

脳波・心拍変動解析による機械学習を用いた ヒューマンエラーの予測手法の提案

齋藤 勇斗¹ 高田 尚子²

Muhammad Nur Adilin Bin Mohd Anuardi¹ 菅谷 みどり¹

概要：近年、ヒューマンエラーによる労働災害が増加している。ヒューマンエラーの精神的状態の分析では、生体情報を用いた手法が提案されている。しかし、精神状態を生体情報で計測し、その結果に基づき正確な予測を行う研究はまだ十分ではない。本研究では、無意識的な精神状態として脳波・心拍で計測し、また疲労やストレス値をアンケートで補う事で、エラーの予測モデルの構築およびリアルタイム予測を行う事を目的とする。実験では、協力者に数日に渡りストループ課題を実施してもらい、その際に得られた脳波・心拍とアンケート結果を用いて、個人ごとのエラーの予測モデルを提案した。結果として、脳波・心拍とアンケート結果の一部指標において、エラーに関係するという事が得られ、これをエラーの予測モデルに組み込んだ。さらに、エラーの予測をリアルタイムで行い、ヒューマンエラーを未然に防止する事ができるかどうかを検証した。結果として、エラーと予測された際、74%でエラーの発生が確認された。

キーワード：ヒューマンエラー、生体情報、機械学習

1. はじめに

平成28年度の製造業における労働災害は約2万8千件であり、うち97.1%が労働者の不安全行動に起因している[1]。不安全行動とは、労働者の安全を阻害する可能性のある行動を意図的に行う行為であり、不安全行動による労働災害の発生には、ヒューマンエラーが大きく関係しているとされる。ヒューマンエラーとは、意図しない結果を生じる人間の行為であり、意図しない不安全行動による労働災害を抑制するためには、その要因として重要なヒューマンエラーを事前に防ぐ必要がある。ヒューマンエラーを事前に防ぐためには、ヒューマンエラーに関わる複数の要素や、人の事前の精神状態など、ヒューマンエラーと関連のある事項を客観的に分析することが求められる。特に近年では、こうした精神状態との関わりが指摘されており[2,3]、これらの研究が重要である。

ヒューマンエラーを発生させる際の精神的状態の分析では、無意識の精神状態を検知可能である生体情報を用いた手法が提案されている[2,3,4]。生体情報は、無意識に発生する精神状態を取得できることから、高度に発展し複雑な判断を求められる状況下で、重大な事故を防止するための予測などを行う事に適していることが指摘されている[3]。

Markらは、高度に複雑化した作業におけるヒューマンエラーを定量的に評価出来る指標として生体情報に着目した[2]。EEG、心拍、皮膚電気など計20種類の生体情報を取得しながら実験協力者に軍事シミュレーターのタスクを行い、

生体情報の変化を記録した。その生体情報からタスク時の作業負荷の度合いなど、心理学的アプローチだけでは知り得ない情報を得ることが出来たとされている[2]。しかし、評価項目が多すぎることから一般的な評価実験には適さない課題がある。吉田らは、中枢神経計測指標として脳波、自律神経指標として心拍変動値を用いて生体計測の指標数を減らしつつ、ミスの事前状態と事後状態とミスの相関を調査し、ミスの事前状態が脳波・心拍の状態と相関があることを明らかにした[4]。しかし、吉田らは予測モデルの構築は行っておらず、正確な予測を行うには十分とはいえない課題がある。さらにエラーを未然に防ぐリアルタイムシステムについては言及されていない[4]課題があった。

そこで、本研究では、エラーの予測モデルの構築およびリアルタイム予測を行う事を目的とした。実現のために、吉田らが用いた精神状態推定の代表的な計測指標である脳波・心拍計測値をもとに、ロジスティック回帰により予測モデルを構築した。実験では、協力者に数日に渡りストループ課題を実施してもらい、その際に得られた脳波・心拍とアンケート結果を用いて、個人ごとのエラーの予測モデルを作成した。さらに、エラーの予測をリアルタイムで行い、ヒューマンエラーを未然に防止する事ができるかどうかを検証した。結果として、エラーと予測された際、74%でエラーの発生が確認された。

本論文の構成は次の通りとする。2節にて関連研究とその課題、3節にて目的と提案、4節にて予備実験に向けての開発、5節にて実験1、6節にて設計と実装、7節にて機械学

¹ 芝浦工業大学
Shibaura Institute of Technology.
² (株)日立製作所
Hitachi Ltd.

習とリアルタイム予測, 8 節にてまとめと今後の課題を述べる。

2. 関連研究

2.1 ヒューマンエラーに関する研究

加藤らは、看護師のヒューマンエラー要因の明確化を目的とし、医療事故の抑制を目指した。具体的には、実際にあった3つの医療事故について P-mSHELL モデルと有識者の意見を踏まえてエラー要因抽出を行った[5]。P-mSHELL モデルとは河野らがヒューマンエラー分析モデルである SHEL モデルに patient 要素を加えて医療用に提案したモデルである[6,7]。3つの事例研究について、看護師のヒューマンエラーによる医療事故の事例から「役に立たない手順書」と「不十分なコミュニケーション」がいずれの事例においても抽出され、エラー誘発要因を明確化したとされている[5]。しかし、創造的で高度に複雑化した作業におけるエラーについて、この認知モデルを用いた分析を行う場合、そのモデルは複雑化しエラー要因の抽出が出来たとしても他の事例への活用が困難である可能性がある。今後人が従事する作業は創造的であることや、複雑度が高いなど、高度なものとなる事が予測される。そうした作業におけるヒューマンエラーを統計的かつ簡易的に予測・管理する手法が必要である。

これに対して、近年、ヒューマンエラーの精神的状態の分析では、生体情報を用いた手法が様々提案されている。Mark ら[2]は、EEG、心拍、皮膚電気など計20種類の生体情報を取得しながら実験協力者に軍事シミュレーターのタスク(レーダー上の不審な飛行機の発見タスク)を行った。タスク負荷は、一次タスク中に出現する航空機の量と種類を変えたり、二次言語記憶タスクの有無を変えたりする事で操作した。その結果、生体情報でタスク中の認知活動の変化が有意に確認され、生体情報にミスの影響が現れる事を示した。しかし、評価項目が多すぎる事から一般的な評価実験には適さない課題がある。

これに対して、岸本らは[4,11] 作業環境に対して不快感を抱いている場合、ヒューマンエラーの発生率が上昇するという点に着目し、脳波および心拍を用いたヒューマンエラー調査を行い、作業中のヒューマンエラーにおける生体計測による評価方法の検討を行った。実験では、ヒューマンエラーが発生しやすい環境下での作業における生体情報の特徴調査を元に計算課題を実施した。また、ワークロード測定に用いられる NASA-TLX[12]を負荷作業時における作業環境の評価として用いた。エラーの前後で変化する事を元にパターン化し、そのパターンの出現度を評価した。その結果、心拍の一部のパターンのみ有意差が確認された。しかし、単純なパターン化のみで評価している事から、予測に十分な判定モデルを検討する事はなされていない課題がある。

