

テキスト黙読時における視線情報を用いた集中状態・非集中状態の識別に関する調査

An Investigation on Eye Gaze-Based Recognition Between Concentration and Non Concentration During Silent Reading

田中 咲希¹⁾ 辻 愛里²⁾ 藤波 香織²⁾

Saki Tanaka Airi Tsuji Kaori Fujinami

1 はじめに

労働人口の減少に伴いオフィスワークの生産性向上が求められる [1] が、生産性低下の一要因として集中力の低下が挙げられる [2]. 作業中の集中力の低下をシステムが検知し、作業に介入して休憩を促す等で集中力を回復させることが可能となれば、生産性の低下を抑止できると考えられる. 本研究では、PC 作業において普遍的に存在すると想定されるテキスト黙読を対象とした、視線情報による集中状態・非集中状態の識別を行う.

2 関連研究

2.1 集中度推定

集中度推定を行う際には、集中度を推定可能な仕組みを用意し、評価実験によってその仕組みの有効性を検証する方法が考えられる. そのため、集中度推定では集中の定義と独自の推定手法の他に、評価実験における実験参加者の集中度に対する正確な評価が必要となる. 宮城ら [3] は、「認知資源を一定期間対象に割り当てている」状態を集中状態と定義し、Uema ら [4] は集中度を「パソコンで文書を作成するなどの単一作業にどれだけ深く取り組むかを示すもの」と定義している. 深澤ら [5] はテレワーク時の作業集中性と通知への気づきやすさとの関係性の調査に関する研究において、作業集中性をアンケートの回答 (心理評価) および作業達成率 (行動評価) によって評価した. しかし、アンケートによる集中度の評価は前述の通り主観的である点で正確性に欠ける. 集中度の客観評価を行った研究として、Yoshihashi ら [6] は講義聴講中の学生の集中状態を、スクリーンの方を向いて話を聞こうとする姿勢を取っている状態と定義して、学生の集中状態・非集中状態の識別を行ない、自己評価における曖昧さを回避している. しかし、スクリーンを向いている状態を集中状態とする定義は現実的でない. 例えば、他者同士の会話内容に注意が向けられ授業に集中できない場合や、疲労により授業への注意が散漫になっている場合のように、顔がスクリーンに向いていても非集中状態である場合が想定される. また、前

述の宮城ら [3] は計算課題の解答時間によって全作業時間に対する集中時間比率を算出したが、集中時間比率は解答という時間的に分割可能なマイクロタスクのない作業には適用不可能である. 以上の観点から、顔の向きに注目した集中度推定手法は厳密性に欠ける. また、課題解答時間に注目した集中度推定手法は PC 作業全般に対する評価には適さない. 一方で、視線情報は心理状態推定に有用であるという報告 [7] があることから集中度推定への活用にも期待でき、どの PC 作業においても取得可能である.

2.2 視線を用いた集中度推定

集中度を含めた人間の内在状態を扱う研究では、主にアンケートなどを用いた自己評価尺度によって内在状態を判断している [8, 9, 10, 11] が、社会的に望まれる評価を行う [12], 主観に左右されるため個人によって基準が異なる、誤った判断による報告が行われるといった懸念がある. 客観的な内在状態推定手法としての視線情報を使用した研究では、主に固視、サッカード、スムーズパスト、瞳孔径の 4 種類が対象として取り上げられる [13, 14, 15, 16, 17]. このような眼球運動を意図的に制御することは困難であることから、視線情報の採用によって正確な内在状態推定が期待できる. また、画面下部に設置するアイトラッキングデバイスを使用する場合には、ユーザに与える負荷が少ない. 視線情報を用いた集中度推定としては、Kanematsu ら [18] は瞬きの情報を取得可能なデバイスである JINS MEME [19] を使用した集中度推定を行った. しかし、この研究では集中度の相対的な変化のみに注目しており、集中状態・非集中状態の識別は行われていない.

2.3 読書中の視線による内在状態推定

本研究では一般的にオフィスワーク中に生じる作業としてテキスト黙読を対象とする. 読書中の視線から読者の内在状態を推定する研究として、Faber ら [15] は読書中の視線から計算された視線特徴量を用いて機械学習を行い、読者の理解度の低・中・高を識別するモデルを構築した. また、Southwell ら [20] は、読書中の視線から計算された視線特徴量を用いて機械学習を行い、読者のマインドワンダリング [21] 発生有無を識別するモデルを構築した. ここで、マインドワンダリングとは、現在のタスクから関係のない考えに注意が移る現象であ

1) 東京農工大学 大学院 生物システム応用科学府 生物機能システム科学専攻. Department of Bio-Functions and Systems Science, Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology.

2) 東京農工大学 大学院 工学研究院 先端情報科学部門. Division of Advanced Information Technology and Computer Science, Institute of Engineering, Tokyo University of Agriculture and Technology.

