星があると確信した
- c) 望遠鏡の技術の向上に努めた
- d) さらに多くの惑星を人々は探すようになった

(b) 小テスト画面

図1: ミニ講義システムの画面

3.2 顔特徴量の算出

顔特徴量を算出するため、特徴点の抽出が必要である。本研究では、OpenFace を利用し顔特徴点を抽出した。OpenFace で定義されている全ての特徴点の中から、眉の傾きや目の高さ、口の幅と高さを計算し、表1に示す7つの特徴量を算出する。

表1: 特徴量の定義と算出式

特徴量	算出式
f_1	左眉の傾き $(y_{21} - y_{20}) / (x_{21} - x_{20})$
f_2	右眉の傾き $(y_{24} - y_{22}) / (x_{24} - x_{22})$
f_3	左目の高さ $y_{41} - y_{37}$
f_4	右目の高さ $y_{46} - y_{44}$
f_5	口の幅 $x_{64} - x_{60}$
f_6	口の高さ $x_{66} - x_{62}$
$pitch$	頭の垂直方向の回転角度 R_x

3.3 傾き動作特徴の抽出

人間は相手の話を集中して聞いているとき、意識せず傾くことが多く見られる。したがって、傾き動作は集中度と正の相関があると考えられる。本研究では、頭の垂直方向の回転角度 (pitch) から、傾きの動作特徴を図2に示した手順により抽出した。FFT 変換後の周波数領域のデータに対し $f=0\text{Hz}$ のトレンド成分と $f>5\text{Hz}$ のノイズ成分にフィルタをかけて除去し、逆高速フーリエ変換 (IFFT) で時間領域に戻し、傾きの動作特徴を抽出する。

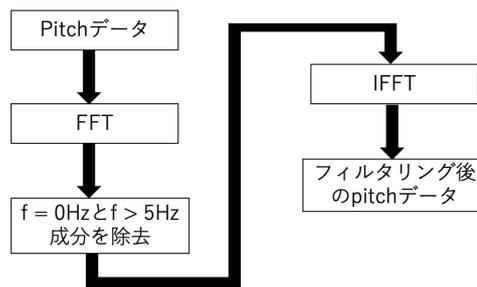


図2: 傾き動作特徴抽出

3.4 視線・視点情報の抽出

OpenFace は、視線情報を抽出するために、顔のランドマーク検出と 3D モデルベースの推定技術を組み合わせた手法を用いている。まず、顔検出器により画像または動画フレーム内で顔を特定し、その後、68 個の顔の特徴点 (ランドマーク) を高精度に推定する。このランドマーク情報をもとに、目の領域と瞳孔の位置を特定し、視線方向を計算する。視線推定では、3D 顔モデルを用いた頭部姿勢 (head pose) の推定と、瞳孔の位置を考慮した眼球の 3D モデリングを行うことで、目の向きが決定される。さらに、カメラキャリブレーション情報 (例: 焦点距離や光軸) を利用することで、視線ベクトルをカメラ座標系または頭部座標系で正確に算出する。これにより、視線の方向や焦点の位置を 3 次元空間で表現することが可能となる。視線情報の出力は、左右それぞれの視線角度 (ピッチ角とヨー角) や視線ベクトルとして提供される。これらの情報は、ヒューマンコンピュータインタラクション、注意力分析、および視線追跡の分野で広く応用

されている。

本研究では、図3のように視線の方向ベクトルと、目の位置ベクトルを取得し、視線の方向ベクトルを、「目とカメラの距離」と同じ長さまで引き伸ばした先を視点とし、視点とカメラの差が「ほぼ0」であればカメラ目線として検知する。