これに対して、吉田ら[3]は、岸本らの実験手順の不備な

どを見直した上、作業者にストループ課題の負荷作業を与えた。これは、脳波、心拍を用いて、ヒューマンエラーの起こる前と後の比較を行った。その結果、脳波、心拍の値に有意なヒューマンエラーの予測式や相関値を得た[10]。また、これを活用する事で、エラーの予防策・善後策へ応用可能である事を確認できた。このように、吉田らは、適合的指標選択によるヒューマンエラー予測手法の検討した。しかし、ヒューマンエラーの起こる前と後の比較では、負荷作業の前、または後の安静時の状態を用いており、ヒューマンエラーが発生した過程の生体情報の変化を詳細に考慮できているとはいえない課題がある。また、生体情報の個人差を考慮していない課題がある。

2.2 課題

吉田らの研究では、ヒューマンエラーの事前、事後の安静時の状態を、エラーの事前状態と定義して予測を行った[2]。ヒューマンエラーの負荷作業の前の安静時の状態や、後の安静時の状態は、それぞれ、ヒューマンエラーの事前状態と事後状態とする一方、作業開始後の心理的な過程についての変化を考慮することはできない。例えば、作業中に連続でエラーが発生した場合とそうでない場合では、心理的な過程が異なると考えられる。この事から、ヒューマンエラーの解析や、予測モデルの構築のためには、エラー発生について短い時間区間での評価が必要であると考えられる。さらに、吉田らの研究では、ヒューマンエラーに関する事前、事後状態を、複数人のテスト結果をもとに統計分析を行なっていることから、個人差を考慮できていない課題がある。作業者や労働環境によって生体情報は大きく異なるため、個人差を考慮したヒューマンエラーの予測モデルを構築する必要がある。

3. 目的と提案

3.1 目的

先行研究におけるヒューマンエラーの予測における生体情報を用いた研究では、次の課題が挙げられる。

1. 評価項目や実験機器が多い
2. ヒューマンエラーが発生した過程の生体情報の変化を詳細に考慮できていない
3. 個人差を考慮できていない

本研究の目的は、上記課題を解決し、ヒューマンエラーの予測モデルの構築およびリアルタイム予測を行う事とする。実現のために、本研究では脳波・心拍の計測値と、疲労状態などの事前アンケートから、エラー時の精神状態を予測するための機械学習モデルを作成する。また、ヒューマンエラーの予防策として、エラー予測モデルを用いたリアルタイム予測システムを提案する。分析にあたり、課題点を改善する。改善策を次に述べる。

- (1) 実験機器を減らす。
- (2) ヒューマンエラーの状態を時間区間で定義する。

(3) 個人差を考慮した予測モデルを構築する。
本研究では、特にヒューマンエラーの調査において、ストループ課題を用いたテストを行う。ストループ課題は、文字意味と文字色というように同時に目にする情報が干渉し合う現象をもとに作成されたテストであり、微妙な精神状態や疲労などの影響を受けやすい。テストにて発生したエラーをヒューマンエラーとして本目的を達成するものとした。さらに (1)については、装着が容易な脳波計(Mindwave Mobile 2)、および、心拍計(Pulse Sensor)を用いてデータを収集する。(2)については、エラーの事前、事後状態をエラーの発生について詳細な時間区間で評価する。(3)については、説明変数として、エラーと関係のある脳波・心拍と事前アンケートを検討する。モデル構築のための機械学習にはロジスティック回帰を用いた。また、負荷作業中における、エラーのリアルタイム予測を行う。このシステムを提案するにあたり、分析で得られたエラーの予測モデルを用いる。さらに、エラーのリアルタイム予測を行うシステムの精度を検証した。

3.2 評価指標

エラー時の精神状態を分析・検証する上で、評価指標は脳波・心拍と事前アンケートを用いる。

3.2.1 脳波・心拍

本研究では、必要最小限の生体情報を用いる。吉田、岸本らの研究では、中枢神経系の指標である脳波と、自律神経系の指標である心拍変動指標を用いることで、脳の覚醒状態と身体の緊張・リラックスといった生体変化を評価指標により計測した[3,4]。このことから、本研究においても用いるものとした。

ヒューマンエラーの研究において、脳波は様々な研究で用いられている[10]。精神活動や意識の状態に伴って変化する事から、ヒューマンエラーの際の詳細な精神状態を分析する事ができると考えられる。本研究で対象とする脳波指標と心理状態を、表 1 に示す。

表 1 取得・算出した脳波指標
Figure 1 EEG indices acquired and calculated.

指標	周波数帯(Hz)	心理状態[12][13]
低 α	8 - 9Hz	リラックス, 平穏, 意識的
高 α	10 - 12Hz	リラックス, 良い集中
低 β	13 - 17Hz	思考, 事故及び環境の認識
高 β	18 - 30Hz	警戒, 動揺
低 γ	31 - 40Hz	記憶, 高次精神活動
中 γ	41 - 50Hz	視覚情報処理

これに対し、自律神経を計測する手法として心拍変動指標(HRV)がある。HRV は心臓の自律神経の緊張の指標であり、ヒューマンエラーによる緊張の状態を詳細に分析する

事が可能であると考えられる。本研究では、池田[11]らの研究などで、精神的な状態を判定するために用いられていた心拍変動指標 pNNx を用いる[11]。pNNx は、心拍変動の揺らぎから、副交感神経の変動を評価する事ができ、リラックス状態を計測する事ができる。pNNx の算出方法は、隣接する心拍間隔の絶対値が x ms を超えた割合から算出する。

3.2.2 疲労蓄積度

本研究の目的にある、ヒューマンエラー時の精神状態を予測するモデルを機械学習にて作成するために、生体情報以外に、疲労蓄積度を用いる。時間外労働や一定時間以上の時間外労働による疲労の蓄積は、ヒューマンエラーを引き起こす重大な要因であると考えられる。この事から、本研究におけるエラーに関わる事前状態、発生過程における精神状態を評価できると判断した。評価のための事前アンケートとして厚生労働省が発表している労働者の疲労蓄積度自己診断チェックリスト[16]を用いた。

3.2.3 ストレス度

疲労蓄積度と同様に、人の自覚精神状態を示す方法として、ストレス度の調査がある。ストレス度の調査方法として、厚生労働省が発表している職業性ストレス簡易調査票[17]を用いる。この調査票は、回答者が現在抱えているストレスの要因や度合について分析する事を目的として作られている。また、心理的ストレス反応として、ネガティブな反応のみでなく、ポジティブな反応も含めて調査する事ができるとして、信頼性や妥当性が高いといわれている。この事から、本研究における、エラーに関わる事前状態、発生過程における精神状態を評価できると判断した。

