How Does Cognitive Fatigue and Mental Workload Influence Alarm Detection in Flight Simulator? Classification of Electrophysiological Signatures with Explainable IA

Massé Eva, Bartheye Olivier and Fabre Ludovic
Centre de Recherche de l'École de l'Air (CREA), Ecole de l'Air et de l'Espace, F-13661, Salon-de-Provence, France

Keywords: Single-Trial Classification, pBCI, Inattentional Deafness, Brain Activity, ERP (Event Related Potentials), Explainable AI.

Abstract: Relevant sounds such as alarms are sometimes involuntarily ignored, a phenomenon called inattentional deafness. This phenomenon occurs under specific conditions including high workload (i.e., multi-tasking) and/or cognitive fatigue. In the context of aviation, such an error can have drastic consequences on flight safety. The present study used an Oddball paradigm in which participants had to detect rare sounds in an ecological context of simulated flight. Cognitive fatigue and cognitive load were manipulated to trigger inattentional deafness, and brain activity was recorded via EEG. Our results showed that alarm omission and alarm detection can be classified based on a time-frequency analysis of brain activity. We reached a maximum accuracy of 76.4% when the algorithm was trained on all participants and a maximum of 90.5%, on one participant, when the algorithm was trained individually. This method can benefit from explainable artificial intelligence to develop efficient and understandable passive Brain-Computer Interfaces, to improve flight safety by detecting such attentional failures in real-time and giving appropriate feedback to pilots, according to our ambitious goal: providing them reliable and rich human/machine interactions.

1 INTRODUCTION

Increased operational capabilities of aircraft had considerably modified the pilots' missions. These changes concern an increase in the time spent onboard and the complexity of technologies or operations, particularly in the military domain. These long periods of intense and sustained cognitive activities induce cognitive fatigue that is one of the major risks of incidents/accidents in aviation (e.g., Dönmez & Uslu, 2018; Marcus & Rosekind, 2017).

Cognitive fatigue has been shown to occur when the costs of cognitive effort to perform the activity are higher than the expected benefits (e.g., Boksem & Tops, 2008; Kurzban et al., 2013). In this case, after performing an effortful task, disengagement from the current task or unwillingness to sustain the effort on a second task is likely (Inzlicht et al., 2014; Müller & Apps, 2019). Previous studies found that cognitive fatigue can impair cognitive performance, leading to impaired ability to suppress irrelevant information (selective attention found by Faber et al., 2012), alter the automatic motor response (online action control found by Salomone et al., 2021), and more generally disrupt attentional processes (Boksem et al., 2005).

The influence of cognitive fatigue on electrophysiological activities has been also reported as, for example, an increase of the spectral power of δ, θ, and α frequency bands but a decrease of the spectral power of β band, as well as a decreased amplitude of ERP components such as P300, N100 and N200b (e.g., Boksem et al., 2005; Barwick et al., 2012; Borghini et al., 2012; Zhao et al., 2012; Wascher et al., 2014 ; Sabeti et al., 2017; Schmidt et al., 2009). However, these findings are not always replicated, and some authors do not report any impairment of performance with cognitive fatigue (e.g., Ackerman & Kanfer, 2009; Boksem et al., 2005; Lorist et al., 2005; Möckel et al., 2015; Trejo et al., 2007). Unknown is whether the decreased performance or electrophysiological changes associated with cognitive fatigue are caused by a progressive deterioration of the cognitive resources or by an inadequate recruitment of unaltered cognitive processes, and this is the issue we address here. Moreover, we aimed at developing a passive brain-

Eva, M., Olivier, B. and Ludovic, F.
How Does Cognitive Fatigue and Mental Workload Influence Alarm Detection in Flight Simulator? Classification of Electrophysiological Signatures with Explainable IA.
DOI: 10.5220/0011958100003622
In *Proceedings of the 1st International Conference on Cognitive Aircraft Systems (ICCAS 2022)*, pages 47-51
ISBN: 978-989-758-657-6

computer interfaces (pBCI) based on explainable classification and explainable machine learning to infer the influence of cognitive fatigue on inattentional deafness in the context of flying. To achieve these ends and following previous studies (Dehais et al., 2019, 2018), we asked participants to perform an alarm-detection task during repeated landing sessions on flight simulator. To accentuate the presence of cognitive fatigue, we also manipulated the mental workload. Most originally, we tested whether a real flight glider-instruction prior the experiment influence performance in the alarm detection task on flight simulator. We hypothesized that (a) cognitive fatigue impairs alarm detection as a function of the mental workload (b) cognitive fatigue modulates electrophysiological activities and (c) these modulations can be used as a predictor of a loss in pilot's efficiency.