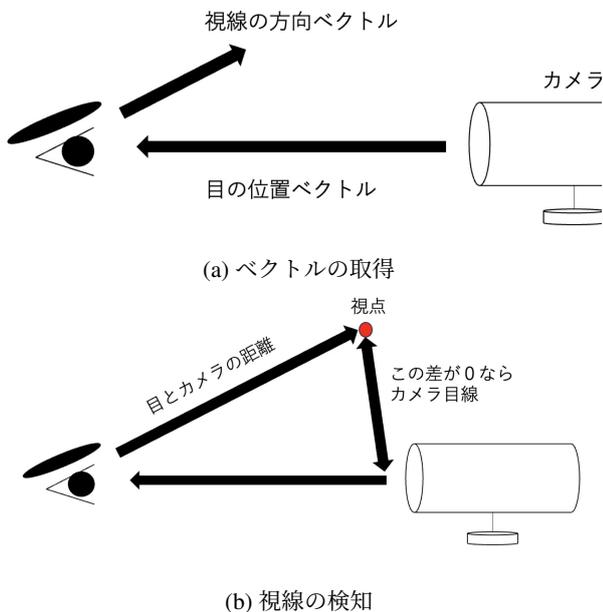


図3: 視線・視点の検出

3.5 鼻尖の温度取得

集中度推定の指標として鼻尖の温度変化を測定する。温度測定には赤外線サーモグラフィーカメラを使用する。測定は空調管理された室内（室温 $22 \pm 1^\circ\text{C}$ 、湿度 40%以下）で行い、被験者はカメラから約 1.5m の距離で正面を向いた姿勢を保持する。照明条件は蛍光灯のみとし、一定の光環境下で測定する。サーモグラフィーカメラで取得した温度画像から鼻尖部分を正確に抽出するため、画像処理ライブラリ OpenCV を用いて顔全体を検出し、鼻尖を ROI (Region of Interest) として設定した。鼻尖領域の温度は 1 秒間に 30 フレーム取得し、それらの温度値を平均化して解析に使用した。ノイズ除去のため、中央値フィルタを適用した後、環境温度との差分を計算することで温度データの補正を行った。

3.6 分類モデルの学習

分類モデルには、TFT を用い、頭の向き、顔特徴点、視線情報、鼻尖の温度変化などの多次元特徴量を入力データとして使用する。これらは集中度に関連すると考えられ、TFT の入力処理ブロックでエンコードされる。

入力シーケンスは Robust Scaler で標準化し、正解ラベルは one-hot ベクトルに変換して学習した。損失関数には交差

エントロピー誤差を用いる。

また、ニューラルネットワークの重みは学習率に応じて更新されていく。しかし、事前に適切な学習率を決めることは非常に困難である。学習率が大きすぎると学習が発散し不安定になり、逆に学習率が小さすぎると学習が進まない。そのため図4に示すように3エポックごとに精度改善を確認し、改善が見られない場合は学習率を 0.5 倍に下げっていくといった学習率の適応的な減少機構を導入する。

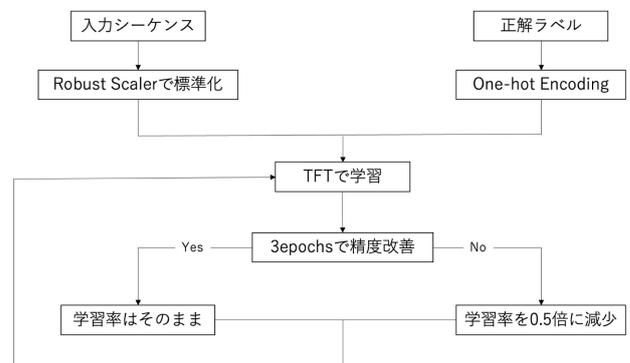


図4: 学習手順

図4の減少機構を導入し、構築した集中度推定システムの全体構成を図5に示す。このシステムは、Web カメラおよびサーモグラフィーを用いてデータを収集し、それらを基に集中度を推定するものである。

まず、Web カメラで被験者の動画を録画し、OpenFace を用いて顔特徴点や表情に関連する特徴量を抽出する。同時に、サーモグラフィーを使用して鼻尖の温度データを計測し、集中度に関連する温度変化を取得する。これら2種類のデータは統合され、学習済みモデルに入力される。

学習済みモデルには TFT を採用しており、これにより複数の時系列特徴量間の依存関係を学習して集中度を推定する。最終的に、推定された集中度の結果を出力することで、被験者の集中状態を定量的に評価する。

