EEG-based performance estimation during a realistic drone piloting task

Marcel Francis Hinss, Vincenzo Maria Vitale, Anke Brock, Raphaelle Roy

EEG-BASED PERFORMANCE ESTIMATION DURING A REALISTIC DRONE PILOTING TASK

Marcel F. Hinss [1], Vincenzo Maria Vitale [1], Anke M. Brock [1], Raphaëlle N. Roy[1]

[1] Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, France.

E-mail: marcel.hinss@isae-supaero.fr

ABSTRACT

Passive brain-computer interfaces (pBCIs) developed within the neuroergonomic field usually aim to improve safety by augmenting human-machine interaction. To accomplish said goal, many pBCIs classify mental states such as mental workload or mental fatigue. An alternative is to forego mental states and aim to predict performance. Despite its drawbacks, we argue that performance estimation is a more goal-oriented approach than mental state estimation. In a realistic experiment, 25 participants had to control an uncrewed aerial system for two hours, continuously switching between target search and navigation. EEG classification accuracies based on mental states and performance were compared. With a Tangent Space Logistic Regression, we could predict an increased likelihood of lapses in the form of missing instructions with an above-chance level accuracy of 62.09 %.

INTRODUCTION

Passive Brain-Computer Interfaces (pBCI), i.e., BCIs that observe brain activity that is not influenced by the presence of a BCI, are a valuable component of neuroergonomics [1]. They promise to provide complex systems, such as cockpits, with valuable information on their user and the Human-Machine Interaction. A machine can then use that information employing adaptation or feedback to improve said interaction [2].

To do so, pBCIs are often trained to detect specific mental states such as mental workload [3] or mental fatigue [4]. The underlying argument for detecting said mental states is their correlation with erroneous or sub-optimal behaviour by the operator. Thus, by detecting, e.g. a high mental workload, the system may adapt itself to reduce workload and increase safety [5]. This approach has certain drawbacks. Mental states as constructs are not observable and vary across definitions [6, 7]. Furthermore, mental states depend on the current context and tasks. Differing task instructions can result in differing brain activity [8]. A high mental workload during an N-Back task may not be comparable to a high workload during a Stroop task [9, 10]. Finally, mental states are not always strong predictors of performance. In the case of mental fatigue, evidence suggests that participants, using e.g. compensatory strategies, can uphold performance despite fatigue [11–13]. So, all these aspects must be accounted for when constructing a mental state-based pBCI to be used in an open-loop adaptation (feedback) or a closed-loop adaptation (interface change) with a complex system [2].

Alternatively, a pBCI could forego the mental state aspect and try to predict a participant's behaviour directly. Performance estimation has been proven to work in several contexts [14–17] and does not suffer from any of the aforementioned issues. It works by assigning labels to the physiological data, using the recorded behavioural data of participants, such as reaction time, accuracies and misses. Performance prediction allows direct observation of the variable we want to optimize with a pBCI, but also faces challenges. To predict performance, we first need to define good and bad performance. Many tasks, such as the Stroop or N-back tasks, involve some measure of correctness and reaction time [9, 10]. Reaction time or accuracy may be considered a valid performance measure in these cases, but only the combination into a global score will provide a complete picture of performance. Combining scores, on the other hand, raises questions about how to weigh each metric. Here, it needs to be considered that these measures are not orthogonal [8]. As mentioned above, the issue gets more complex as we move away from very controlled tasks and move towards more ecologically valid measures that may include several different reaction times and accuracies.

A related challenge is then how these cases should be labelled. Imagine, for example, an experiment where participants continuously perform a task for one hour. A global score of performance is assigned to each minute of the task. The value is continuous from 0 (bad) to 1 (good). How can the data now be divided into a 2-class problem? The 10 worst minutes of performance versus the remaining 50 minutes, the 10 best minutes versus the 10 worst minutes, the good half versus the bad half, or values exceeding a threshold (e.g. >0.8) or subceeding another (e.g.<0.2) are all present plausible approaches. The metric calculation and label assignment issue is further complicated when algorithms are tested to classify the data. Does a chance level classification accuracy mean the algorithm doesn't work or that the label assignment is sub-optimal?