表 1: 本研究と従来の研究との比較

研究	集中状態の評価	眼球運動の利用	主観評価	客観評価	内在状態識別
本研究	○	○		○	○
Uema ら [4]		○			
深澤ら [5]	○		○		
Yoshihashi ら [6]	○			○	○
宮城ら [3]	○			○	
Kanematsu ら [18]	○	○		○	
Faber ら [15], Southwell ら [20]		○	○		○

る。これらの研究から、読書中の視線は内在状態推定に有効であり、読書中の視線情報による集中状態・非集中状態の識別には可能性がある。そこで、本研究ではこれらの研究を参考に視線特徴量を使用する。

本研究と従来の研究との比較を表 1 に示す。本研究では関連研究の知見から読書中の視線情報による集中状態・非集中状態の識別を目指す。

3 集中状態および非集中状態の定義

2.1 節で述べたように、集中という概念は曖昧であるため、本研究においても集中の明確な定義が必要である。ここで、集中状態に対する非集中状態は 1 つの決まった状態ではなく、複数の状態で構成されていると考えられる。具体的には、メインタスクに従事している作業には次のような 4 つの状態が考えられる。

1. メインタスクとは別のことを考えている
2. メインタスクではなく別の作業に集中している
3. メインタスクと同時に別の作業を行う (阻害状態)
4. 疲れていて集中できない (疲労状態)

1 の状態を実験で再現するためには実験参加者の思考を制御する必要があり、2 の状態はメインタスクとは別の作業が、メインタスクに関連のある作業であるかを判定する必要がある。例えば、メインタスクとして資料作成を行っているにも関わらず、別の作業として SNS 閲覧に集中している場合、作業者はメインタスクへ集中していない状態であると想定されるが、別の作業として資料作成の関連書籍の閲覧に集中している場合、作業者が非集中状態であるとは必ずしもいえない。したがって、1 の状態と 2 の状態を実験条件として設定するのは困難である。そこで、実験条件として設定が容易かつ検出可能性が見込まれる 3 の阻害状態および 4 の疲労状態を本研究における非集中状態とした。ここから、本研究での定義は、「一定期間 1 つの事物に一定量以上の認知資源を割り当てること」とする。図 1 は読書を例にこの定義を図示したものである。読書という 1 つの事物に一定量以上の認知資源を割り当てるとき、作業者は集中状態であるとみなされる。一方で、読書とラジオ視聴のように複数の事物に認知資源を割り当てるとき、作業者は非集中状態であると考えられる。さらに、読書という 1 つの事物であっても、疲労によって一定量以上の認知資源を

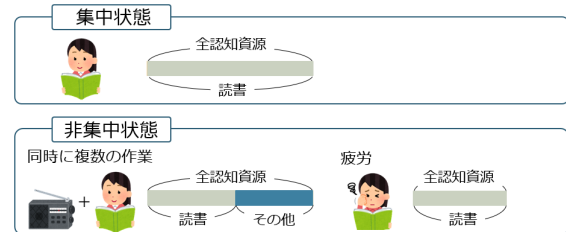


図 1: 集中の定義

割り当てられないとき、作業者は非集中状態であると想定される。このとき、阻害状態と疲労状態での視線特徴が同じであると断定することはできない。したがって、本研究では阻害状態と疲労状態を分離して扱い、集中状態、阻害状態および疲労状態の 3 クラス分類を試みる。

4 黙読作業時の視線運動計測実験

4.1 実験概要

本実験では集中条件、阻害条件および疲労条件において 30 分ずつ黙読タスクを実施する実験を行い、それぞれの条件下の視線運動を収集する。次に、得られた視線情報を用いて機械学習を行うことで、各視線情報に対応する参加者の状態を予測する。

4.2 実験環境

実験参加者が使用した作業台を図 2 に示す。参加者は参加者用作業台に配置したディスプレイに表示された文章を黙読した。作業者の視線データを収集するため、Tobii 社製アイトラッカーである Tobii Pro nano[22] を PC の画面下部に設置することで、使用者に非接触な形で視線運動を取得した。なお、サンプリングレートは 60Hz に設定した。実験中の音声を聴くタスクを行う際や、雑音による黙読タスクへの阻害を回避するため、ノイズキャンセリング機能の付いたヘッドフォンを使用した。

本実験におけるメインタスクでの黙読は、内容理解を目的とした黙読を想定しており、読者が文章の内容を認識しないまま目を通すだけの行為とは区別される作業である。そこで専門分野の高度な前提知識を必要とせずに黙読可能であると考えられる高校生新聞オンライン [23] の記事を採用した。

4.3 被験者概要

本実験の被験者として 20 代の学生 13 名 (メガネ有り : 5 名, メガネ無し : 8 名) が参加し、実験参加には謝

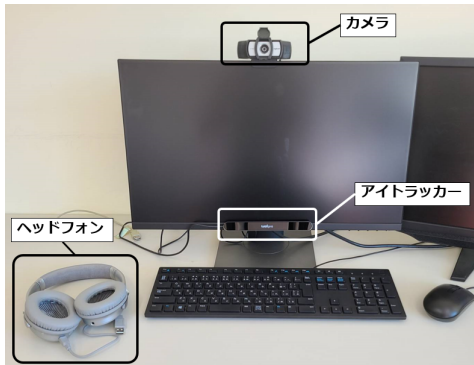


図 2: 参加者用作業台

パターンA

説明	10分
アンケート1	2分
練習1	1分
集中条件	15分
テスト	7分
休憩	15分
練習2	3分
阻害条件	15分
テスト	7分
アンケート1	2分
疲労条件	15分
テスト	7分
アンケート1	2分
アンケート2	4分

パターンB

説明	10分
アンケート1	2分
練習1	1分
阻害条件	15分
テスト	7分
休憩	15分
練習2	3分
集中条件	15分
テスト	7分
アンケート1	2分
疲労条件	15分
テスト	7分
アンケート1	2分
アンケート2	4分