3.3 実験課題

本研究では、目的に示したように、脳波・心拍変動値と、疲労・ストレス状態の事前アンケートからデータを分析し、エラー時の精神状態を予測するための機械学習モデルを作成する。目的を達成するために、ヒューマンエラーにつながる精神負荷テストを行い、その際のデータを取得する。精神負荷テストには、ストループ課題を用いた。

開発したストループ課題のインターフェイスを図 1 に示す。ストループ課題においては、作業者は、文字の”意味”と文字の”色”が異なる 2 パターンにおいて、正しい意味もしくは色を回答する。意味の理解と、色の理解という 2 つの情報が干渉し合うことで、精神負荷がかかる状態となる。



図 1 実験に用いるストループ課題の UI

Figure 1 UI of the Stroop task used in the experiment.

例えば図 1 において文字色を答える場合、文字色である

「赤」に文字意味である「みどり」が干渉する。また文字意味を答える場合、文字意味である「みどり」に文字色である「赤」が干渉する。ストループ課題をヒューマンエラー発生傾向に用いる妥当性については、吉田[3]、重森[18]らの研究で示されている。このことから負荷作業に用いる事は妥当だと判断した。本研究では、ストループ課題を汎用的に利用するため、ブラウザ上で実行できるように開発を行った。

4. 予備実験

4.1 目的と概要

提案に示した(2)で、ヒューマンエラーに関する状態を時間区間で定義する、と述べたように、ヒューマンエラーについて詳細な時間区間での評価を実現する必要がある。そのために、本節では、取得したデータから、ヒューマンエラーの発生前後における時間区間の定義を検討する。

まず予備実験では、脳波計、脈拍計から生体情報を取得する。次に、連続するヒューマンエラーの前後で、複数の時間区間での計測結果を比較・検証した。この検証には、各ヒューマンエラーの前後と初回安静時の比較を行う。初回安静時の状態を用いる理由は、初回の安静時は精神負荷課題を遂行する前であり、その人の現状での生体情報の基準値として考える事ができるためである。これらの比較により、生体情報が基準値から離れる時にヒューマンエラーが発生するという仮説を検証可能であると考えた。実験は1人につき1回行った。実験協力者は4名(20代成人男性2名：実験協力者AとD、女性2名：実験協力者BとC)である。

4.2 実験手順

予備実験は以下の手順で行った。

- 3.3で述べた課題プログラム起動後、脳波計、脈拍計を装着する
- 課題の説明を読む
- 90秒間安静にする(以降安静1)
- 1分間ストループ課題を実施する(文字意味を答える)
- 90秒間安静にする(以降安静2)
- 1分間ストループ課題を実施する(文字色を答える)
- ストループ課題の結果をダウンロードする

このストループ課題プログラムにおいて、未回答および誤答をヒューマンエラーとする。実験環境は静寂が保たれている部屋で行うように指示した。

6のストループ課題の処理は、回答すれば次の問題が出題され、未回答の場合は1.2秒後に次の問題に移動する。また、問題はランダムで32問出題され、次の問題が出題されるまでの時間は500msから2000msの間で決まる。次の出題までの時間を短い時間に設定し、課題の暗記を防いだ。

記録データとして1)次のフローに移動する時の時刻、2)問題回答をした時刻、3)問題回答に要した時間、4)問題回答の正誤判定、5)問題回答の正答率、6)ストループ効果の有る問題の平均回答時間、7)ストループ効果の無い問題の平均回答

時間とした。

4.3 結果分析と考察

実験協力者Aは脈拍計系から取得した値が異常値であった。また、実験協力者C、Dにおいては、実験時、静かな環境静寂が保たれていなかった事から、実験協力者Bについてのみ考察する事とした。予備実験のデータ分析を行うために、(1)ヒューマンエラーの起こる前と、起きた後の2つの時間区間の定義を行った。また、(2)ヒューマンエラーの発生前後と初回安静時の比較方法として、一元配置分散分析を行った。

まず(1)は、吉田らの研究で用いられているエラーの事前事後を、エラー発生について詳細な時間区間での評価を行うため、新たに定義した「事前状態」と「事後状態」のグルーピングを行った。グルーピングでは、時間区間あたり前後10秒ずつをヒューマンエラーの事前状態と事後状態とした。また、連続(前のエラーから4秒以内)でエラーを起こした場合、その時間区間をミス群と定義する方法として以下2つの手法を検討した。1)2個以上の連続のヒューマンエラー部分を群として1つのヒューマンエラーとしてまとめる(以降ミス群前後10秒区間)。2)連続で別のヒューマンエラーが発生しているとする(以降ミス前後10秒区間)。

この2つの時間区間について、実験協力者Bの実験データを用いて比較を行った。比較は、安静時間区間と、エラーで定義した時間区間(1)2)の各ヒューマンエラーを起こした前後10秒間の合計20秒間のデータを対象とし、一元配置分散分析および下位検定(Turkey法)を実施した。この時に、1)2)の異なる方法でエラー区間を分類した場合に、有意確率 $p<0.05$ に該当する結果の割合を算出した。この有意割合について二群の比率の差の検定を行った結果を図2に示す。

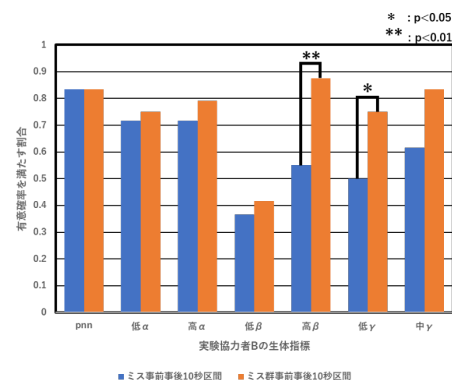


図2 実験協力者Bの生体情報において図7で示した定義の比較

Figure 2 Comparison of the definitions shown in Figure 7 in the biometric data of the experimental collaborator B.