このシステムは、視線情報や温度変化といった非接触で取得可能な特徴量を効率的に統合し、高精度で集中度を推定できる点が特徴である。

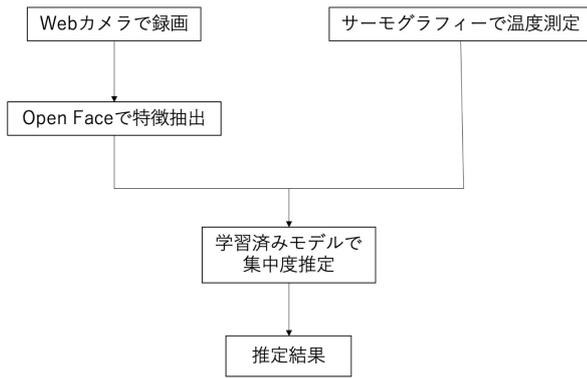


図 5: 集中度推定システムの構成

4. 評価実験

4.1 特徴量の評価

4.1.1 使用データセットの詳細

動画を複数用意し、友人 28 名を対象としてデータ収集実験を行った。被験者が受講している時の様子を Web カメラで録画した。

動画を見てもう際にその動画の内容に関する知識などがあると推定に影響を与えると考えることができるため、動画は聴講者が内容についてはあらかじめ知らない内容にした。

収集されたデータに対し、Augly を用い水増し (data augmentation) を行った。動画データの水増しは通常フレームごとにアフィン変換やノイズ付加などを適用しデータ量を増やすが、アフィン変換により 2 次元空間上データの特徴が大きく変わってしまうため、本研究ではノイズ付加のみを行った。

水増し後の動画データに対し、5 秒間 (150 フレーム) のウィンドウを 1 秒 (30 フレーム) ずつずらしてサンプリングした。これにより表 2 に示す 4274 個の入力シーケンスデータが得られた。

表 2: 作成したデータセット

テストの結果	データ数
低	1247
中	2114
高	913
合計	4274

4.1.2 テスト告知とテスト結果の相関

本研究では、集中度を測定するために、被験者に事前知識の無い動画を視聴してもらい、その後、その動画内容に関するテストを実施した。動画の内容は被験者にとって新しい情報を含むよう設計され、視聴中の集中度がその後の理解

や記憶にどのように影響を与えるかを評価することを目的とした。さらに、被験者の集中度における動機付けの役割を検証するため、実験条件として以下の 2 つを設定した。

1 つ目の条件では、動画視聴前に「視聴後に内容に関するテストを行う」と告知を行い、被験者にテストへの備えを意識させた。一方、2 つ目の条件では、テストの存在を一切告知せずに動画を視聴させた。その後、両条件の被験者に同じ内容のテストを実施し、得点を比較した。

テスト告知をした時の平均の正答率は 87.5 %、テスト告知をしなかった時の平均の正答率は 32 % になった。さらに、視覚的な結果の明確化を図るため、被験者の正答率に応じて結果を色分けした。具体的には、正答率が 80~100 % の場合を赤、50~80 % の場合を緑、それより低い場合を青で示した (図 6)。この色分けにより、各条件における被験者のパフォーマンス分布が直感的に理解しやすくなり、告知が与える影響の傾向をより明確に視覚化することができた。

ラファエロ、アテナイの学堂	ゴッダの法制と小惑星帯の発見	官渡の戦い	相対性理論	マネジメント	想像を超えた宇宙の広さ	砂漠の植物
100%	100%	60%	60%	80%	100%	60%
100%	100%	100%	90%	100%	100%	80%
80%	100%	60%	80%	80%	80%	60%
90%	80%	80%	100%	60%	80%	80%
100%	90%	100%	80%	100%	80%	80%

(a) テスト告知をした時

ラファエロ、アテナイの学堂	ゴッダの法制と小惑星帯の発見	官渡の戦い	相対性理論	マネジメント	想像を超えた宇宙の広さ	砂漠の植物
20%	20%	20%	20%	0%	10%	40%
0%	20%	0%	60%	10%	10%	60%
10%	0%	0%	20%	60%	10%	10%
20%	60%	10%	60%	80%	60%	20%
0%	10%	10%	10%	10%	0%	0%