図 3: 1 日の実験手順

金が支払われた。また、本実験は東京農工大学倫理審査委員会の承認を受け実施した（承認番号：211102-0360）。

4.4 実験手順

図 3 に 1 日の実験手順の例を示す。参加者への過度な負担を避けるため、1 条件における黙読時間を 15 分とし、本実験を 2 日に分けて行った。本実験では、参加者の空腹による集中への影響および満腹時の眠気による集中への影響を避けるため、本実験の実施時間を 13 時～17 時の間とした。集中条件、阻害条件および疲労条件において、メインタスクとしてテキスト黙読を行わせた。阻害条件ではメインタスクと同時に阻害タスクを行わせた。阻害タスクの内容は、認知負荷の分析に関する複数の研究 [17, 24, 25] において、認知負荷をかけるために用いられる n-back タスク [26] を採用した。疲労条件では、直前のタスクにより疲労を蓄積させたうえで、テキスト黙読のみ行わせた。順序効果を相殺するため、図 3 の 2 通りの順序を用意し、パターン A に 7 名、パターン B に 6 名を割り当てた。実験中の黙読の有無を確認するため、各作業後毎に内容理解を問うテストを実

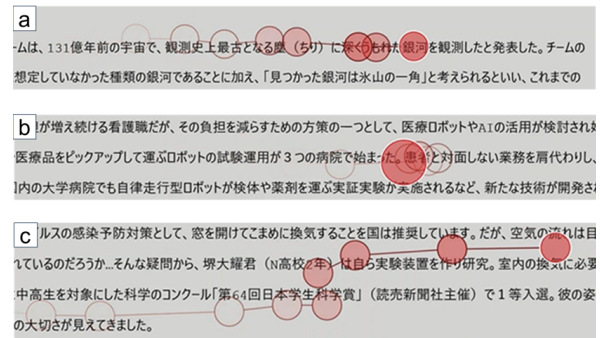


図 4: 収集データ例 (a: 集中状態, b: 阻害状態, c: 疲労状態)

施した。また、疲労度を確認するための質問紙調査（アンケート 1）と集中度変化の諸要因について確認するための質問紙調査（アンケート 2）を実施した。実験では、作業内容および作業環境へ慣れることを目的としたタスク練習を本実験の前に実施したが、疲労が本実験へ影響することを避けるため、別日に行った。タスク練習実施日に阻害タスクである n-back タスクの n の値も定めた。これは、1-back タスクとメインタスクを同時に実行する作業を 1 分間行い、黙読テキストを理解できずに目で追うだけになってしまった参加者のみ 0-back タスクに変更して再度 1 分間の並行作業を行うことで決定した。n の決定後、改めて 10 分間の並行作業練習を行った。

4.5 データ収集結果

図 4(a)～(c) はそれぞれ集中状態、阻害状態および疲労状態における視線データの一例を示す。図 4 の赤い印が視線遷移を表す。このうち、赤い丸が固視点を表し、線がサッカードを表す。また、赤い色の濃度はその時点からの時間的な距離を表しており、濃いほど新しい視線動作であることを示す。例えば、図 4(a) と (b) を比較すると、(b) では固視点がより狭い範囲に留まっており、(a) と (c) を比較すると、(c) では上の行への視線遷移、すなわち読み返す視線移動がみられる。

5 識別モデル

5.1 識別手法概要

図 5 に本研究の識別手法概要を示す。まず実験にて得られた視線情報に後述する前処理を施し、ウィンドウ単位（インスタンス）に区切る。次に、インスタンスごとに特徴量を計算する。その後、機械学習のために全データを訓練データとテストデータに分けて、訓練データを使用し外れ値を計算し、訓練データおよびテストデータにおいて外れ値に該当する数値を変更する処理を行う。そして、訓練データにのみ標準化処理およびデータ均衡化処理としてクラス間のインスタンス件数を均衡化するサンプリング処理を行う。最後に、訓練データを用いた機械学習により識別モデルを構築し、テストデータに対して集中状態、阻害状態および疲労状態の識別を行う。

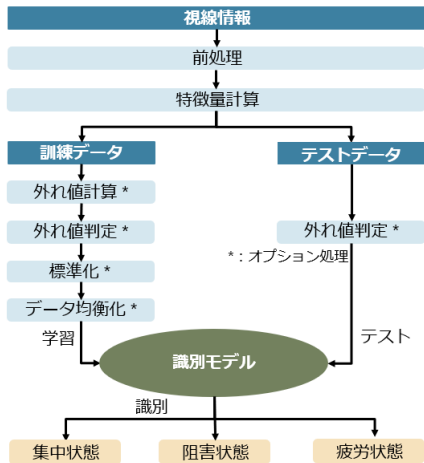


図 5: 識別手法概要

なお、外れ値の処理およびデータ均衡化処理については使用・不使用で複数の識別モデルを構築する。また、RandomForestClassifier 関数内で標準化処理が行われるため、分類器が RandomForest の場合のみ追加の標準化処理を実行しない。

