Automation surprise in the neuroadaptive cockpit

Juliane Pawlitzki[1,2,*], Oliver Klaproth[3], Laurens R. Krol[1,2], Thorsten O. Zander[1,2]

[1]Neuroadaptive Human-Computer Interaction, Brandenburg University of Technology, Cottbus – Senftenberg, Germany

[2]Zander Laboratories B.V., Amsterdam, the Netherlands

[3]Airbus Central R&T, Hamburg, Germany

*Corresponding author: juliane.pawlitzki@gmail.com

Background. Operating an aircraft is multi-dimensional and complex. The pilot has to 'aviate, navigate, communicate' – stay airborne, manage the aircraft's course and talk to Air Traffic Control. To facilitate these tasks, automation was introduced to cockpits (Billings, 1997). When this automation fails, the consequences are annoying at best and life-threatening at worst (Endsley & Kiris, 1995). Errors in the automation can be surprising and distracting, resulting in *automation surprises* (Boer & Dekker, 2017). These can cause confusion in the pilot, which in turn can lead to human error, a major cause of aviation accidents (Lyssakov, 2019). Identification of this confusion and its cause potentially improves the interaction of human and machine (Dehais et al., 2015).
In a previous study (Krol et al., 2018), we showed that it is possible to record a pilot's cognitive reaction to flight-relevant events via electroencephalography (EEG; Berger, 1929), use passive Brain-Computer Interfaces (pBCI; Zander & Kothe, 2011) to determine different levels of event criticality and report the interpretation back to the cockpit in real-time. This procedure can be used to adapt the cockpit to the pilot's cognition, resulting in a *neuroadaptive cockpit* (Krol et al., in press).
In this study, we developed a more specific classifier that reliably detects a pilot's cognitive reaction to surprising and/or erroneous flight-relevant events which could be critical to the continuous task of operating the aircraft.

Methods. EEG activity and eye movements of 13 test pilots (all male), aged 44-62 years (mean 54) with 7210 ± 4809 hours of flight experience, were recorded. We used a 32-channel mobile, wireless EEG system[1] and binocular eye tracking glasses[2] in a two-part experiment.

In the first part, participating pilots conducted 10 blocks of a newly devised training paradigm. We intended to calibrate different classifiers for surprising events (S-classifier), erroneous events (E-classifier) and events which were both surprising and erroneous (AS-classifier), corresponding to possible automation surprises. Thus, we devised a combination of training paradigms, the *interaction oddball paradigm*. This paradigm consists of 2 separate parts, and classifiers are trained on different parts of the resulting data. To evoke cognitive states which correspond to surprise and/or error, we simulated a computer program which needs to be taught when to count a tone and when to ignore it.
A sequence of 50 tones was presented in each of 10 blocks. Each tone could either be a standard (probability 70%-80%), a non-target (10%-15%) or a target (10%-15%). This represents a standard oddball-paradigm (Friedman et al., 2001). It was found that the target tones evoke surprise in the participant (Squires et al., 1975). Participants were instructed to state verbally after every tone whether it was a target ('Yes') or not ('No'). The computer then gave acoustic feedback: 'Count' or 'Ignore'. Since the voice recognition was (unbeknownst to the participant) simulated, feedback was independent from the participant's assessment. This enabled us to control the number of occurring errors in the feedback. In the first 7 blocks, there was a 14%-18% probability of incongruent feedback, i.e. the computer replying 'Ignore' after a 'Yes' or 'Count' after a 'No'. This corresponds to rare, surprising errors. In the last 3 blocks, incongruence probability was 38-40%, corresponding to frequent errors.

---

[1] Brain Products LiveAmp
[2] SMI ETG 2w

In the second part of the experiment, pilots performed a flight session in a fixed-base simulator similar to current-generation Airbus A350 cockpits. The session consisted of 5 5- to 25-minute scenarios with 1 to 4 flight-critical events each.

On the training EEG data, eye components were identified with an independent component analysis (ICA; Makeig et al., 1996) and removed for classifier calibration. For each individual participant, we trained 3 different classifiers based on a windowed-means approach combined with regularized LDA (Blankertz et al., 2011). The classifiers were trained on different parts of the data:

- the S-classifier was trained on data following a standard (no surprise) and following a target (surprise);
- the E-classifier was trained on data from blocks 8-10 following congruent (no error) and incongruent feedback (frequent error);
- the AS-classifier was trained on data from blocks 1-7 following congruent (no error) and incongruent feedback (rare error).

We then applied these classifiers to the flight session data of the same participant. Every event obtained three different assessments, regarding surprise, error and automation surprise. Due to simultaneous eye-tracking and self-assessment of the pilot during subsequent debriefing, it was possible to assess whether an alert during the flight session was expectable for the pilot. Thus, it was possible to compute application accuracies of the calibrated classifiers, and compare them statistically.

Results. All classifiers reached averaged training accuracies significantly above chance: 83% for the S-classifier, 85% for the AS-classifier and 74% for the E-classifier. Figures 1(a),(b),(c) show sensitivity, specificity and accuracy of the classifiers applied to flight data. Significant differences are indicated using asterisks. Although they were trained on similar data (feedback with rare/frequent errors), significant differences ($F(2,30)=4.51$, $p<0.05$) occurred between the AS- and the E-classifier. The S- and AS-classifier gave similar outputs, despite being focused on different cognitive reactions.