(b) テスト告知をしなかった時

図 6: テスト告知の有無とテストの正答率の関係

図 6 から、動画視聴前にテストの告知を行った場合、告知を行わない場合と比較して、被験者のテストスコアが有意に高いことが確認された。この結果は、テストの告知が被験者の集中度を高め、動画視聴中の情報処理や記憶保持を促進する可能性を示唆している。特に、告知が無い場合は動画の内容に対する記憶が散漫になる傾向が見られ、集中度の低下がパフォーマンスに影響を与えているのではないかと考えることができる。

これらの結果から、集中度の測定や評価を行う際には、実験環境における動機付けの有無が重要な要因として影響を与えることが示された。本研究では、テストの告知というシンプルな条件の違いが被験者の集中度に与える影響を確認することができた。

4.1.3 視線・視点情報の評価

本研究では、視線の移動量と注視時間の割合に着目し、これらの特徴量を集中度の指標として出力し、学習に活用した。

視線の移動量は、視線ベクトルの変化量を時間経過ごとに計算することで算出し、被験者が視線を頻繁に動かしているか、あるいは特定の方向に集中しているかを定量的に評価した。また、注視時間の割合は、視線が一定の方向に向

いている時間を全視聴時間で割ることを求め、視線の安定性を評価する指標として利用した。

これらの特徴量を用いて、分析した結果、視線の移動量が少なく注視時間の割合が高い被験者ほど、正答率が高い傾向が確認された(図7)。具体的には視線の移動量が少ない被験者は視覚的注意を効率的に集中させており、動画視聴中の情報を効果的に処理していることが示唆された。一方、視線の移動量が多い被験者では、視覚的注意が散漫になりやすく、記憶や理解が低下する可能性が示された。

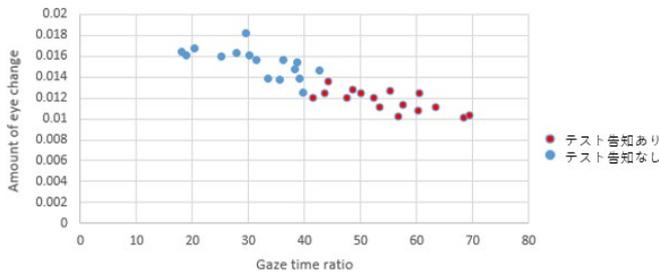


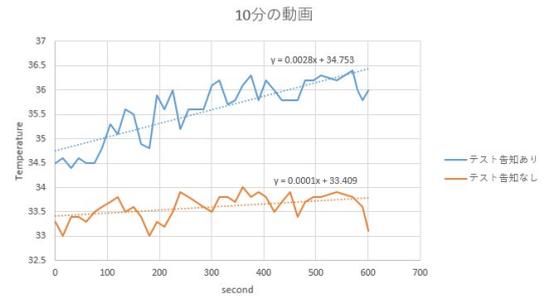
図7: 注視時間割合と平均視線変化量のグラフ

視線の移動量と注視時間の割合が動画視聴中の集中度を評価するための有力な指標であることを確認することができた。

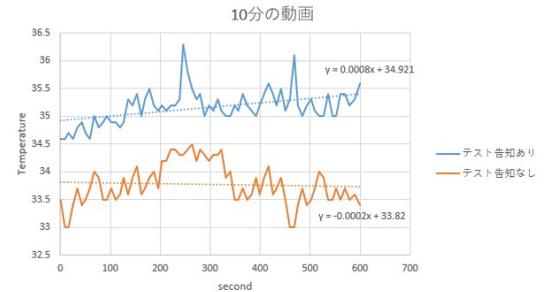
4.1.4 鼻尖の温度変化

本研究では、動画視聴中の集中度の評価指標の一つとして、鼻尖の温度変化に着目した。集中度が交感神経活動に影響を与えることで皮膚温度に変化が生じる可能性があることが先行研究で示唆されており、鼻尖の温度はその変化を敏感に反映する部位であると考えられる。そこで、動画視聴中の被験者の鼻尖温度を計測し、テスト正答率との関連を分析した。

温度データは動画視聴中に1秒ごとに記録され、視聴前後の温度差および視聴中の温度変化の傾向を解析した。実験の結果、テスト正答率が高い被験者ほど動画視聴中に鼻尖温度が上昇する傾向があることが確認された。一方で、テスト正答率が低い被験者では、鼻尖温度の変化がほとんど見られない、あるいはむしろ低下する傾向が見られた(図8)。