図2について、ミス群前後10秒区間の方が、有意割合が高い結果となった。また生体指標の高βと低γにおいては、ミス群前後10秒区間とミス前後10秒区間において、有意

な差を得られた。このことから、ミス群前後 10 秒区間はミス前後 10 秒区間に比べて、特徴が得られやすいといえる。したがって、今後はこの値を用いて分析するものとした。エラーに関わる生体反応の変化を検討するためには、エラー発生とその際の脳波・心拍変動値の間に、統計的に有意な変化が必要である。先ほどの結果から、有意に差があることが示されたが、それが、生体指標ごとに一貫した傾向を持つかどうかについては、明らかではない。我々は、脳波の値がミスの後は上昇する、心拍変動値は低下するといった一貫した傾向があるかを理解するために、ミス前後の相関分析を行った。

同じく実験協力者 B のデータについて「事前状態」と「事後状態」の心拍変動 pNN10、脳波 (α , β , γ) 値の変化との相関を分析した。各ミス群 10 秒区間の「事前状態」のデータと「事後状態」のデータについてミス群ごとに番号 (ラベル) をふり、13 回のミス群について、前後の 10 秒の相関分析を、ピアソンの積率相関分析にて行った。その結果、有意確率 5%未満の信頼性のある相関は得られなかった。この理由として、10 秒 (10 個) のデータ量ではデータが少ないことが要因であると考えた。

4.4 課題と解決策

実験では先に述べたデータ量の不足の課題と、また、2つ以上のエラー発生を起点とすると、データ取得位置が可変となり、自動的な解析が行いにくいという課題があった。そこで次の実験では、これらの課題を解決するため、負荷作業中の時間を固定 (周期) 区間で分割したデータを自動的に解析することとした。前項では2つ以上のエラー群をまとめ、前後の 10 秒としたが、今回は、固定で 1 分間の課題を 20 秒ずつ区間分割する。統計データは、区間ごとに 20 秒間の生体情報 (20 データ) とエラー数とする。これにより、区間ごとに解析するデータが 20 個と、エラー数が固定的に取得できる。このデータを対象として脳波・心拍変動値とエラー数の相関分析を行うものとした。このことから、次の評価実験 1 では、エラーに関するデータ量を増やすため、ストループ課題プログラムの改善を行った。改善点は以下の 2 つである。1) 予備実験では、ストループ課題が 32 問であったのに対し、実験 1 では 60 秒間 (約 50 問) に変更した。2) プログラムを実行した際に、課題を行う回数は、予備実験が 2 回に対し、実験 1 では 6 回とした。

5. 評価実験 1

5.1 概要

ヒューマンエラーの予測モデルの構築の実現のため、エラーに関わる生体指標の検討分析と検証を目的とする。特に、2つの課題を検証した。

- 1) 予測モデルの説明変数に用いる生体情報
 - 2) 予測モデルの説明変数に用いるアンケート項目
- 1)については、脳波・心拍の生体情報とエラー数の相関分析

を行う。2)については事前アンケートにおいて、精神負荷作業に影響を及ぼす指標の検証をした。この検証には、精神負荷作業中のパフォーマンス指標との相関を用いる。これは、エラーと関係があると考えられるからである。

実験方法は、脳波計と心拍計を装着し、前述の改善した述べたストループ課題を、1日1回ずつ計5日間行った。実験協力者は4名 (20代成人男性2名:実験協力者 A と D, 女性2名:実験協力者 B と C) である。実験手順は予備実験と同一とした。

5.2 実験結果

実験協力者 B の心拍指標の時系列データを図 4 に示す。図について、黒い縦線がエラー発生時を表し、折れ線グラフが心拍指標を表す。さらに、赤い丸で囲まれた部分に着目すると、エラー発生後に心拍指標が低下しており、緊張している。この様に心拍指標の時系列データから、エラー発生時の特徴が見られた。

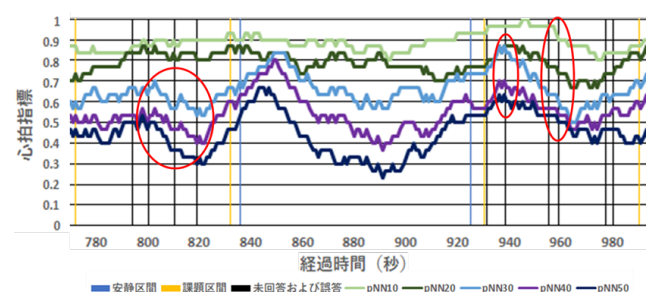


図 3 心拍指標の時系列データ

Figure 3 Time series data of heart rate index.

5.3 解析と考察

5.3.1 脳波・脈拍とエラー数の相関

予測モデルの説明変数に用いる生体情報を検討する。4 人の実験協力者における脳波・脈拍の生体指標情報とエラー数の相関分析を SPSS にて行った。相関分析はピアソンの積率相関係数を用いた。結果を表 2 に示す。

表 2 の列方向は、実験協力者 A, B, C, D の属性を示す。行方向は、脳波・脈拍の生体指標を示す。色の塗られている枠は、有意な相関係数を得られた事示す。各指標に着目すると、中 γ , pNN10, 20, 30, 40 においては、4/4 人 (全員) で有意確率 5%, 1%未満の相関係数を得た。また、低 β , 高 β , 低 γ , pNN50 において、3/4 人で有意な相関係数を得た。これらの指標は、1 名を除く実験協力者で有意な相関係数を得る事ができたことから、エラー時に特徴が見られやすい指標であると考えられる。

表 2 生体情報とエラー数の相関

Table 2 Correlation between biometric data and errors.

** : p<0.01
* : p<0.05

	良い集中成分		緊張(集中)成分					リラックス成分				
	低 α	高 α	低 β	高 β	低 γ	中 γ	pNN10	pNN20	pNN30	pNN40	pNN50	
A	相関係数	0.011	-0.074**	-0.024	-0.077**	-0.095**	-0.110**	-0.205**	-0.153**	-0.169**	-0.110**	-0.075**
	有意確率	0.653	0.002	0.317	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	度数	1813	1813	1813	1813	1813	1813	1813	1813	1813	1813	1813
B	相関係数	-0.019	0.027	-0.053*	-0.166**	-0.176**	-0.164**	-0.137**	-0.183**	-0.205**	-0.214**	-0.174**
	有意確率	0.429	0.255	0.023	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	度数	1824	1824	1824	1824	1824	1824	1824	1824	1824	1824	1824
C	相関係数	-0.054*	-0.044	-0.062**	-0.052*	-0.028	-0.048*	0.054*	0.104**	0.075**	0.057*	-0.015
	有意確率	0.024	0.063	0.009	0.029	0.240	0.045	0.023	0.000	0.002	0.017	0.523
	度数	1769	1769	1769	1769	1769	1769	1769	1769	1769	1769	1769
D	相関係数	-0.079**	-0.080**	-0.066**	0.004	0.057*	0.111**	0.147**	0.117**	0.117**	0.145**	0.132**
	有意確率	0.001	0.001	0.005	0.852	0.015	0.000	0.000	0.000	0.000	0.000	0.000
	度数	1820	1820	1820	1820	1820	1820	1820	1820	1820	1820	1820