In many cases, predicting any change in performance may be helpful, whether it is the likelihood of committing

an error, missing a trial or the speed at which a participant responds. Still, trying out unlimited combinations of labels may also create a global performance score not because of its usefulness but because of its ability to be classified.

To the best of our knowledge, no study has yet evaluated performance prediction using a pBCI applied to a prolonged realistic drone task. Hence, this experiment tested whether an EEG-based pBCI can predict meaningful performance metrics from participants performing a complex Uncrewed Aerial System (UAS, drone) piloting task. Moreover, extended mission duration makes UAS pilots vulnerable to mental fatigue and the associated risks [18]. Participants were asked to switch between a search and a navigation task during the experiment for two hours. This work expands on a previous protocol that focused only on a visual search task –and which did not yield above chance level performance estimations– [19] by adding a second navigation task and making the overall performance more difficult and longer. The long duration allows for comparing performance-based labelling and more traditional labelling based on subjective fatigue scores and Time-On-Task (TOT). Our goal is to illustrate how label assignment impacts classification accuracies, particularly in the absence of an absolute performance definition, for such a realistic task.

METHODS

Participants:

25 Participants (7 female, mean age 23.54 years (std 2.7), 11 English speaking & 14 French speaking) were recruited and completed the experiment. From the subsequent analysis, one participant had to be dropped due to inconsistencies and missing data in the recordings.

Procedure: Participants who agreed to participate signed the informed consent forms and were equipped with the EEG sensor.Next, participants completed a battery of questionnaires. They then completed a training phase of 16 minutes before a five-minute resting state was recorded (30-second intervals of eyes open and closed). They then started the 120-minute main phase of the experiment. After completing the main phase, participants filled in another battery of RSME, KSS, SPS, and VAS questionnaires. The ethical committee of Toulouse (Comité éthique de l'Université de Toulouse) approved the experimental protocol (Dossier 2022-501).

Materials:

Task: The UASOS task (Fig. 1) combines some of the fundamental aspects of UAS operations with a Task-Switching protocol to allow the investigation of cognitive flexibility [20]. The task requires participants to alternate between tasks on a trail-based system. On average, every 7 seconds (with a +-1000ms jitter, ~1020 trials during the main condition), written instructions appear on a widget in the middle of the participant's visual field to indicate the current task. To ensure adequate performance, small pretests were conducted to calibrate these parameters. Participants work on two main tasks, with two modes each. The Navigation task (NAV) requires the participant to navigate the UAV either using headings (heading mode, HDG) or waypoints (Waypoint mode, WPY). The design was balanced with an equivalent number of trials in all tasks and modes.

Figure 1: Experimental setup. Left Screen: Search Task. Right Screen: Navigation Task. Center Top: Flight director with task information. (Note: The text in the centre of the screens is feedback only displayed during the training phase.)

In the heading mode of the NAV task, participants receive a heading instruction (e.g. 350) in each trial. Using a joystick, they then turn the UAS in said direction. For the WPY mode, they receive a waypoint consisting of a letter and a number (e.g. F13). They must choose the corresponding waypoint using a trackball on a grid overlaying the navigational display. The other task is the SRC task. This task was adapted from previous work [19], and integrated into the overall protocol. Participants see a 3x3 grid of black-and-white images that visual filters may further distort. They are instructed to search either People or Vehicles. If they detect a target on one picture, they select the corresponding picture using a numpad. For all tasks, reaction times and the correctness of responses are recorded. The instructions on the flight director widget tell the participant which mode to perform at the onset of each trial.

The task was coded in Python and presented on two identical computer screens. A detailed description of the experimental environment can be found in [21].