2 METHODS

Twenty-four pilot-students were recruited for the experiment (*mean age*: 22.6 years old, *SD* = 2.0; flight experience: 75.6 hours, *SD* = 79.6 hours, including 44.7 hours of glider experience, *SD* = 58.9 hours). Half of the participants had normal daily activities without any training flight during the whole day of the experiment (NFBE group), and the other half had an instruction flight just before the experiment (IFBE group).

In a first time, participants were asked to evaluate their level of subjective fatigue (with Visual Analogous Scales of fatigue and sleepiness, VASf, and VASs), sleepiness (Karolinska's Sleepiness Scale), and alertness (Samn-Perelli scale). Then, they performed a Stroop task and an arithmetic task to assess their cognitive control.

In a second time, participants had to perform 6 identical and successive landings in a flight simulator (on a glider) while performing an alarm detection task (i.e., auditory Oddball). In this task, they had to detect rare sounds (i.e., target) and press a key on the joystick as fast and accurate as possible. In average, 100 sounds were played during a landing, among which 75 were standard sounds (i.e., to be ignored) and 25 had to be responded to. The landing was composed of 2 phases: a low cognitive load phase (i.e., corresponding to the downwind leg of the approach: they had to pilot the glider and perform the Oddball task) and a high cognitive load phase (i.e., composed of the base leg, the final and the landing: they had to pilot the glider, perform the Oddball task, and perform a backward counting task). Brain activity was recorded by a Bionic-EEG (32 passive electrodes).

After the experimental task, they had to perform again the cognitive tests and the subjective evaluations.

3 MAIN FINDINGS

3.1 Alarm Detection Task

Performance was analyzed with mixed-design ANOVAs, 2 (Group: NFBE, IFBE) x 2 (Time on Task: beginning—the first three landings, end—the last three landings) x 2 (Cognitive Load: low, high) with group as the only between-participants factor.

In the low cognitive load condition, participants were faster and more accurate to detect alarms than in the high cognitive load condition. A lack of attentional resources is thus associated with higher rates of inattentional deafness. Surprisingly, we found better alarm detection in the IFBE group than in the NFBE group. One possible explanation to this finding is that participants of the IFBE group were more trained to detect alarms due to the prior instruction flight, compared to NFBE group participants. No difference of alarm detection rate was observed throughout successive landings.

3.2 Electrophysiological Signatures

Data were analyzed with mixed-design ANOVAs, 2 (Group: NFBE, IFBE) x 3 (Electrode: Fz, Cz, Pz) x 2 (Sound: target, standard) x 2 (Cognitive Load: low, high) x 2 (Time on Task: beginning, end) with group as the only between-participants factor.

3.2.1 ERP Analyses

We found that the amplitude of the P300 component was higher with target sounds (i.e., rare alarms) than with standard sounds (i.e., frequent sounds to ignore) only in the low cognitive load condition. Moreover, for target sounds, we found an increase of the P300 amplitude under high cognitive load condition compared to the low cognitive load condition.

No effect of cognitive fatigue was observed on the amplitude of the P300 component. Possibly, our task was not sufficiently difficult to increase cognitive fatigue and to observe modifications on ERPs.

3.2.2 Frequency Analysis

Results showed that the spectral power of δ band tended to vary as a function of the temporal window. The spectral power was larger at the beginning compared to the end of the task. The effect of cognitive load was observed on the β spectral power for the NFBE group. β spectral power was larger for high cognitive load condition than for low cognitive load condition. These results are correlated with slower latencies in alarm detection observed under high cognitive load conditions.

3.3 Subjective Scales and Cognitive Tests

No differences were observed between the beginning and the end of the experimental session for the Visual Analogous Scale of Fatigue, the Samn-Perelli scale and the Karolinska scale.