反対に、低 α 、高 α において、2/4 人のみ有意な相関係数を得た。これは、半数の実験協力者でしか特徴が見られないことから、エラー時に特徴が見られにくい指標である。また 3.3.1 で述べた様に、表 2 の上部に着目すると、生体情報の指標は、良い集中成分、緊張(集中)成分、リラックス成分に分ける事ができる。

上記結果から、緊張(集中)成分(低 β 、高 β 、低 γ 、中 γ)、リラックス成分(pNN10, 20, 30, 40, 50)ではエラー時に特徴が見られやすく、良い集中成分(低 α 、高 α)は、エラー時に特徴が見られにくいと説明する事も可能である。

5.3.2 指標とエラーの相関

本研究では、 α 波(低 α 、高 α)を良い集中成分、 β 波、 γ 波(低、高)を緊張(集中)成分、pNNxをリラックス成分とした。また、A,B,C,D の正解率は、それぞれ A(85%), B(80%), C(94%), D(85%) であった。結果より、実験協力者 A と D の正解率が 85%、B の正解率が 80%と同程度であるのに対し、C の正解率は 94%と高い結果となった。C は、表 3 では、今回実験対象となった 4 名の中では、エラー前後の相関係数の有意確率が 5%未満の割合が低く(約 73%)、そもそもエラーが少ないこと、また、エラーによる精神的な変化が少ないと考える事ができる。これに対して、実験協力者 C や D は、エラー前後の相関が、11 指標のうちの 80%-90%に有意に現れており、エラー前後に有意に精神的な変化が発生していることがわかる。このことから、エラー発生率が高い協力者は、その前後の生体変化にも相関の高い変化が現れていると言える。

次に、生体情報とエラー数の相関係数値の傾向について分析する。表 2 の 11 の指標の相関係数に着目すると、ほとんどの指標に個人差があり、相関の傾向にはばらつきが見られる。例えば pNN10 から 50 のリラックス成分において、実験協力者 A と B は負の相関であるのに対し、実験協力者 C と D は正の相関である。しかし脳波の緊張(集中)成分では、実験協力者 D 以外において、エラー前後の生体変化では、負の相関関係にある。これは、緊張(集中)が低下すると、エラーが増加する事を示している。

5.3.3 事前アンケートと相関値の傾向

本研究では、作業パフォーマンス評価として、課題中のエラーのばらつきを調査した。データのばらつきは、慣れや疲労との相関を確認することで、精神に与える影響を明らか

にすることを目的とした。事前アンケートは、疲労蓄積度、ストレス度とした。

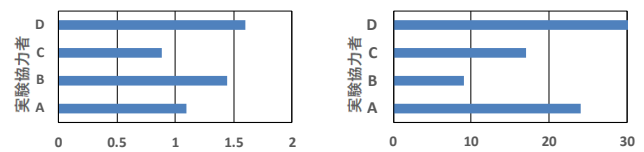


図 4 作業パフォーマンス評価(課題中のエラーの分散) (左) と疲労蓄積度

Figure 4 Work performance evaluation (variance of errors in the assignment) (left) and Level of subjective symptoms of fatigue (right)

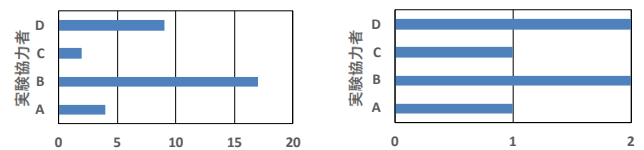


図 5 心理的ストレス度 (左) 身体的ストレス度 (右)

Figure 5 Psychological stress level (left) Physical stress level (right)

図 4, 図 5 に着目すると、実験協力者 D は課題実行時のエラー数の分散が大きく、かつ、疲労蓄積度、身体ストレス度が高い事がわかった。D はミス前後の生体情報において有意確率が最も高い (90%) が得られている。このことから、疲労蓄積や身体的ストレスが、エラー時の一貫した反応に関連していると考えられる。

また、図 5 より、エラーの特徴は、脳波においては良い集中成分よりも、緊張(集中)成分である β 、 γ 、また、心拍変動にみられるリラックス成分に有意な結果が得られている。これらのことから、予測モデルの説明変数に用いる生体情報として、緊張(集中)成分とリラックス成分、事前アンケートを用いることにした。

5.3.3 エラーの分散とストレス度

次に、課題実行時と事前状態の関係について分析した。課題時のデータには、作業パフォーマンス評価である、課題実行時の 20 秒区間のエラーの分散値を用いた。事前状態には事前アンケート結果を用いた。課題時と事前状態の関係の分析には、SPSS によるピアソンの積率相関分析を用いた。その結果、作業パフォーマンスと、心理的ストレス度は、相関係数 0.754(p<0.246)、また、身体的ストレス度は、相関係数 0.946(p<0.054)と、高い正の相関を示した。

次に、連続する日程で行った作業における作業パフォーマンス評価のそれぞれの日と、心理的ストレス、身体的ストレスについて 4 人のデータをもとに相関分析を行った。これは、日毎に異なる課題への慣れの状態や、疲労の状態などの影響を調査するためである。実験日ごとの作業パフォーマンス評価と事前アンケートの相関係数の結果を表 3 に示す。

表 3 日毎の作業パフォーマンスと心理・身体ストレス
の相関係数の比較

Table 3 Each correlation coefficient in Figure 20.

実験日数	1 日目	2 日目	3 日目	4 日目	5 日目
心理的ストレス	0.97	0.96	0.16	0.21	0.28
身体的ストレス	0.75	0.69	0.63	0.03	0.67

表 3 の身体的ストレスに着目すると、1 日目から 3 日、5 日目には、0.6 から 0.75 の正の相関関係が得られた。4 日目のみ、ほぼ相関がみられない状態があった。これに対して心理的ストレスと作業パフォーマンスの相関は、1 日目、2 日目は 0.9 以上であったが、3 日目以降は 0.16, 0.21, 0.28 と大きく低下した。これは、3 日目以降で課題に慣れるに連れて、不慣れによる心理的ストレスの緊張が減り、エラーの分散が減ったと考えられる。これらのことから、課題に慣れていくにつれて、エラーと心理状態との関係性は薄くなる傾向がみられる一方、身体状態は、3 日目に相関が低くなる以外、ほとんど一定の関係性にある事がわかった。

5.3.4 熟練度による学習効果と、漫然の状態

次に実験への慣れをさらに詳細に分析するため、実験協力者 A,B,C,D 全員の課題ごとの正解率の推移を評価した。課題の試行回数と正解率の推移を図 6 に示す。

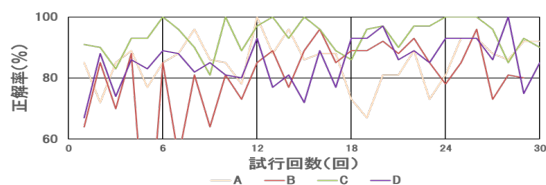


図 6 試行回数と正解率

Figure 6 Number of attempts and percentage of correct answers.