Questionnaires: Participants answered 5 questionnaires at varying moments. At the beginning of the experiment, participants completed the demographics questionnaire, and their handedness was also assessed using the shortened Edinburgh handedness questionnaire [22]. Next, the Karolinska Sleepiness Scale (KSS), a 9-point Likert and the Samn-Perelli Fatigue (SPF) 7-point Likert scale were used to assess fatigue [23, 24]. Participants also filled in the RSME scale [25] that evaluates participants' mental effort invested in the task. The versions in which all items are labelled were used [26]. The translated versions of the KSS and SPS questionnaires originate from the ICAO [27]. Participants also responded to two Visual Analogue Scales (VAS) scales: cognitive fatigue (VAS-

F) and drowsiness (VAS-D). The entire battery was presented a second time following the completion of the experiment. The VAS scales were also filled in at 19-minute intervals during the main experimental phase.

EEG: Using an active AG-AgCl electrode system with an ActiCHamp amplifier (Brain Products, Gmbh), EEG data was recorded from 64 electrodes. The international 10-20 system was used for electrode placement [28]. Data were recorded at 500 Hz, and impedances were kept below 50 kΩ. Data was streamed and synchronized using the LabStreamingLayer (LSL) [29].

As part of the data validation, we performed a frequency analysis of the EEG data independently from the mental state prediction. For this, a zero-padded channel was added to the EEG data before an average referencing, with a subsequent removal of the zero-padded channel. Extreme values were clamped following the method proposed by [30]. The data was then cut into 5-second non-overlapping epochs. The power of each frequency band was calculated by band-pass filtering the signal and calculating the root mean square for each electrode. Using the parameters suggested in [31], the power of the theta (4-8Hz) alpha (8-12Hz) bands were extracted. For the statistical analysis the data was then averaged into 10-minute epochs over three clusters of electrodes (i) Frontal: F3, F1, Fz, F2, F4, FC3, FC1, FCz, FC2, FC4; (ii) Central: C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, CP4; and (iii) Parieto-Occipital: P3, P1, Pz, P2, P4, PO3, POz, PO4, O1, Oz, O2.

Performance and Mental State classification:

EEG preprocessing: The EEG data for each participant was cut into 5-second non-overlapping epochs, to allow for robust covariance matrix estimation, and to be independent of the task trials. The epoched data was then referenced and filtered between 2-36 Hz using the *mne.filter()* function. Data points of each channel that exceeded 20 std of the robustly scaled data were clamped to the value equal to +/- 20 std; for a detailed explanation of this method, see [30]. Finally, the data was resampled to 125Hz.

Label creation: The performance metrics used misses (Miss), reaction times (RT) and accuracies (Acc) of all subtasks. Across subtasks, all values were first normalized to give equal importance to each subtask. Next, averages of misses, RT and accuracies were calculated. For each value, the best and worst 33% were used to assign labels—the global performance (OVR) score combined misses, RT, and accuracy. Three different mental state-based labels were created. The time-on-task used the first and last 33% of each recording, respectively. The VAS scores were used for the other two approaches. Using the drowsiness and cognitive fatigue scales, the blocks corresponding to the most extreme values of each scale were used to label the data. Adjusted chance levels were calculated for each label-type based on [32].

Classification: The data from each participant was divided into an 80/20 split for training/testing datasets. Next, the covariances were computed using OAS or LWF, and the data was projected to the target space. We compared performances of logistic regression (Log Reg), Support Vector Machine (SVM) and Random Forest (RF) Classifiers. Hyper-parameters for each classifier were optimized using 5-fold cross-validation using Bayesian search.

Statistical Analysis:

The general inference criterion is a p-value of $p < .05$. In multiple comparisons, we adjusted that criterion according to the adjusted Bonferroni method. Assumptions for each statistical test were checked and accounted for if not satisfied. Outlier detection was performed based on the interquartile range criterion. This was done for trials grouped by condition.