Performance in the Stroop task was analyzed with mixed-design ANOVAs, 2 (Group: IFBE, NFBE) x 2 (Session: pre, post) x 2 (Congruency: congruent, incongruent), with group as the only between-participants factor. We found a significant congruency effect (i.e., better performance on congruent trials compared to incongruent trials). Most interestingly, the NFBE group performed better than the IFBE group (i.e., 96.3% vs. 94.6%). The interference score increased after the experimental task only in the IFBE group. These results suggest a decrease of cognitive control for participants of the IFBE group compared to the NFBE group. Performing the same activity before and during the experiment could lead participants to be less accurate particularly when they had to inhibit automatic responses. The control of automatic response and more generally the cognitive control seems to depend on the nature of the preceding task.

To summarize, we cannot conclude that cognitive fatigue is responsible for the observed modulations of electrophysiological activities. However, it is possible that the manipulation of cognitive load during sustained activity influences brain activity, as suggested by the modulation of the δ and β frequency bands. These manipulations could have resulted in a significant modulation of the subjective cognitive fatigue in other conditions (i.e., longer runs, more complex weather conditions, more landings...). In the low cognitive load condition, participants benefit from more attentional resources to process target sounds than in in the high cognitive load condition. These differences do not exist for standard sounds that must be ignored. In other words, cognitive fatigue could seem to impair performance as a function of attentional resources available. The frequency analysis can also be explained in term of decrease in attentional resources, but the differences between the beginning and the end of experiment could also reflect a lack of motivation at the end of the experiment.

3.4 Single Trial Classification of Alarm Detection or Omission and Decision-Making Tress

To compare electrophysiological signals between alarm detections and alarm omissions, we focused our analyses on the high cognitive load condition. Data were analyzed with 2 (Group: IFBE, NFBE) x 2 (Time on Task: beginning, end) x 3 (Electrode: Fz, Cz, Pz) x 2 (Response: hit, miss) ANOVAs with group as the only between-participants factor. 80% of trials were used to train classifiers and 20% were used to test them.

The spectral power of the δ frequency band and the α frequency band was larger for hit trials compared to miss trials. The differences between hit and miss trials were significant only at the beginning of the session. Moreover, the spectral power of the mid-β frequency band in the NFBE group was larger for hits than for missed alarms at the beginning of the session. We then classified trials with respect to alarm omission or detection and we reached a maximum averaged performance of 76.4% (range: 57.7% — 90.5%) in participant-specific single-trial classification from the spectral power of δ and α frequency bands with Support Vector Machines classifier. Frequency features, and more specifically δ and α bands, implemented in a support vector classifier formed an efficient tool to assess auditory alarm misperception in simulated flight conditions.

4 CONCLUSIONS

A way to improve the experimentation domain consists of putting the classification work above in a virtuous cognitive loop; to do so, we need explainable classification methods to be able to interpret the knowledge acquired. Such an understandable information, which can be either numerical, symbolic, or logical, constitutes the support of rich human/machine interactions and justifies the interpretability criterion providing a good level of confidence at the operational level. For instance, the Classification and Regression Trees (CART) algorithm delivers logical rules as the criteria separating alarm omission and detection from values on the four centered electrodes Cz, Pz, Oz and Fz. Starting from a normalized form of these rules which is easily explainable as a Boolean expression, we can generate the appropriate code in a static context or in dynamic context. In a static context, once EEG values are available, one can predict attention failure regardless to the software involved as the implementation context (Python, Java, Matlab, …); in a dynamic context, this is more interesting: since one can define an active role for electrodes taken as agents with a dedicated level of knowledge. That way, one can reengineer completely these rules according to electrodes as both actuators and sensors. That way, one can improve the experimentation domain by of putting the classification work in the loop thanks to a multi-agent model. Active electrodes become virtual agents (sensors and actuators) connected together thanks to logical connectors as firing rules. including other actuators (red light alarm, sounds, …). Domain-specific scenarios and doctrines can be defined. thanks to explainable classification. From that situation awareness, one can expect connect more powerful automatic decision mechanisms. In effect, abnormal behavior detection is the first step of the sense-making process relayed by decision-making. For instance, the purpose is to trigger a sequence of actions to be engaged, whether these actions are automatic or not. As a use-case, one can mention the situation in a cockpit characterized by a loss of attention of the pilot and his/her inability to continue his/her current mission. That is, the operator did not consciously detect the alarm although his brain processed the signal. It is therefore necessary to inform the operator that he has omitted the alarm (by feedback) and to adapt the work environment with the explainable AI to help him in his task so that he comes back in the loop.

REFERENCES

Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: An empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied*, *15*(2), 163-181.