5.2 前処理

特徴量計算に必要な処理として、計算に必要なデータの抽出、ウィンドウの分割、スライディングウィンド処理、そして欠損データの除外を行った。この4つの処理を総じて本研究における前処理と呼ぶ。まずアイトラッカーによって得た視線データから視点の座標、左右の眼球の座標、左右の瞳孔径、有効性、眼球運動の種類、眼球運動時間および固視座標などを抽出した。ここで、有効性とはアイトラッカーが認識した眼球の認識の有効性を表す。また、タイムスタンプの時刻の単位は μs 、間隔は約 16ms (60Hz) で眼球運動時間の単位は ms とした。次に、各データのウィンドウ単位での分割によりインスタンスを生成した。さらに、ウィンドウの分割点をウィンドウサイズの半分ずらして分割するスライディングウィンド処理を行った。例えばデータを 4 秒単位で分割する場合、図 6 の分割 1 として示した時刻での分割に加えて、分割 2 として示した時刻での分割も行うことで、インスタンス数を約 2 倍に増やすことができる上に、一方の分割点をまたぐ形で存在する特徴をもう一方の分割点によるインスタンスにて取得可能となる。

データ分割後、表 2 に示す 5 種類に該当するデータを欠損データとして除外した。データ収集の際、作業開始・終了および次ページへの遷移はキーボードを用いて行ったが、操作時の視線が黙読中の視線とは異なる可能性があるため、ページ遷移前後 3 秒間のデータを除外した。同様の理由で、作業開始直後 3 秒間のデータならびに作業終了直前 10 秒間のデータを除外した。なお、作業開始時は参加者が作業を開始してから作業時間の計測を開始したのに対し、作業終了時は観察者の合図、すな

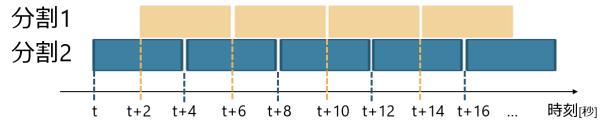


図 6: スライディングウィンド処理

表 2: 欠損データ

番号	欠損データの種類
1	キーボード押下げ前後のデータ
2	一部のデータが欠けたインスタンス
3	インスタンス内の固視が 5 回未満のデータ
4	計算不能な特徴量を生むインスタンス

わち作業時間の計測終了後に参加者が作業を終了した。このため、作業終了時にのみ、計測終了時点から作業終了時点までの時間に比較的時間差がみられた。したがって、作業終了直前のみ除外するデータの時間を 10 秒に設定した。これらの欠損データ除外処理によって、インスタンス内の一部のデータが欠損した場合、そのインスタンスを除外した。加えて、表 2 の 2 に示したデータに対して、そのウィンドウにおいて参加者は黙読を行っていなかったと判断し、そのインスタンスを除外した。ここで、あるインスタンスにおいてサッカーがみられなかった場合、サッカーに関する特徴量が計算不可能となる。このように、一部の特徴量が計算できなかった場合のインスタンスについても除外した。

5.3 特徴量計算

読書中の内在状態を視線により推定する先行研究 [15, 20] で用いられている視線特徴を参考に、表 3 に示す合計 13 種類の視線特徴を使用した。表 3 のうち、固視時間、サッカー時間、サッカー振幅、サッカー速度、絶対サッカー角度分布、相対サッカー角度分布、左目の瞳孔径および右目の瞳孔径の 8 種類の視線特徴に関しては、最大値、最小値、平均値、中央値、標準偏差、尖度、歪度、範囲の 8 つの記述統計量を計算し、その結果を特徴量として用いた。ゆえに、特徴量の次元数は 8 つの記述統計量を計算する 8 種類の視線特徴と残りの 5 種類の視線特徴の合計で 69 次元となった。また、固視の定義は [15] を参考に最大 100ms 以上の間、57 ピクセル (視野角約 1 度) の範囲内で連続した注視点とした。特徴量計算後、平均値から標準偏差の 3 倍以上離れたデータを外れ値として定義し、外れ値をその範囲の最大値または最小値の近い方の値に置き換えた。最大値および最小値の計算では訓練データのみ使用し、置き換えの処理は訓練データとテストデータの両方に適用した。

6 識別性能調査実験

6.1 各パラメータの組み合わせでの交差検証

表 4 に集中状態・非集中状態識別システムにおけるパラメータをまとめる。表 5 に示すように、ウィンドウサ

表 3: 視線特徴の種類

眼球運動	視線特徴	説明
固視	固視時間 [ms]	インスタンス内のある固視にかかった時間
	固視回数	あるインスタンスにおけるすべての固視の回数
	固視分散	インスタンス内の全国視座標の平均と各固視座標との距離の二乗平均の平方根
サッカード	サッカード時間 [ms]	インスタンス内のあるサッカードにかかった時間
	サッカード回数	あるインスタンスにおけるすべてのサッカードの回数
	サッカード振幅 [pixel]	2つの固視点間のピクセル数
	サッカード速度 [pixel/ms]	サッカード振幅/サッカード時間
	相対サッカード角度分布	2つのサッカードの間の角度
	絶対サッカード角度分布	2つの固視点間の線分と x 軸の間の角度
	水平サッカード割合	x 軸の上下に 30 度以内の角度を持つサッカードの割合
	固視時間/サッカード時間	固視時間とサッカード時間との比
瞳孔	左目の瞳孔径 [mm]	左目の瞳孔の大きさ
	右目の瞳孔径 [mm]	右目の瞳孔の大きさ