Subjective: To analyse the subjective results, we compared the SPS and KSS scores from the beginning to the end. For this, we used a paired samples T-Test. We also performed a one-way repeated measures ANOVA for both VAS scales. For one participant, the questionnaires at the end of the experiment were not recorded.

Behavioral: The behavioural analysis was divided into three sections for the (i) Search Task, (ii) Navigation Task - Heading Mode, and (iii) Navigation Task - Waypoint mode, respectively. An overall analysis was not possible due to the differences between tasks. Due to the randomized order of the tasks, missing data occurred in some blocks as single participants did not engage in a task in a given block. In this case, the missing values were replaced with the list-wise mean. This occurred in 0.55 % of the behavioural data.

For the search task, reaction times, F1 score and misses were used as dependent variables in repeated measures within-subjects ANOVA with Task (searching humans or searching vehicles) as an independent variable and time on task (19-minute blocks) for repeated measures.

To analyse the heading task reaction times, turning direction misses and deviation were the dependent variables of repeated measures within-subjects ANOVA with TOT (19-minute blocks) for repeated measures.

The Waypoint mode was evaluated using reaction time, correct choices, and misses as dependent variables, again TOT was the independent variable for repeated measures.

EEG Frequency: The extracted powerbands were compared across blocks in a repeated measures ANOVA for each cluster.

Classification: To analyze the classification results, a 2-way ANOVA with factors Classifier and Label-type was performed on the dependent variable of accuracy.

RESULTS

Subjective Data: Scores for the KSS, RSME and SPS measures all showed significant increases in values comparing the beginning and the end of the experiment KSS: $t(23) = -6.912$, $p < .001, d = -1.411$; SPS: $Z = -4.000$, $p < .001, r = -0.933$; RSME: $t(23) = -6.380$, $p < .001, d = -1.302$). The assumption of normality was violated for the SPS test (Shapiro-Wilk

$W = 0.875$, $p = 0.007$); therefore, the Wilcoxon result is reported. VAS scores on both cognitive fatigue and drowsiness showed linear increases over time (Cognitive: $F(2.374, 49.849) = 25.979$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.553$ and $\eta_p^2 = 0.553$; Drowsiness: $F(4.294, 90.184) = 12.159$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.367$ and $\eta_p^2 = 0.367$, see Figure 2 a-d).

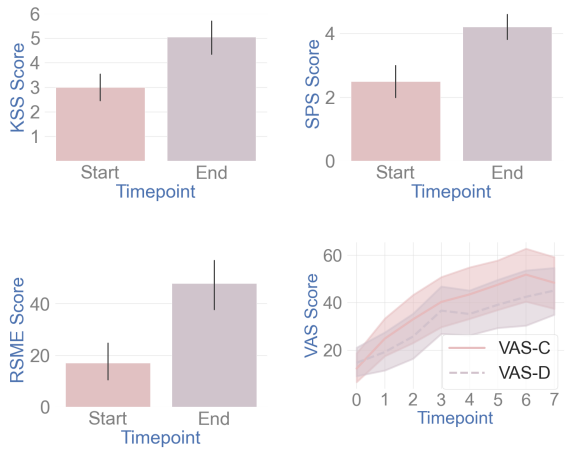

Figure 2: Subjective results: a) KSS Scores comparing the beginning and end of the experiment. b) SPS Scores comparing the beginning and end of the experiment. c) RSME Scores comparing the beginning and end of the experiment. d) VAS scores throughout the experiment for both the Cognitive Fatigue and the Drowsiness scale. Timepoint 0 is before the start and before the training, and then each subsequent point occurred every 19 minutes into the experiment. The last point occurred after the completion of the experimental phase.