Barwick, F., Arnett, P., & Slobounov, S. (2012). EEG correlates of fatigue during administration of a neuropsychological test battery. *Clinical Neurophysiology*, *123*(2), 278-284.

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2005). Effects of mental fatigue on attention: An ERP study. *Cognitive Brain Research*, *25*(1), 107-116.

Boksem, M. A. S., & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Research Reviews*, *59*(1), 125-139.

Borghini, G., Vecchiato, G., Toppi, J., Astolfi, L., Maglione, A., Isabella, R., Caltagirone, C., Kong, W., Wei, D., Zhou, Z., Polidori, L., Vitiello, S., & Babiloni, F. (2012). Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices. *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 6442-6445.

Dehais, F., Duprès, A., Blum, S., Drougard, N., Scannella, S., Roy, R., & Lotte, F. (2019). Monitoring Pilot's Mental Workload Using ERPs and Spectral Power with a Six-Dry-Electrode EEG System in Real Flight Conditions. *Sensors*, *19*(6), 1324.

Dehais, F., Dupres, A., Di Flumeri, G., Verdiere, K., Borghini, G., Babiloni, F., & Roy, R. (2018). Monitoring Pilot's Cognitive Fatigue with Engagement Features in Simulated and Actual Flight Conditions Using an Hybrid fNIRS-EEG Passive BCI. *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 544-549.

Dönmez, K., & Uslu, S. (2018). The Relationship Between Flight Operations and Organizations in Aircraft Accidents; The Application of the Human Factor Analysis and Classification System. *Anadolu University Journal of Science and Technology A - Applied Sciences and Engineering*, 1-1.

Faber, L. G., Maurits, N. M., & Lorist, M. M. (2012). Mental Fatigue Affects Visual Selective Attention. *PLoS ONE*, *7*(10), e48073.

Inzlicht, M., Schmeichel, B. J., & Macrae, C. N. (2014). Why self-control seems (but may not be) limited. *Trends in Cognitive Sciences*, *18*(3), 127-133.

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36*(6), 661-679.

Lee, K. A., Hicks, G., & Nino-Murcia, G. (1991). Validity and reliability of a scale to assess fatigue. *Psychiatry Research*, *36*(3), 291-298.

Lorist, M. M., Boksem, M. A. S., & Ridderinkhof, K. R. (2005). Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cognitive Brain Research*, *24*(2), 199-205.

Marcus, J. H., & Rosekind, M. R. (2017). Fatigue in transportation : NTSB investigations and safety recommendations. *Injury Prevention*, *23*(4), 232-238.

Möckel, T., Beste, C., & Wascher, E. (2015). The Effects of Time on Task in Response Selection—An ERP Study of Mental Fatigue. *Scientific Reports*, *5*(1), 10113.

Müller, T., & Apps, M. A. J. (2019). Motivational fatigue : A neurocognitive framework for the impact of effortful exertion on subsequent motivation. *Neuropsychologia*, *123*, 141-151.

Sabeti, M., Boostani, R., & Rastgar, K. (2017). How mental fatigue affects the neural sources of P300 component? *Journal of Integrative Neuroscience*, *17*, 1-19.

Salomone, M., Burle, B., Fabre, L., & Berberian, B. (2021). An Electromyographic Analysis of the Effects of Cognitive Fatigue on Online and Anticipatory Action Control. *Frontiers in Human Neuroscience*, *14*, 615046.

Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., & Kincses, W. E. (2009). Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accident Analysis & Prevention*, *41*(5), 1087-1093.

Trejo, L. J., Knuth, K., Prado, R., Rosipal, R., Kubitz, K., Kochavi, R., Matthews, B., & Zhang, Y. (2007). EEG-Based Estimation of Mental Fatigue : Convergent Evidence for a Three-State Model. In D. D. Schmorrow & L. M. Reeves (Éds.), *Foundations of Augmented Cognition* (Vol. 4565, p. 201-211). Springer Berlin Heidelberg.

Wascher, E., Rasch, B., Sänger, J., Hoffmann, S., Schneider, D., Rinkenauer, G., Heuer, H., & Gutberlet, I. (2014). Frontal theta activity reflects distinct aspects of mental fatigue. *Biological Psychology*, *96*, 57-65.

Zhao, C., Zhao, M., Liu, J., & Zheng, C. (2012). Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accident Analysis & Prevention*, *45*, 83-90.