また、実験全体の正解率の累積の変化を線形近似した。この時の傾きを表 4 に示す。結果では、実験協力者 4 名の全ての値が、試行回数の増加につれ増加し、実験協力者全員の正解率の上昇がみられた。これは、課題を繰り返し行う事による、学習効果だと考えられる。結果の中でも、特に実験協力者 B は、0.74 と傾きが最も大きく、学習効果が最も高い。この理由として、1 日目の実験で課題の内容を理解しないまま行い、多くのエラーをし、その後、理解により結果が改善されたと考えられる。

表 4 図 20 における各相関係数

Table 4 Each correlation coefficient in Figure 20.

実験協力者	A	B	C	D
線形近似の傾き	0.10	0.74	0.13	0.35

実験日ごとによる課題への慣れを評価する。同様に実験日ごとの正解率の累積の変化を線形近似した。この時の傾きの値を表 5 にまとめた。表の色が塗られている部分は、傾きが負で、正解率が低下した部分を表す。全体に、3 日目以降から正解率の低下が見られ、最終日には実験協力者全員で正解率の低下が見られた。これは、慣れにより漫然状態が引き起こされた傾向であると考えられる。

これらのことから、学習効果による正解率の上昇が見られる一方、実験日ごとに考察すると、慣れによる漫然状態によって、正解率の低下が見られたと考えられる。

表 5 正解率の累積の変化を線形近似したときの傾き(日毎)

Table 5 Slope of a linear approximation of the cumulative change in the percentage of correct answers (daily)

実験試行日(日 目)	1	2	3	4	5
A	0.54	0.14	-2.77	1.54	-0.29
B	-2.14	4.23	0.89	-2.03	-0.57
C	1.83	0.60	-2.46	0.77	-2.34
D	3.06	0.43	2.43	-0.94	-2.29

5.4 課題と改善

5.4.1 ストループ課題

ここまでのところで、生体情報に影響したと考えられる指標として、正解率の差があげられる。正解率が 9 割を超えた場合、生体情報の変動が少なく、エラーに関わる生体情報が少なくなる。そこで個人差を考慮して、設問の難易度を調整する事で、ヒューマンエラーに関する状態のデータ数を増やすことを考えた。具体的には、次の手順で設定を行った。

1. 各実験協力者に練習課題を行ってもらい、その際の平均回答時間を個人ごとに記録する。
2. 記録した個人ごとの平均回答時間を少数第 2 位以下で切り上げる。
3. 2 で算出した各回答時間を、各実験協力者の課題の出題時間とする。

これにより、個人ごとのヒューマンエラーの状態を適切に評価でき、予測精度を高める事もできると考えた。

また、ストループ課題の中で、色を回答する課題においては、学習者が、文字を読まずに色のみを回答するようなノウハウを得てしまい、エラーが発生しない問題が生じた。そのため問題数に比重を設ける事とした。具体的には、色を回答する課題を 1 分間、色の意味を回答する課題を 5 分間とし、これらの結果を用いるものとした。

6. 設計・実装

6.1 ストループ課題の概要

実験システムの改善によりストループ課題の出題時間を、実験協力者ごとに変更した。そのため、評価実験 1 よりも、実験協力者ごとの問題数の差が出ると考えられる。問題数があると、分析時のデータの信頼性に問題が発生するこ

とから、問題数の差を少なくするため、次の問題が出題されるまでの時間を 500ms で固定した。また、実験 1 では、1 回の実験で、1 分間の色の回答課題、色の意味を回答する課題を 3 回ずつ行った。これを評価実験 2 では、前者が 1 分、後者が 5 分になるように変更した。

6.2 システムの概要

エラーの予測モデルを構築し、リアルタイム予測を行うシステムの概要を図 7 に示す。エラーの予測モデルの構築処理は、前処理とモデル作成を行う 2 つのプログラムで構成される。リアルタイム予測を行う処理は、エラーリアルタイム予測プログラムで行う。

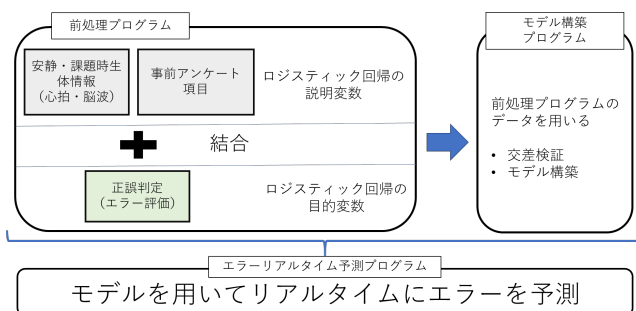


図 7 実験 2 に用いるシステムの概要

Figure 7 Overview of the system used in Experiment 2

6.2.1 ロジスティック回帰

ヒューマンエラーの予測モデルを構築するにあたりロジスティック回帰を用いた。ロジスティック回帰は、非線形回帰分析であり、質的変数を扱うこともできる。今回、説明変数として、生体情報以外にも、事前アンケートを用いたことから、本手法が適していると考えた。本手法は事象であるエラーの起こりやすさを複数の変数で予測する。

6.2.2 データ前処理

本研究では、ミス群前後 10 秒区間において、特徴が有意に得られやすい事を述べた。しかし 2 度目の評価実験においては難易度が上昇し、短い区間でエラーが発生する結果となり、10 秒では、全てエラーになり学習が困難となった。そこで 10 秒を 4 秒に変更し、本区間を対象にミスの判定を行った(以降ミス群前後 4 秒区間)。

次にエラーリアルタイム評価プログラムのためのモデルの構築を行った。言語は python、ライブラリは pandas, numpy, matplotlib, scikit-learn を用いた。処理の内容は、交差検証による学習データの妥当性評価、エラーの予測モデルの構築とした。

前節で説明した前処理に加えて、学習時にもデータの前処理を行う。前処理の内容は、標準化とデータの均衡化である。標準化により、最適な解決法を見つけるまでのステップが少なくなり、分類の精度も向上する。また、均衡化は、データの偏りを軽減するために行った。既に前処理においても、データの偏りを軽減するため、ミス群前後 4 秒区間にしたが、データが偏るため、オーバーサンプリング手法を用い

て、データが均衡になるように前処理を加えた。オーバーサンプリング手法には、SMOTE(Synthetic Minority Over-sampling TEchnique)を用いた。SMOTE は、あるサンプルに対して近傍の K 個のデータから判別を行い、少数派のサンプル数を増やすアルゴリズムである。