Behavioral Data:

SRC Task: Searching humans resulted in significantly larger reaction times and more misses (RT: $F(1,24) = 214,872$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.900$ Misses: $F(1,24) = 34.308$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = .588$). Surprisingly, the F1 score was slightly higher for searching humans (F1: $F(1,24) = 108.011$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = .818$) . Time did not significantly affect performance on any metric (RT: $F(4.117,120) = .381$, $p = .827$, Greenhouse-Geisser corrected, $\eta^2 = .0160$ F1:$F(4.056,120) = 2.268$, $p < 0.066$, Greenhouse-Geisser corrected, $\eta^2 = 0.086$ Miss:$F(2.024,) = .751$, $p = .479$, Greenhouse-Geisser corrected, $\eta^2 = .030$.), see Figure 3 a-d.

NAV task HDG mode: RT, correct turn and misses were all influenced by time (RT: $F(4.1,60) = 41.711$, $p < .005$, Greenhouse-Geisser corrected, $\eta^2 = 0.635$, correct turn: $F(4.207,60) = 3356$, $p = .011$, Greenhouse-Geisser corrected, $\eta^2 = .123$, miss: $F(3.327,60) = 4.185$, $p = .007$, Greenhouse-Geisser corrected, $\eta^2 = .148$). Contrast analysis revealed significant cubic effects for RT, correct turn and misses (RT: $F(5,60) = 99.569$, $p < .005$, Greenhouse-Geisser corrected, $\eta^2 = 0.806$,

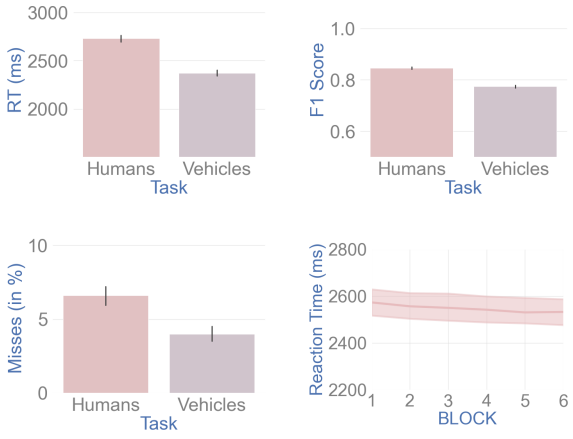

Figure 3: Behavioral Results of the SRC task: a) Reaction time by Mode. b) F1 scores by Mode. c) Misses by mode. d) Reaction Times over time

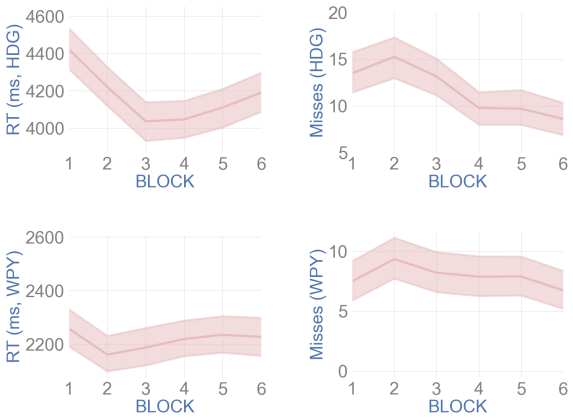

Figure 4: Behavioral Results of the NAV task: a) HDG: Reaction times over time b) HDG: Misses over time. c) WPY: Reaction times over time. d) WPY: Misses over time.

correct turn: $F(5,60) = 7.627$, $p = .011$, Greenhouse-Geisser corrected, $\eta^2 = .231$, miss: $F(5,60) = 5.253$, $p = .031$, Greenhouse-Geisser corrected, $\eta^2 = .180$). See Figure 4 a & b.

SRC task WPY mode: Reaction times were also influenced by time (RT: $F(3.078,60) = 2.707$, $p = .05$, Greenhouse-Geisser corrected, $\eta^2 = 0.101$). Contrast tests revealed a significant cubic effect (RT: $F(1,60) = 10.395$, $p = .004$, Greenhouse-Geisser corrected, $\eta^2 = 0.302$). See Figure 4 c & d.