(a) Aさんの温度変化



(b) Bさんの温度変化

図8: 2人の温度変化の例

さらに、被験者全員の鼻尖温度変化の傾きを算出し、その結果を表3に示した。ここで、温度変化の傾きは、視聴中の温度データを線形回帰分析により近似し、その回帰直線の傾きとして計算した。この傾きは、被験者ごとの温度変化の方向性および強さを示すものであり、集中度の指標として活用できる。

表3では、各被験者について、テストの告知を行った条件と行わなかった条件での傾きが比較されている。この結果、テストの告知を行った条件では、ほとんどの被験者で正の傾きが確認された一方、告知を行わなかった条件では、正の傾きが少ない傾向が見られた。このことは、テストの告知が鼻尖温度の変化を促進し、集中度を高める効果がある可能性が考えられる。

表3: 被験者全員の温度の傾き

	テスト告知ありの動画	テスト告知なしの動画
平均値	0.00389	-0.00098
最小値	0.0006	-0.0036
最大値	0.0168	0.0001

4.2 集中度評価

本研究では、テストの告知、視線情報、鼻尖の温度変化の3つの指標を用いて集中度を評価した。

テストの告知により、被験者の動機付けが高まり、集中

度とテスト正答率が向上する傾向が確認された。視線情報の分析では、告知条件では視線移動が少なく注視時間が長い一方、非告知条件では視線移動が多く注意が分散していた。鼻尖の温度変化では、告知条件では温度が上昇する傾向があり、非告知条件では変化が少ないか低下するケースがあった。

これらの結果から、テストの告知は集中度を操作する手法として有効であり、視線情報や温度変化は集中度評価の指標として有用であることが示された。さらに、顔の特徴点や傾き情報を加えることで、より多面的な評価が可能になると考えられる。

4.3 システムの構築

本研究では、動画視聴中の被験者の集中度を評価するために、視線情報、鼻尖の温度変化、顔特徴点、アクションユニット、傾き動作を解析するシステムを構築した。収録した動画は、被験者が特定の条件下で動画を視聴している様子を撮影したものであり使用する特徴量を抽出するための基盤として使用された。

動画データに対する解析の流れとしては、まず被験者の顔全体を検出し、視線ベクトルや注視時間、視線移動量を算出した。次に、サーモグラフィーを使用して時間ごとの温度変化を算出した。これらのデータは、構築した学習済みモデルに入力され、被験者の集中度を推定するための特徴量として活用された。実際に学習に使用したデータとは異なる35個のテストデータをモデルにかけ推定を行った。出力には、「集中している」と考えられる確率が出力されている。

このシステムでは、入力された動画データから抽出された特徴量に基づき、被験者の集中度を推定している。次に、テストデータ全体の出力結果をまとめた。ここで表4の基準の値で色分けを行い、視覚的に示した。

表4: 基準の値

注視時間割合	平均視線変化量	温度変化の傾き
50%以上：赤	0.0125以下：赤	0.0001以下：青
50%以下：青	0.0125以上：青	0.0001~0.001：緑
		0.001以上：赤

表4の基準で色分けしたテストデータの全体の結果を図9に示す。

動画	推定結果	注視時間割合	平均視線変化量	温度変化の傾き	テストの正答率
k_m_1	0.297987	28.91%	0.0163	-0.005	0%
k_m_2	0.223472	18.88%	0.0164	-0.0036	10%
k_m_3	0.270123	25.14%	0.016	-0.0012	0%
s_m_1	0.269421	39.77%	0.0154	-0.00087	10%
s_m_2	0.089393	17.94%	0.0164	0.0001	0%
s_r_1	0.467096	41.29%	0.0121	0.0006	60%
shi_ma_1	0.643515	48.41%	0.0124	0.0003	60%
ya_ma_1	0.193281	34.30%	0.0146	-0.0024	0%
ya_ma_2	0.13235	38.77%	0.0154	-0.0028	10%
yo_so_1	0.581134	44.13%	0.0123	0.0009	60%
ko_so_1	0.057407	38.80%	0.013	0.0002	60%
yo_so_2	0.125215	20.81%	0.0155	-0.0023	10%
to_so_1	0.81213	26.41%	0.0153	0.0004	40%
ya_so_1	0.307969	49.34%	0.0131	-0.0019	20%
shi_so_1	0.319421	47.81%	0.0136	-0.0002	80%
ka_m_1	0.836163	46.05%	0.0129	0.0005	20%
ka_so_1	0.248452	38.72%	0.0151	-0.003	0%