イズの大きさによってインスタンス総数は異なる。インスタンス件数の影響を含まない純粋なウィンドウサイズの効果を確かめるために、各ウィンドウサイズにおけるインスタンスについて、集中状態が 3490 件、阻害状態が 3508 件、疲労状態が 3446 件となるよう件数を統一した。この処理はすべての識別モデルに対して実行した。

識別モデルの性能を最も向上させる分類器、ウィンドウサイズおよび外れ値処理の有無は未知である。そのため、これらのパラメータをそれぞれ変更し、合計 40 通りの識別モデルを構築した。データ均衡化についても同様に変更し延べ 120 通りの識別モデル作成を検討したが、外れ値処理無しかつオーバーサンプリングを行った場合の組み合わせの識別モデルにて学習が収束しなかったため、サンプリングを行わず 40 通りの識別モデルを構築した。その後、最も高精度であった識別モデルにおける分類器、ウィンドウサイズならびに外れ値処理の組み合わせに対して、データ均衡化手法を変更し、3 種類の識別モデルの生成および検証を行った。

表 4: 識別処理におけるパラメータ

項目	内容
分類器	Random Forest, SVM, kNN, Naive Bayes
ウィンドウサイズ	4 秒, 6 秒, 8 秒, 10 秒, 12 秒
データ均衡化	サンプリング無し, オーバーサンプリング アンダーサンプリング
外れ値処理	有り, 無し
特微量グループ	すべて, 固視, サッカード, 瞳孔

パラメータチューニングを含めた交差検証の手順を図 7 に示す。まず、各クラスのデータの割合が均一となる

表 5: インスタンス件数

ウィンドウサイズ	集中状態	阻害状態	疲労状態	合計
4 秒	11203	11184	10789	33176
6 秒	7366	7346	7206	21918
8 秒	5433	5450	5325	16208
10 秒	4262	4294	4224	12780
12 秒	3490	3508	3446	10444

ようにデータセットを 10 分割し、このうちの 1 個をテストデータ、残りの 9 個を訓練データ 1 とした。次に、パラメータチューニングのための層化 4 分割交差検証を行った。これは前述の訓練データ 1 を各クラスのデータの割合が均一となるように 4 分割し、このうちの 3 個を訓練データ 2 として機械学習を行い、残りの 1 個である検証用データを用いて予測を行うというものである。この処理を検証用データを変更して 4 回実行し、4 回の処理を 1 度の層化 4 分割交差検証として、パラメータの数だけ交差検証を繰り返した。パラメータチューニング後、各層化 4 分割交差検証において最も高精度であったときの訓練データ 2 に対して、平均を 0、分散を 1 とする標準化処理およびサンプリング処理を行った上で、改めて機械学習を行い最良モデルを作成した。なお、層化 4 分割交差検証では訓練データに対する標準化処理およびサンプリング処理は実行しなかった。最後に、構築した最良モデルと最初に用意したテストデータを使用して分類を行った。この一連の処理をテストデータを変更して 10 回繰り返し、各交差検証におけるすべての正解データと予測結果を使用して前述の各評価指標を算出し、Random Forest 使用時は延べ 120 回、kNN 使用時は延べ 200 回の学習および予測を行った。

また、分類器の使用には scikit-learn 1.0.1 の関数を用いた。Random Forest のパラメータチューニングでは GridSearchCV 関数によって n_estimator の値 (決定木の数) を 10, 100, 1000 から選定し、機械学習では RandomForestClassifier 関数によって学習した。SVM による分類では svc 関数を使用し、パラメータのうち kernel を 'linear' に、probability を 'True' に設定し、その他のパラメータを既定値に設定した。ここで、SVM ではハイパーパラメータチューニングが可能であるが、チューニングを導入した場合の一部の識別モデルの学習が収束しなかったため、本研究では実行しなかった。kNN を用いる際は Random Forest と同じくパラメータチューニングを行った。具体的には、k の値とし

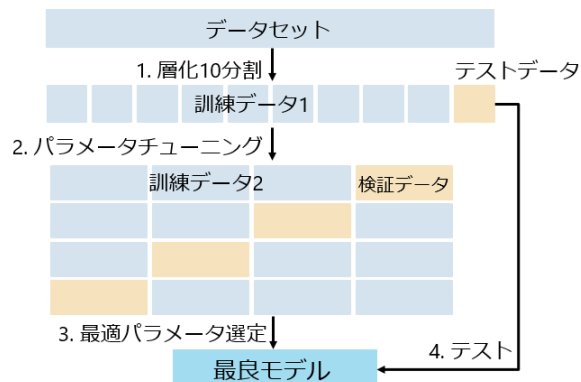


図 7: パラメータチューニングを含めた検証手法

て 1, 2, 3, 4, 5 から最もパフォーマンスの良いパラメータを求め、そのパラメータを使用した識別モデルを構築した。NaïveBayes のパラメータは GaussianNB 関数で設定された既定値を使用した。

データ均衡化におけるオーバーサンプリング手法として、近傍の少数クラスのデータから新しいデータを作成する処理である SMOTE[27] を用いた。SMOTE における近傍のデータ数は 5 個に設定した。アンダーサンプリングの際は訓練データ内でクラスごとのデータの件数を最も少ないクラスの件数に統一した。識別モデルの評価指標として、適合率、再現率、F 値ならびに各評価指標のマクロ平均と正解率を使用した。