6.2.3 エラーリアルタイム予測プログラムの開発

エラーをリアルタイムに予測するプログラムを開発した。このプログラムの処理は、エラー予測モデルの検証としてエラーのリアルタイム予測と評価を行う。エラーの予測モデルの構築、検証には、ロジスティック回帰を用いる。学習に使用するデータは、前処理プログラムから出力されるデータを用いる。更にエラーに関わる事前状態には、ストレス度を加え、他にも体調、安静時の生体情報の平均を加えた。

処理の概要を図 8 に示す。図の左側は、エラーリアルタイム予測プログラム、右側はストロープ課題のプログラムのフローチャートを示す。フローチャートの説明を次に述べる。

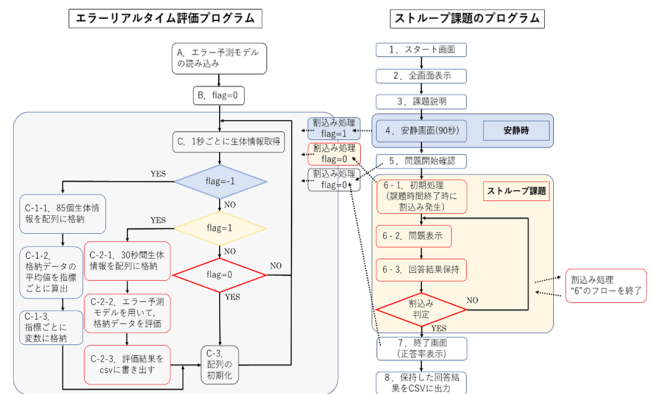


図 8 エラーリアルタイム予測プログラムとストロープ課題のプログラムのフローチャート

Figure 8 Error real-time prediction program and Flowchart of the Stroop assignment program

(処理A) は、構築したモデルを読み込む。

(処理B) は、変数 flag を初期化する。変数 flag は、ストロープ課題プログラムにおいて、安静時の計測は-1、課題時の計測は 1、それ以外での計測は 0 となるよう、割込み処理で得られる値を格納する変数である。

(処理C) は繰り返し行う処理である。

(処理C-1) は、安静画面の際に行われる処理である。この処理の目的は、安静時の生体情報の平均を算出するためである。

(処理C-2) は、安静時の生体情報を配列に格納する。

(処理C-3) は、課題中に行われる処理で、エラーの予測を行う。

(処理C-2-1) にある様に、予測は 30 秒ごとに行われる。

(処理C-2-2) では、モデルを用いて、事前状態と予測前 30 秒間(30 個)の生体情報から、エラーの有無の予測(30 個)

を行う。事前状態は、事前アンケート後、手動でプログラムに入力した。この 30 個のエラー判定において、多く分類された判定をエラーの予測結果とした。

(処理C-2-3) は、予測結果を評価し、結果として出力する。

評価方法は、上記の予測結果と、次の予測までに、実際にエラーが発生したか否かを比較した。

(処理C-3) は、予測用の生体情報を格納する配列を初期化する処理である。

7. 機械学習によるリアルタイム予測

7.1 概要と目的

エラーリアルタイム予測を行い、エラーの予測モデルの検証・評価を行う事を目的とする。そこで、評価実験 2 では次の 3 つを検証した。

- 1) 交差検証によるモデルの評価
- 2) 実験データを用いて、モデルの再評価
- 3) エラーのリアルタイム予測の検証

1)については、開発したプログラムを用いて、3 日間の実験データで構築したモデルの妥当性評価を行う。2)については、4 日目の実験データをテストデータとして、モデルの再評価を行う事で、モデルの妥当性を再確認する。3)については、エラーに分類された割合が半数超えた際、実際にエラーが発生したかを評価する。

実験方法として、脳波計と心拍計を装着しストループ課題を、1 日 1 回ずつ計 4 日間行った。その内、3 日間はモデル構築用のデータに、最後の 1 日を上記検証のための実験として行った。実験協力者は 4 名(20 代成人男性 2 名：実験協力者 A と D, 女性 2 名：実験協力者 B と C)である。また、感染症を防ぐため、各自の自宅で実施していただいた。実験手順は評価実験 1 と同じであるため省略する。

7.2 解析と考察

実験協力者 A はプログラムの不具合により、エラーリアルタイム予測時の、データを取得する事ができなかった。そのため、実験協力者 A は交差検証によるモデルの評価のみを行った。

7.2.1 交差検証によるモデルの評価

エラー予測モデルの妥当性の評価のため、10 人の実験協力者を対象に同様のテストを 3 日間にわたり実施した。実施したテスト時に収集した 3 日間のデータを用いて、学習データ 8 割、テスト用データ 2 割とし交差検証を行った。モデル構築の妥当性の評価には、再現率、適合率、決定係数、精度を用いた。結果を表 6 に示す。

表 6 より、再現率の平均は 74.5%で、最大は 89.3% となった。10 人の実験協力者のうち、7 割を超えた再現率は 7 人で得られたことから、モデルにはある一定の妥当性があるといえる。3 割にあたる H,I,J の 3 名については、80%を超える結果となった。

表 6 使用したモデルの交差検証結果

Figure 6 Cross-validation results for the model used.

実験協力者	再現率	適合率	決定係数	正確率
A	70.0	28.9	0.66	64.0
B	65.6	41.7	0.64	65.3
C	71.4	13.7	0.68	70.4
D	73.1	20.9	0.64	64.6
E	65.8	90.3	0.63	65.0
F	74.1	86.6	0.77	76.3
G	69.5	90.4	0.65	67.0
H	80.8	86.1	0.75	78.5
I	85.4	57.1	0.71	68.2
J	89.3	74.8	0.78	81.0
平均	74.5	59.1	0.69	70.0

また、再現率の結果の分散は小さく、ばらつきが少ないことがわかる。これに対して、適合率の平均は 59.1%で、最大は 90.4%となり、再現率よりも平均は低く、最大は高い結果となった。E,F,G,H は 80%を超える適合率であったのに対して、A,C,D はそれぞれ 28.9%、13.7%、20.9% と 3 割未満と低い値となった。本提案システムは、エラーを防止するための安全に関わるシステムであることから、再現率の高いモデルが望まれる。今回、再現率の平均が比較的高い結果となったこと、また、決定係数も 0.69 と比較的高いことから、システムの有効性を示すことができたといえる。また、各実験協力者の再現率と適合率のばらつきが、A,C,D をのぞき小さいことがわかった。