(a) テスト告知なしの推定結果

動画	推定結果	注視時間割合	平均視線変化量	温度変化の傾き	テストの正答率
sa_r_1	0.865196	56.58%	0.0103	0.0152	100%
sa_r_2	0.967581	55.22%	0.0121	0.00389	90%
sa_r_3	0.998216	50.13%	0.0119	0.0147	100%
sa_wa_1	0.896621	68.20%	0.0102	0.0168	100%
shi_ra_1	0.915984	58.58%	0.0115	0.014	100%
shi_ra_2	0.926503	60.40%	0.0108	0.0164	80%
shi_ra_3	0.994448	65.22%	0.0101	0.0159	80%
ya_wa_1	0.889218	54.35%	0.012	0.0148	100%
yo_ma_1	0.567033	50.13%	0.0123	0.0002	60%
yo_ra_1	0.9351203	60.52%	0.0123	0.0154	88%
yo_ra_2	0.6573225	48.41%	0.0124	0.0003	65%
yo_sa_1	0.5354805	52.10%	0.0125	0.0008	60%
ko_wa_1	0.292044	69.42%	0.0104	0.0004	80%
ko_wa_2	0.851175	58.57%	0.0119	0.0133	80%
ko_wa_3	0.935239	56.49%	0.0116	0.0141	100%
to_r_1	0.916956	65.58%	0.0122	0.0044	100%
ko_r_1	0.584022	51.84%	0.0127	0.0002	60%
ha_r_1	0.905068	57.72%	0.0107	0.0129	80%

(b) テスト告知ありの推定結果

図9: テストデータの推定結果

図9は、集中度の推定結果と注視時間割合、平均視線変化量、温度変化の傾き、テスト正答率を示したものである。35個のテストデータのうち32個で推定結果と正答率が一致していることが確認された。一致率は約91.4%となった。これは本システムの高い精度を示しており、注視時間や視線変化量、温度変化も集中度推定に有効であることが確認された。

5. 考察

5.1 非接触における集中度推定について

集中度を推定するための従来のアプローチには、脳波測定装置や心拍計などの接触型デバイスを用いるものが多く存在する。しかし、これらの手法は被験者に直接デバイスを装着させる必要があるため、装置の装着が被験者の自然な行動を妨げる可能性があるほか、測定環境を限定するという制約がある。一方で、非接触手法は、こうした課題を克服するための有効なアプローチとして注目されている。

非接触手法の最大の利点は、被験者に物理的負担を与えず、自然な行動を観察しながらデータを取得できる点にある。この特性により、教育現場やオンライン学習環境、リモートワーク、さらには運転中や作業中といった多様な状況での応用が可能となる。また、カメラやサーモグラフィーといった一般的なデバイスを用いることで、比較的低コストでの導入も期待される。

非接触手法を用いた集中度推定には、視線情報、顔表情、温度変化といった外部から観測可能なデータが活用される。本研究では、これらのデータを基に集中度を推定するアプローチを検討したが、これにより、従来の接触型手法と比較して高い利便性と実用性が確認された。一方で、外部環境要因や個人差といった課題も存在する。例えば、照明条件や撮影距離の違いが視線情報の精度に影響を与えるほか、温度変化の測定精度もデバイスの性能に依存する部分がある。

また、非接触手法では、集中度を直接測定するのではなく、間接的な指標に基づいて推定を行うため、データの多次元的な統合や精度向上のための機械学習モデルの工夫が必要となる。今後は、これらの課題に対応する技術の開発や、様々な環境下での汎用性の検証が求められる。

非接触手法は、装置の物理的負担やコストの削減に加え、自然な行動を観察しながらデータを取得できる点で優れており、集中度推定の新しい方向性を示している。本研究の結果を踏まえ、更なる技術的改良や応用範囲の拡大が期待される。