分類器、ウィンドウサイズ、外れ値処理の有無を変更して全体層化 10 分割交差検証を実施した結果、分類器が RandomForest, ウィンドウサイズが 12 秒, 外れ値処理有りの場合に最も高い F 値が得られた。外れ値処理を実行した場合の交差検証の結果として得られた F 値を表 6 に示す。この組み合わせにおいてデータ均衡化手法を変更して交差検証を実施したところ、サンプリング(データ均衡化) 無しの際に F 値が 0.932 となり最高性能を示した。

表 6: パラメータ変更時の識別結果

	4 s	6 s	8 s	10 s	12 s
Random Forest	0.616	0.702	0.782	0.860	0.932
SVM	0.436	0.451	0.454	0.474	0.477
kNN	0.553	0.631	0.731	0.818	0.912
Naïve Bayes	0.394	0.412	0.414	0.418	0.421

以上の検証では、各ウィンドウサイズにおけるインスタンス件数を統一した場合の識別モデルの性能について分析したが、同じ作業時間のデータを用いる場合にはウィンドウサイズが短いほど多くのデータを取得・利用可能となり、データ数の増加が精度向上の上では強みとなりうることから、データ数の影響を排除して検証についても行うべく、インスタンス件数を統一せず上記の手順と同様の識別モデル選択を行った。その結果、ウィ

ンドウサイズが 4 秒, 分類器が RandomForest, 外れ値処理無し, サンプリング無しの場合の F 値が最も高かった。外れ値処理無し, サンプリング無しの場合かつ, 分類器とウィンドウサイズを変更した場合の交差検証結果を表 7 に示す。

表 7: インスタンス件数が異なる場合の交差検証結果

	4 s	6 s	8 s	10 s	12 s
Random Forest	0.949	0.920	0.925	0.927	0.930
SVM	0.435	0.452	0.467	0.474	0.479
kNN	0.937	0.905	0.910	0.907	0.911
Naïve Bayes	0.338	0.378	0.396	0.336	0.407

インスタンス件数によらない識別モデルの性能を調査するため、以降はインスタンス件数を統一した場合に最も高評価であった組み合わせとして、分類器を Random Forest に、ウィンドウサイズを 12 秒に設定し、外れ値処理の実行およびサンプリング無しの組み合わせにより識別モデルを構築する。また、この最も高い評価を得る組み合わせを最良の組み合わせと呼び、構築されたモデルを本稿における最良モデルとする。

6.2 各種交差検証

6.1 節で決定した最良モデルに対し、全体層化 10 分割交差検証の他に、2 種の交差検証を実行した。1 人分の参加者のデータを全データセットとする個人内層化 10 分割交差検証を行い、テストデータと訓練データが同一人物のデータで構成される理想条件下での識別性能を調査した。さらに、1 人分の参加者のデータをテストデータ、その他の参加者のデータを訓練データとする 1 人抜き交差検証を行うことで、識別モデルにとって未知の被験者に対する識別性能を調査した。個人内層化 10 分割交差検証における混同行列を図 8 に、1 人抜き交差検証における混同行列を図 9 に示す。

表 8: 各交差検証の結果

	正解率	適合率	再現率	F 値
個人内層化 10 分割交差検証	0.945	0.945	0.945	0.945
全体層化 10 分割交差検証	0.913	0.911	0.913	0.912
一人抜き交差検証	0.428	0.427	0.427	0.425

6.3 特徴量グループによる精度の違い

6.1 節で決定した最良モデルにおいて、表 3 に示した特徴量のうち、機械学習にてすべての特徴量を使用した場合、固視に関する特徴量のみを使用した場合、サッカーに関する特徴量のみを使用した場合および瞳孔に関する特徴量のみを使用した場合の 4 通りの識別モデルを構築した。なお、固視時間/サッカー時間ほどの種類にも属さない特徴量として扱った。それぞれの識別モデルに対する全体層化 10 分割交差検証結果を表 9 に示す。

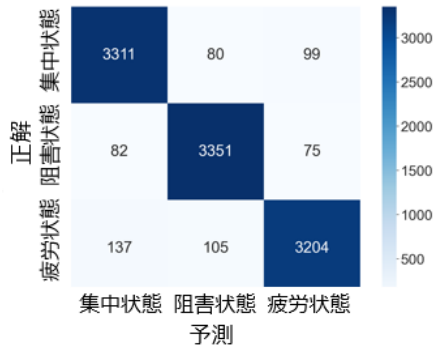


図 8: 個人内交差検証における混同行列

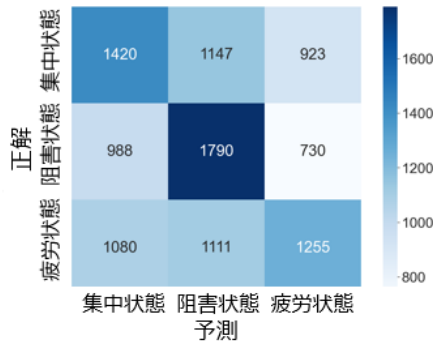


図 9: 一人抜き交差検証における混同行列

7 考察

7.1 パラメータ選択について

6.1 節の結果においてウィンドウサイズに注目すると、すべての分類器においてウィンドウサイズが 12 秒の場合に最も高い評価が得られた。したがって、ウィンドウサイズの違いのみに着目すると 12 秒と設定するのが最も適切であると考えられる。さらにウィンドウサイズの大きさに比例して F 値の上昇がみられるため、12 秒より大きいウィンドウサイズにおいてさらに高い F 値が得られる可能性がある。このような F 値の増加傾向が見られた原因として、ウィンドウサイズが大きいことの利点であるインスタンス内のデータ数の増加が挙げられる。したがって、時間的に連続した視線情報をより多く活用することによって識別性能が向上すると考えられる。