また、モデルの正確率の平均が 70.0%と比較的良好な結果となった。予測モデルにおいてこの値を達成できていることは評価に値する。

7.2.2 ヒューマンエラーのリアルタイム予測の検証

4 日目に行ったエラーリアルタイム予測の結果の評価をする方法として、30 個の予測データにおいて、エラーに分類された割合が半数超えた際、実際にエラーが発生したかを評価する。ロジスティック回帰で得られたモデルを用いて、事前状態と予測前 30 秒間(30 個)の生体情報から、エラーの有無の予測(30 個)を行う。事前状態に含まれるストレス度と体調の値は、事前アンケートを行った後、プログラムに入力した。この 30 個のエラー判定において、多く分類された判定をエラーの予測結果とした。割合の算出には、次に示す式の値を用いた。

$$\frac{\text{エラー(不正解)に予測された個数}}{\text{正解に予測された個数} + \text{エラー(不正解)に予測された個数}}$$

予測結果の評価方法は、上記の予測結果と、次の予測までに、

実際にエラーが発生したか否かを比較した。例えば、予測結果がエラーと判定された場合、次の予測までの 30 秒間に、実際にエラーが発生したかを評価する。

結果を表 7 に示す。表 7 は 3 名 (10 名では) 全員の予測結果を合わせた 80 個(実験協力者 A と E 以外は 10 個ずつ)のエラーリアルタイム予測の結果を表す。エラーが発生すると予測された 79 個のうち、59 個において実際にエラーが発生した。すなわち、適合率は 74%であった。また、実際にエラーが発生した 59 個全てについて、エラーが発生すると予測された。ただ、1 個のエラーが発生しないと予測されたが実際にエラーが発生した。このことから、再現率は 100%であった。つまり、エラーのリアルタイム予測の有効性が示されたといえる。

表 7 リアルタイムエラー予測結果の評価

Table7 Evaluation of real-time error prediction results

	エラーの発生予測	エラー発生なし予測
エラーが発生	59	1
エラー発生なし	20	0

8. まとめ

本研究では、提案手法によりモデルを構築する事ができた。モデルにはエラーの発生に関係する脳波・心拍とアンケート結果の一部指標を用いる事で構築を行った。また、構築したモデルを用いて、リアルタイムエラー予測の有効性を示す事ができた。

今後の課題として次の 2 つが考えられる。エラー予測モデルの精度向上、リアルタイムエラーの予測の更なる有効性の調査がある。前者については、新たな事前状態の指標や、データ数を増やす必要がある。具体的には、環境的要因を考慮した指標が必要であると考え。また、時間的要因に左右されないためにも、データ数を増やす事が重要であると考え。さらに、ロジスティック回帰以外の機械学習手法も検証する必要がある。理由は、環境的、時間的要因を考慮した際、説明変数間の相互作用や重要性の考察が必要だからである。そのため、説明変数間の相互作用を考慮した分類が可能であり、変数の重要度を考察できるランダムフォレストなど、他の機械学習手法を比較する事も検討する。

後者については、今回の実験で、エラーが発生しないと予測されたケースと、実際にエラーが発生しなかったケースにおいて、有効性を示す評価は出来ていない。この要因として、課題の難易度を上げた事が考えられる。具体的な改善策としては、30 秒間に 1 回エラーするか否かの難易度が望ましいと考える。さらに、体調によって作業パフォーマンスが異なった。このことから、課題の難易度にも機械学習を用いて、適切な難易度を調節できると良いと考えられる。具体的には、事前状態から回答時間を設定し、課題の回答状況から、リアルタイムに回答時間を調節する。

参考文献

- [1] 厚生労働省. “職場のあんぜんサイト：労働災害原因要素の分析 (平成 28 年製造業)”
<https://anzeninfo.mhlw.go.jp/user/anzen/tok/link/tok1-701-3.xlsx>.
- [2] Mark St. John, David A. Kobus, Jeffrey G. Morrison & Dylan Schmorow, "Overview of the DARPA Augmented Cognition Technical Integration Experiment", International Journal of Human-Computer Interaction, Volume 17, pp. 131-149. 2004
- [3] 吉田怜司, “適合的指標選択によるヒューマンエラー予測手法の検討”, 芝浦工業大学修士論文 2020 年度, 2020.
- [4] Taro Kishimoto, Reiji Yoshida, Midori Sugaya, Examination of Evaluation Method During Work by Bioinstrumentation, Second International Workshop on Affective Computing for Requirements Engineering, Sep. 23, 2019
- [5] 加藤淳, “看護師のヒューマンエラーによる医療事故に関する事例研究”, 人間生活文化研究, No. 28, pp.595-609, 2018
- [6] Frank H Hawkins, “Human Factors in Flight,”2019.
- [7] 河野龍太郎, “医療におけるヒューマンエラー”, 医学書院, 2004.
- [8] 河野龍太郎, “医療安全へのヒューマンファクターズアプローチ”, 日本規格協会, 2010.
- [9] 西川雅弥, 西原直枝, 田辺新一, “中程度の高温環境下の長時間作業が作業効率と疲労に与える影響に関する被検者実験”, 日本建築学会環境系論文集, vol.74, no.638, pp.525-530, 2009.
- [10] Soraia M. Alarcão, "Reminiscence therapy improvement using emotional information", Affective Computing and Intelligent Interaction (ACII) 2017 Seventh International Conference on, pp. 561-565, 2017.
- [11] Shaffer F, Ginsberg JP. An Overview of Heart Rate Variability Metrics and Norms. Front Public Health. 2017;5:258. Published 2017 Sep 28. doi:10.3389/fpubh.2017.00258
- [12] National Aeronautics and Space Administration : Nasa TLX, <https://humansystems.arc.nasa.gov/groups/tlx/>, 2020.
- [13] Yuhei Ikeda, Ryota Horie, Midori Sugaya, Estimate Emotion with Biological Information for Robot Interaction, 21st International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES-2017), Marseille, France, 6-8, Sep, Procedia Computer Science, Vol.112, pp.1589-1600, Year 2017.
- [14] Chee-Keong Alfred, Wai Chong Chia and Siew Wen Chin, “A Mobile Driver Safety System: Analysis of Single-Channel EEG on Drowsiness Detection”, International Conference on Computational Science and Technology (ICCST).2014.
- [15] 平井章康, 吉田幸二, 宮地功, “簡易脳波計による学習時の施行と記憶の比較分析”, マルチメディア, 分散協調とモバイル(DICOMO2013)シンポジウム, 2013.
- [16] 安全衛生情報センター. “労働者の疲労蓄積度自己診断チェックリスト”
http://www.jaish.gr.jp/td_chk/pdf/chk_list1.pdf
- [17] 厚生労働省. “ストレスチェック等の職場におけるメンタルヘルス対策・過重労働対策等”
https://www.mhlw.go.jp/bunya/roudoukijun/anzeneisei12/dl/stress-check_j.pdf.
- [18] 重森雅嘉. “ヒューマンエラー尺度としてのストループ課題”, 日心第 70 回大会. 2006.