Discussion and conclusion. Our findings indicate that automation surprises do not exclusively evoke surprise in the pilot, but also error. This distinction is especially relevant regarding surprising errors resulting from flawed automation. While different alarms or messages from air traffic control can result in surprise, we want to focus on surprising, erroneous alerts. Alerts which cannot be anticipated and startle the pilot need to be detected separately from surprising events which are not necessarily flight-relevant. In an earlier study (Krol et al., 2018), we showed that it is possible to train a classifier using an abstract task and apply it in a different, realistic scenario, yielding meaningful estimates of the pilot’s cognitive state. This finding could be reproduced here. Additionally, we have devised a novel paradigm to train a classifier which detects both surprise and error and shows advantages at identifying flight-critical events over conventional classifiers.

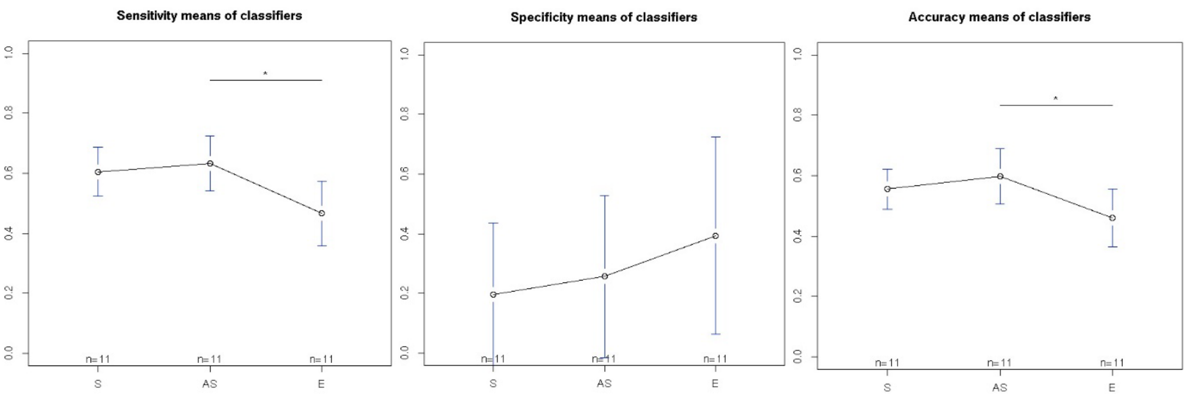

Figure 1: a) Sensitivity means of classifiers S, AS, E; b) Specificity means of classifiers S, AS, E; c) Application accuracy means of classifiers S, AS, E. Asterisks indicate significant differences (p < 0.05), bars represent standard errors of the mean.

References

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, *87*, 527–570.

Billings, C. E. (1997). *Aviation Automation: The Search for a Human-Centered Approach*. Lawrence Erlbaum Associates.

Blankertz, B., Lemm, S., Treder, M. S., Haufe, S. & Müller, K.-R. (2011). Single-trial analysis and classification of ERP components - A tutorial. *NeuroImage*, *56*(2), 814–825.

Boer, R. J. de & Dekker, S. (2017). Models of Automation Surprise: Results of a Field Survey in Aviation. *Safety*, *3*(3).

Dehais, F., Peysakhovich, V., Scannella, S., Fongue, J. & Gateau, T. (2015). "Automation Surprise" in Aviation: Real-Time Solutions. *CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computer Systems*, 2525–2534.

Endsley, M. R. & Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors*, *37*(2), 381–394.

Friedman, D., Cycowicz, Y. M. & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, *25*(4), 355–373.

Krol, L. R., Klaproth, O., Vernaleken, C., Wetzel, I., Gaertner, J., Russwinkel, N. & Zander, T. O. (2018). Towards a Neuroadaptive Cockpit: First Results. In *Proceedings of the 3rd International Mobile Brain/Body Imaging Conference*.

Krol, L. R., Zander, T. O., Klaproth, O., Vernaleken, C. & Russwinkel, N. (in press). Towards Neuroadaptive Modelling: Assessing the Cognitive States of Pilots through Passive Brain-Computer Interfacing. *S. H. Fairclough & T. O. Zander (Eds.), Current Research in Neuroadaptive Technology.*

Lyssakov, N. (2019). Human factor as a cause of aircraft accidents: II International Scientific-Practical Conference "Psychology of extreme Professioens" (ISPCPEP). *Advances in Social Science, Education and Humanities Research*, *321*.

Makeig, S., Jung, T.-P., Bell, A. J. & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. In D. S. Touretzky, M. C. Mozer & M. E. Hasselmo (Hrsg.), *Advances in Neural Information Processing Systems* (Bd. 8, S. 145–151). MIT Press.

Squires, N. K., Squires, K. C. & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401.

Zander, T. O. & Kothe, C. A. (2011). Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *Journal of Neural Engineering*, *8*(2), 25005. https://doi.org/10.1088/1741-2560/8/2/025005