一方で、表 7 に示したインスタンス件数非統一時の結果では、分類器によって最高精度を示すウィンドウサイズが異なった。SVM ではインスタンス件数統一時と同様に、ウィンドウサイズの拡張に伴う精度向上傾向が見られた。本研究では SVM のパラメータを既定値としたが、ハイパーパラメータチューニングにより精度向上が考えられる。RandomForest および kNN では、ウィンドウサイズが 4 秒の場合に最高精度が示され、ウィンドウサイズが 12 秒の場合に二番目に高い精度が示された。こうした精度の変化は、トレードオフの関係にあるインスタンスの数と 1 つのインスタンスが持つデータの数が影響したと考えられる。NaiveBayes では、ウィンドウサ

表 9: 種別特徴量選択時の全体層化 10 分割交差検証結果

	正解率	適合率	再現率	F 値
すべて	0.932	0.932	0.932	0.932
固視	0.908	0.908	0.908	0.908
サッカー	0.926	0.926	0.925	0.925
瞳孔	0.911	0.912	0.911	0.911

イズが 12 秒の場合に最高精度が示された。こうした分類器およびウィンドウサイズの変化が識別性能に及ぼす影響を明確にするためには、ウィンドウサイズの指定範囲を拡張した追加検証が必要である。

7.2 個人依存性について

6.1 節に示した結果において、全体層化 10 分割交差検証ならびに個人内層化 10 分割交差検証結果では全ての評価指標が 0.9 を超える高評価が得られたが、一人抜き交差検証の結果は他の 2 つの交差検証と比較して著しく低評価であった。これは、個人によって読む速さや読み返す頻度が異なることから視線移動に個人差が生じたためであると考えられる。この原因を考察するため、一人抜き交差検証における各クラスの適合率、再現率ならびに F 値を表 10 に示す。ここで、全クラスを対象に正解率を求めた結果、0.428 であった。表 10 において、阻害状態におけるすべての評価は他の状態の評価よりも高い。さらに、阻害状態の適合率と再現率を比較すると、再現率のほうが高い。したがって、作成した識別モデルは正解ラベルが阻害状態であるインスタンスの正確な予測に長けており、阻害状態の検出に強いと想定される。この原因として、n-back タスクのメインタスクへの阻害が大きいことが予想される。実作業環境にて n-back タスクと同様の作業と並行して行う場面は少ないことから、メインタスクへの阻害が比較的小さいタスクを用いた場合の識別性能についても調査が必要である。

表 10: 一人抜き交差検証における各クラスの識別精度

	適合率	再現率	F 値
集中状態	0.407	0.407	0.407
阻害状態	0.442	0.510	0.474
疲労状態	0.432	0.364	0.395

7.3 特徴量について

6.3 節で述べた特徴量の種別選択時の識別結果について考察する。表 9 に示した特徴量選択時の全ての評価指標において、サッカー由来の特徴量を使用した場合の精度が他の 2 種の特徴量を使用した場合の精度に比べ高評価となった。固視由来の特徴量の次元数が 10 次元、瞳孔径由来の特徴量の次元数が 16 次元であるのに対し、サッカー由来の特徴量の次元数は 42 次元と最大であることから、特徴ベクトルが持つ情報量の多さが一要因であると考えられる。

8 おわりに

本研究では、オフィスワーク中の集中状態・非集中状態の識別を目的とし、集中状態を一定期間1つの物事に一定以上の認知資源が割り当てられていることと定義した上で、非集中状態が複数の異なる状態から構成されると考え、阻害状態と疲労状態を対象として設定した。特に、PC作業中に普遍的に存在するテキスト黙読作業をメインタスクとして想定し、前述の3状態を促す実験条件をそれぞれ設定し、視線データ計測実験を実施した。得られたデータに対し前処理を施し、識別モデルの構築ならびに予測を行った結果、最大でF値が0.945となった一方で、未知のユーザに対する識別性能を示す一人抜き交差検証でのF値は0.425に留まった。今後は、主に未知のユーザに対する識別精度の向上と、テキスト黙読タスク以外のタスクへの提案手法の応用を検討する。未知のユーザに対する精度向上を行う上で課題となるのは視線特徴の個人化であり、この影響を低減させる可能性のある処理としては標準化処理が考えられる。例えば、読む速さが速い作業者と遅い作業者との個人差については、視線特徴のうちのサッカード速度を標準化することにより緩和されうることから、適用を検討する。また、本研究で構築した識別モデルの黙読タスク以外の、例えばキーボードやマウスを用いた入力を伴う作業に対する汎化性能は未知であるため、個人間および作業間の汎化性能向上を目指す。