5.2 構築したシステムについて

構築した集中度推定システムは、視線情報・温度変化・顔特徴点・アクションユニットなどを用い、Temporal Fusion Transformer(TFT)により集中度を推定する。TFTは時系列データの依存関係を捉えるのに優れ、高精度な推定を可能にした。

評価実験では、学習に使用したデータとは異なる35個のテストデータをモデルに入力した結果、32個のテストデータで正しい推定が行われた。91.4%という高い一致率は、TFTモデルが非接触の特徴量から得られる情報を適切に処理し、高精度で集中度を予測できていることを示している。

本システムにおいて、TFTモデルは集中度推定のために学習され、テストデータに対して高い精度で集中度を予測する結果が得られた。具体的には、視線情報や温度データ、表情データの時間的な変化をモデルに入力し、各被験者の集中度を確率値として出力する。この推定結果は、学習者や作業者の集中度をモニタリングし、リアルタイムでフィードバックを提供する応用に有用であると考えられる。

5.3 他のモデルとの精度比較

本研究では、集中度推定において Temporal Fusion Transformer (TFT) を用いた手法と、従来の LSTM (Long Short-Term Memory) を用いた手法を比較した。両手法の精度を評価するために、35個のテストデータを用いて実験を行った。その結果、TFTモデルは32/35のデータで正しい集中度の推定を行い、約91.4%の一致率を示した。一方、LSTMモデルは26/35のデータで正しい推定を行い、約74.3%の一致率となった(表5)。

TFTは複数の特徴量を同時に処理し、時間的な依存関係

表 5: LSTM モデルと TFT モデルの比較

モデル名	精度
TFT	91.4 %
LSTM	74.3 %

を捉える能力が高いため、集中度推定において優位であった。一方、LSTMは長期的な依存関係を捉える能力に限界があり、精度が低下した。以上の結果から、TFTは集中度推定においてより適したアプローチであることが確認された。

6. 結論

6.1 まとめ

本研究では、非接触手法を用いた集中度推定システムを構築し、その有効性を検証した。視線情報、鼻尖の温度変化、顔特徴点、アクションユニット、顔き動作といった複数の特徴量を統合し、Temporal Fusion Transformer(TFT)を用いて被験者の集中度を推定した結果、35個のテストデータ中32個で正しい推定を行い、高い精度(約91.4%)を達成した。また、LSTMモデルとの比較においても、TFTが優れた性能を示すことが確認され、非接触で得られる特徴量を活用する手法の有効性が示された。

6.2 今後の課題

集中度の表れかたは被験者によって異なり、視線の動きや表情の変化、体温変化の傾向に個人差が存在する。本研究では一般的な特徴量を用いて推定を行ったが、より高い精度を目指すには、個別のプロファイリングを行い、個人差に適応したモデルを構築する必要がある。具体的には、被験者ごとのデータを用いた転移学習や、パーソナライズされた特徴量選択手法の検討が必要であり、今後の課題として、非接触手法による集中度推定の実用化と社会的な価値の向上が期待される。精度向上や応用範囲の拡大など今後の研究の発展が望まれる。

7. 参考文献

参考文献

- [1] A. Mehrabian, Nonverbal Communication, Aldine Publishing Company, 1972.
- [2] 濱谷 尚志, 内山 彰, 東野 輝夫, 「種々のセンサを併用した集中度センシング法の検討」, 情報処理学会研究報告, 2015.
- [3] 徳田 正安, 浜田 宏一, 荒井 正之, 「同期型オンライン授業における集中度推定システムの検討」, 情報処理学会, 2023.
- [4] 田中 亮磨, 古賀 道雄, 鹿嶋 雅之, 佐藤 公則, 渡邊 睦, 「時系列動画解析に基づく人物内部状態推定に関する研究」, 電子情報通信学会技術研究報告.PRMU, パターン認識・メディア理解 114(90), 61-66, 2014-06-12
- [5] Ralf C. Staudemeyer, Eric Rothstein Morris, 「Understanding LSTM – a tutorial into Long Short-Term Memory Recurrent Neural Networks」, 2019, arXiv:1909.09586