EEG: The alpha frequency band showed significant effects throughout the experiment. In all three clusters, an increase in alpha power was observed (Frontal : $(F(1,4.6) = 3.184$, $p = .012$, Greenhouse-Geisser corrected, $\eta^2 = 0.117$; Central: $(F(1,6.051) = 3.995$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.143$; Parieto-Occipital: $(F(1,4.899) = 2.508$, $p = .035$, Greenhouse-Geisser corrected, $\eta^2 = 0.095$). Theta power did not show any significant change over time (see Figure 5 a-d.

Classification: The 2-way ANOVA showed a signifi-

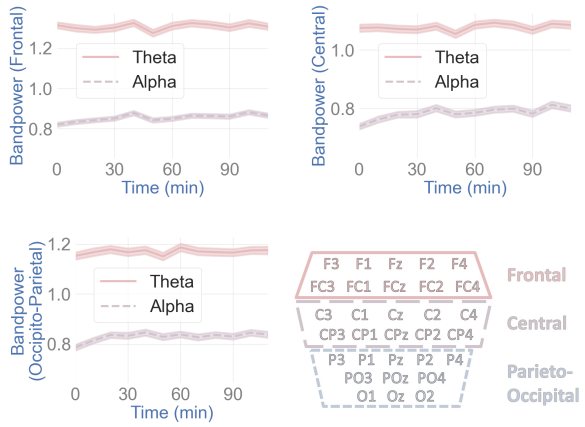

Figure 5: Frequency power in the theta and alpha power bands over time for the a) Frontal Cluster; b) Central Cluster; c) Parieto-Occipital Cluster. d) Topology of the 3 clusters.

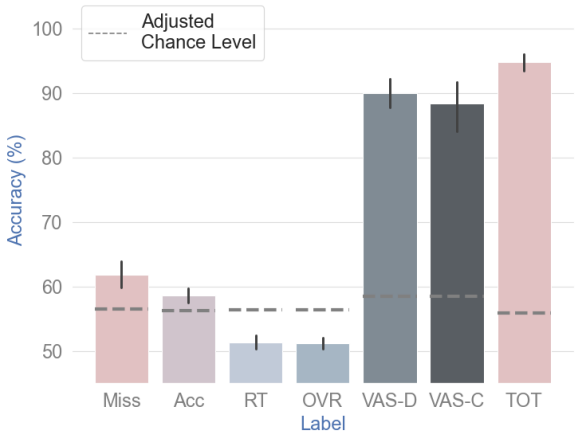

Figure 6: Classification results: Accuracy by Label Type with adjusted chance levels.

cant effect of Label type ($F(6,24) = 342.12$, $p < .001$, $\eta^2 = 0.81$). There was no significant effect of classifier ($F(7,24) = 2.398$, $p > .09$, $\eta^2 = 0.01$). The highest classification accuracy was obtained using the TOT labels, with an average accuracy of 94.86% across all classifiers. However, both VAS scales, the performance labelling based on misses and accuracy, also performed above their respective chance levels. This resulted in a 62.09 % accuracy for detecting misses using the tangent space logistic regression (see Figure 6).

DISCUSSION

To use pBCIs in complex environments, the output of a pBCI needs to have some predictive value. The results presented here compare mental-state estimation and performance estimation using EEG data. The subjective and EEG analyses both point toward increased mental fatigue over time. Yet, while some behavioural metrics, such as reaction times in the navigation task, seem to show a similar trend, variability in overall performance is not best explained by TOT.