謝辞

本研究はJSPS科研費21K17787の助成を受けた。

参考文献

- [1] 総務省. 働き方改革とict利活用-第2節, 2027. <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/pdf/n4200000.pdf> 最終アクセス2022/01/25.
- [2] 内山皓介, 宮城和音, 石井裕剛, 下田宏, 大林史明, 岩川幹生. 作業への集中に着目した知的生産性評価ツールの開発. 2013.
- [3] 宮城和音, 内山皓介, 大林史明, 岩川幹生, 石井裕剛, 下田宏. 知的生産性評価のための集中指標の提案. ヒューマンインタフェース学会論文誌, Vol. 16, No. 1, pp. 19–28, 2014.
- [4] Yuji Uema and Kazutaka Inoue. Jins meme algorithm for estimation and tracking of concentration of users. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pp. 297–300, 2017.
- [5] 深澤伸一, 下村義弘. 周辺視野における通知の動きの種類が気づきやすさと作業集中性に及ぼす影響. 人間工学, Vol. 57, No. 3, pp. 119–128, 2021.
- [6] Ryosuke Yoshihashi, Daiki Shimada, and Hitoshi Iyatomi. Feasibility study on evaluation of audience's concentration in the classroom with deep convolutional neural networks. In *2014 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, pp. 288–292. IEEE, 2014.
- [7] Pascual Martínez-Gómez and Akiko Aizawa. Recognition of understanding level and language skill using measurements of reading behavior. In *Proceedings of the 19th international conference on Intelligent User Interfaces*, pp. 95–104, 2014.
- [8] Henning Krampe, Lars Johan Danbolt, Annie Haver, Gry Stålsett, and Tatjana Schnell. Locus of control moderates the association of covid-19 stress and general mental distress: results of a norwegian and a german-speaking cross-sectional survey. *BMC psychiatry*, Vol. 21, No. 1, pp. 1–13, 2021.

- [9] Mirosław Pawlak, Mariusz Kruk, and Joanna Zawodniak. Investigating individual trajectories in experiencing boredom in the language classroom: The case of 11 polish students of english. *Language Teaching Research*, p. 1362168820914004, 2020.
- [10] Virginia Lemay, John Hoolahan, and Ashley Buchanan. Impact of a yoga and meditation intervention on students' stress and anxiety levels. *American journal of pharmaceutical education*, Vol. 83, No. 5, 2019.
- [11] Han Zhang, Kevin F Miller, Xin Sun, and Kai S Cortina. Wandering eyes: Eye movements during mind wandering in video lectures. *Applied Cognitive Psychology*, Vol. 34, No. 2, pp. 449–464, 2020.
- [12] Tiffani Razavi. Self-report measures: An overview of concerns and limitations of questionnaire use in occupational stress research. *University of Southampton - Department of Accounting and Management Science, Papers*, 2001.
- [13] Dario D Salvucci and Joseph H Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pp. 71–78, 2000.
- [14] Richard J Krauzlis, Laurent Goffart, and Ziad M Hafed. Neuronal control of fixation and fixational eye movements. *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol. 372, No. 1718, p. 20160205, 2017.
- [15] Myrthe Faber, Robert Bixler, Sidney KD' Mello. An automated behavioral measure of mind wandering during computerized reading. *Behavior Research Methods*, Vol. 50, No. 1, pp. 134–150, 2018.
- [16] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pp. 439–448, 2013.
- [17] Bastian Pfleging, Drea K Fekety, Albrecht Schmidt, and Andrew L Kun. A model relating pupil diameter to mental workload and lighting conditions. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pp. 5776–5788, 2016.
- [18] Hideyuki Kanematsu, Dana M Barry, Tasuya Shirai, Masashi Kawaguchi, Nobuyuki Ogawa, Kuniaki Yajima, Katsuko T Nakahira, Toshiro Kobayashi, and Michiko Yoshitake. Measurements of eye movement and teachers' concentration during the preparation of teaching materials. *Procedia Computer Science*, Vol. 159, pp. 1499–1506, 2019.
- [19] JINS. Jins meme. <https://jinsmeme.com/> 最終アクセス2022/01/24.
- [20] Rosy Southwell, Julie Gregg, Robert Bixler, and Sidney K D'Mello. What eye movements reveal about later comprehension of long connected texts. *Cognitive Science*, Vol. 44, No. 10, p. e12905, 2020.
- [21] Sidney K D'Mello, Caitlin Mills, Robert Bixler, and Nigel Bosch. Zone out no more: Mitigating mind wandering during computerized reading. *International Educational Data Mining Society*, 2017.
- [22] Tobii Pro. Tobii pro ナノ. <https://www.tobiiipro.com/ja/product-listing/tobii-pro-nano/> 最終アクセス2022/01/24.
- [23] 株式会社スクールパートナーズ. 高校生新聞オンライン. <https://www.koukouseishinbun.jp/> 最終アクセス2022/01/25.
- [24] Thomas Kosch, Mariam Hassib, Paweł W Woźniak, Daniel Buschek, and Florian Alt. Your eyes tell: Leveraging smooth pursuit for assessing cognitive workload. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2018.
- [25] Ying Wang, Bryan Reimer, Bruce Mehler, Jun Zhang, Alea Mehler, and Joseph F Coughlin. The impact of repeated cognitive tasks on driving performance and visual attention. In *Proceedings of the Third International Conference on Applied Human Factors and Ergonomics*, 2010.
- [26] Bruce Mehler, Bryan Reimer, and Jeffery A Dusek. MIT AgeLab delayed digit recall task (n-back). *Cambridge, MA: Massachusetts Institute of Technology*, Vol. 17, , 2011.
- [27] Alberto Fernandez, Salvador Garcia, Francisco Herrera, and Nitesh V Chawla. SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *J. Artif. Intell. Res.*, Vol. 61, pp. 863–905, 2018.