While mental state estimation using Time-On-Task or subjective metrics as ground truth performs considerably above chance level, our algorithms could also predict misses with an above chance level likelihood. The algorithm's success with TOT metrics may be attributable to the observed alpha power increase often associated with mental fatigue. [31, 33]. It may also be due to slow drifts and the non-stationarity of the EEG signal [34]. Subjective fatigue scores increased over time, creating similarities between the TOT and VAS labels. The absence of stronger effects in the spectral analysis may have been attenuated due to the complexity of the task [35]. The reaction time-based performance estimation had the lowest accuracy. One possible explanation is that longer reaction times may reflect several processes that are then mixed up. Slow reaction times may be due to fatigue [36] or a speed-accuracy tradeoff [37]. The moderate success of the performance estimation based on misses suggests that spectral EEG features, especially theta power, are sensitive to lapses [38]. Future work could evaluate incorporating Bayesian updating, which may further improve the performance estimation

CONCLUSION

This study highlights the challenges and possibilities of EEG-based performance estimation. The differences between definitions of performance highlight the importance of label assignment. In our opinion, performance scores should (i) be defined as a priori, (ii) be explainable, and (iii) provide real-world value.

REFERENCES

[1] Lotte F, Roy RN. Brain–computer interface contributions to neuroergonomics. In: Neuroergonomics. Elsevier, 2019, 43–48.

[2] Krol, LR., Zander, TO. Passive Bci-Based Neuroadaptive Systems. 2017.

[3] Aricò P, Borghini G, Di Flumeri G, Sciaraffa N, Colosimo A, Babiloni F. Passive BCI in Operational Environments: Insights, Recent Advances, and Future Trends. IEEE Transactions on Biomedical Engineering. 2017;64(7):1431–1436.

[4] Trejo LJ, Kubitz K, Rosipal R, Kochavi RL, Montgomery LD. EEG-Based Estimation and Classification of Mental Fatigue. Psychology. 2015;06(05):572–589.

[5] Zander TO *et al.* Automated Task Load Detection with Electroencephalography: Towards Passive Brain–Computer Interfacing in Robotic Surgery. Journal of Medical Robotics Research. 2017;02(01):1750003.

[6] Young MS, Brookhuis KA, Wickens CD, Hancock PA. State of science: Mental workload in ergonomics. Ergonomics. 2015;58(1):1–17.

[7] Phillips RO. A review of definitions of fatigue – And a step towards a whole definition. Transportation Research Part F: Traffic Psychology and Behaviour. 2015;29:48–56.

[8] Heitz RP. The speed-accuracy tradeoff: History, physiology, methodology, and behavior. Frontiers in Neuroscience. 2014;8.
[9] Kirchner WK. Age differences in short-term retention of rapidly changing information. Journal of experimental psychology. 1958;55(4):352.
[10] Stroop JR. Studies of interference in serial verbal reactions. Journal of experimental psychology. 1935;18(6):643.
[11] Hamann A, Carstengerdes N. Assessing the development of mental fatigue during simulated flights with concurrent EEG-fNIRS measurement. Scientific Reports. 2023;13(1):4738.
[12] Linden D van der. The urge to stop: The cognitive and biological nature of acute mental fatigue. In: Cognitive fatigue: Multidisciplinary perspectives on current research and future applications. APA: Washington, DC, US, 2011, 149–164.
[13] Robert J. Hockey G. Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. Biological Psychology. 1997;45(1):73–93.
[14] Müller KR, Tangermann M, Dornhege G, Krauledat M, Curio G, Blankertz B. Machine learning for real-time single-trial EEG-analysis: From brain–computer interfacing to mental state monitoring. Journal of Neuroscience Methods. 2008;167(1):82–90.
[15] Ko LW *et al.* Single channel wireless EEG device for real-time fatigue level detection. In: IJCNN. Jul. 2015.
[16] Blaha LM, Fisher CR, Walsh MM, Veksler BZ, Gunzelmann G. Real-Time Fatigue Monitoring with Computational Cognitive Models. In: Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience. Springer International Publishing: Cham, 2016, 299–310.
[17] Alhazmi S, Saini MK, El Saddik A. Multimedia Fatigue Detection for Adaptive Infotainment User Interface. In: Human-Computer-Media Communication. ACM: New York, NY, USA, Oct. 2015, 15–24.
[18] Hinss MF, Brock AM, Roy RN. Cognitive effects of prolonged continuous human-machine interaction: The case for mental state-based adaptive interfaces. Frontiers in Neuroergonomics. 2022;3.
[19] Hinss MF, Jahanpour ES, Brock AM, Roy RN. Labeling mental fatigue for passive BCI applications: Accuracy vs applicability tradeoff. Bruxelles, 2023.
[20] Hinss MF, Brock AM, Roy RN. The double task-switching protocol: An investigation into the effects of similarity and conflict on cognitive flexibility in the context of mental fatigue. PLOS ONE. 2023;18(2):e0279021.
[21] Hinss MF, Vitale VM, Phan NT, Roy RN, Brock AM. UASOS: An Experimental Environment For Assessing Mental Fatigue & Cognitive Flexibility During Drone Operations. ACM: Boulder Colorado, USA, 2024.
[22] Veale JF. Edinburgh Handedness Inventory - Short Form: A revised version based on confirmatory factor analysis. Laterality. 2014;19(2):164–177.
[23] Samn SW, Perelli LP. "Estimating Aircrew Fatigue: A Technique with Application to Airlift Operations." Brooks AFB, TX: USAF School of Aerospace Medicine. Technical Report SAM-TR-82-21. 1982.
[24] Åkerstedt T, Gillberg M. Subjective and Objective Sleepiness in the Active Individual. International Journal of Neuroscience. 1990;52(1-2):29–37.
[25] Ghanbary A, Ashnagar M, Habibi E, Sadeghi S. Evaluation of Rating Scale Mental Effort (RSME) effectiveness for mental workload assessment in nurses. Journal of Occupational Health and Epidemiology. 2016;5(4):211–217.
[26] Miley A, Kecklund G, Åkerstedt T. Comparing two versions of the Karolinska Sleepiness Scale (KSS). Sleep and Biological Rhythms. 2016;14(3):257–260.
[27] ICAO. "Manual for the Oversight of fatigue Management Approaches." Tech. Rep. Doc 9966. 2016.
[28] Jasper H. The ten-twenty electrode system of the international federation \textbar CiNii Research. Electroencephalogr. Clin. Neurophysio. 1958:10, 370–375.
[29] Kothe C. Sccn/labstreaminglayer. original-date: 2018-02-28T10:50:12Z. 2015. (visited on 01/11/2023).
[30] Défossez A, Caucheteux C, Rapin J, Kabeli O, King JR. Decoding speech perception from non-invasive brain recordings. Nature Machine Intelligence. 2023;5(10):1097–1107.
[31] Wascher E *et al.* Frontal theta activity reflects distinct aspects of mental fatigue. Biological Psychology. 2014;96:57–65.
[32] Mueller-Putz G, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random: A closer look on BCI results. International Journal of Bioelectromagnetism. 2008.
[33] Boksem MA, Meijman TF, Lorist MM. Effects of mental fatigue on attention: An ERP study. Cognitive Brain Research. 2005;25(1):107–116.
[34] Urigüen JA, Garcia-Zapirain B. EEG artifact removal—state-of-the-art and guidelines. Journal of Neural Engineering. 2015;12(3):031001.
[35] Kamzanova AT, Matthews G, Kustubayeva AM, Jakupov SM. EEG Indices to Time-On-Task Effects and to a Workload Manipulation (Cueing). International Journal of Psychological and Behavioral Sciences. 2011;5(8):928–931.
[36] Csathó, Linden D vd, Hernádi I, Buzás P, Kalmár G. Effects of mental fatigue on the capacity limits of visual attention. Journal of Cognitive Psychology. 2012;24(5):511–524.
[37] Wood CC, Jennings JR. Speed-accuracy tradeoff functions in choice reaction time: Experimental designs and computational procedures. Perception & Psychophysics. 1976;19(1):92–102.
[38] Peiris MTR, Jones RD, Davidson PR, Carroll GJ, Bones PJ. Frequent lapses of responsiveness during an extended visuomotor tracking task in non-sleep-deprived subjects. Journal of Sleep Research. 2006;15(3):291–300.