A Window into the Mind? On Usefulness and Challenges of Neurophysiological Measurements in the Cockpit

A. Hamann[1] and N. Carstengerdes[1]

1 Institut für Flugführung, Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR) Braunschweig, Germany. anneke.hamann@dlr.de

Abstract

Neurophysiological measurements seem a powerful tool for investigating pilots' cognitive states. Such measurements need to be reliable and valid in order to gain interpretable and robust results. Recent research, however, suggests that this may not be as easy as it seems. In this paper, we give a short overview of commonly used neurophysiological measurements, and problems that need to be overcome to make these measurements the valid, reliable source of information we wish for.

Introduction: Aviation, Automation, Adaptation

The history of aviation is one of rapid increases in technological advances and ever higher levels of automation [8]. These developments go hand in hand with decreasing numbers of crew in the cockpit, following the assumption that higher levels of automation result in tasks that can be achieved by fewer human pilots [1]. Historically, the task allocation between human and machine has followed simple guidelines based on the capabilities of the actors [12], resulting in a static allocation. Later on, dynamic allocations arose in the form of adaptable automation which the human can changes based on their current needs, or adaptive automation in which the task allocation is changed by the machine, triggered by pre-defined situational, environmental or task-related factors [26, 31]. With the emergence of more sophisticated machine learning algorithms, the aviation industry and research alike now hope to develop concepts of adaptive automation that are tailored to the pilots and their current needs, and are thus more flexible and more supportive than traditional systems [8]. In order to do so, however, the system needs information about the pilots, their current state and needs. In their quest for detailed information about the human operator, researchers have turned to physiological measurements [2, 3]. Peripheral measurements like electrocardiography or electrodermal activity are often used to indicate stress or high arousal [2]. Neurophysiological measurements like electroencephalography (EEG) are used to gain more detailed insights into the humans' brain activity and cognitive concepts like mental workload or mental fatigue [2]. With the emergence of wearable functional near-infrared spectroscopy (fNIRS) devices, this interest in neurophysiological measurements has peaked again, indicated by a growing body of research in the last years [35]. EEG and fNIRS are promising measurements for the use in the cockpit because of their low intrusiveness, possibility to gain data continuously, and of course their objectivity. Thus, they have huge advantages compared to self-report or performance data.

Yet, there is one question that needs to be answered before sophisticated adaptive assistance systems should be fed with (neuro-) physiological data: How precise are the information we can gain from said measurements?

Mental Workload and the Validity of EEG Measures

To give an example and highlight the problems with valid neurophysiological measurements, in this paper we will focus on one concept: Mental workload. Mental workload is one of the most frequently researched concepts in aviation, despite controversies about its definition and conceptualisation [10, 27]. Generally, mental workload is defined as the part of one's cognitive resources needed to accomplish a task [25]. The more difficult the task, the more cognitive resources one needs to allocate to its accomplishment. If one has no spare resources left, task performance will decline and errors will occur. In aviation, it is generally believed that a medium level of mental workload is ideal [22]: Too low mental workload can border on boredom and can prompt disengagement from the task, causing loss of attention and out-of-the-loop phenomena when the pilot is suddenly asked to actively engage again. On the other hand, too high mental workload can lead to declined performance and errors that could lead to catastrophic outcomes. In sum, it would be ideal to keep the pilots' mental workload at a comfortable medium level to ensure optimal performance. Future adaptive assistance systems could help with this, if the pilots' current mental workload level could be assessed.

There is a large and growing body of research on mental workload assessment using EEG. Researchers usually focus on frequency band analyses for this, i.e. the decomposition of the EEG signal into its frequency bands. Changes in the composition of the signal are used to assess changes in cognitive states, e.g. between an easy and a difficult task. This way, data can be collected continuously and without the need to insert additional stimuli into the cockpit environment to elicit responses (as is done in analyses of event-related potentials). Indeed, the literature suggests a classical, "tell-tale" pattern of changing signal compositions for mental workload: Increasing mental workload is usually indicated by an increase in frontal theta activity, accompanied by a decrease in parietal alpha activity [9, 11, 16]. Yet, upon closer inspection, this "tell-tale" pattern cannot be found in every publication. While this is to be expected in research, the reasons for failing to detect the expected patterns should concern us.

For example, there are other concepts that elicit very similar cognitive activity, such as mental fatigue. Mental fatigue is described as a sense of weariness usually induced by long monotonous, yet demanding tasks [6, 15]. Not unlike mental workload, it is also based on the idea of resource depletion. The longer one needs to focus on a task, the more cognitive resources one spends until these resources are used up. If the individual does not or cannot take a break to replenish their cognitive resources, they will experience a subjective feeling of fatigue and a general unwillingness to spend further effort [15]. In EEG measurements, mental fatigue usually manifests in increasing frontal theta activity, accompanied by increasing parietal alpha activity [33, 34, 36], the same regions involved in mental workload. This is not just inconvenient when one wants to differentiate the concepts. There is research indicating that confounding mental workload and mental fatigue (when a strenuous task is performed over a long period of time) decreases the accuracy of mental workload classification [30]. Why? Because decreasing alpha activity with mental workload and increasing alpha activity with mental fatigue may cancel each other out and result in no detectable changes [30]. And mental fatigue is not the only concept that can interfere with a mental workload assessment. Frequent task switching, for example between navigation and communication tasks, also affects the alpha frequency band [29] and could be responsible for "missing" changes in parietal alpha activity [18]. On top of that, emotional responses, especially negative ones, impact frontal theta activity [16] and have the potential to interfere with the assessment as well.

These examples highlight why valid mental workload assessment may be achievable in the laboratory, but becomes rather difficult rather fast then turning towards realistic tasks. An algorithm trained to spot the typical, "tell-tale" mental workload pattern from laboratory studies can be confused easily when another factor comes into play in real-world settings. And yet, very few studies consider these problems and explicitly control for confounding factors or discuss limitations of their findings in this regard [4, 17]. It seems that an EEG-based system alone lacks validity. Maybe adding another data source could solve the problem?

Combining EEG and fNIRS

Combining fNIRS and EEG measurements has certain advantages. The higher temporal resolution of EEG and the higher spatial resolution of fNIRS measurements complement each other and can lead to more accurate results of when and where exactly changes happen, if analysed accordingly. And there is also the possibility to analyse convergent validity: If both measurements point towards the same result, it is more likely that the finding is true and not influenced by a confounding factor.

fNIRS measures the cortical activation based on changing oxygen concentrations in the cortical blood flow. Overly simplified, increasing brain activity increases the consumption of oxygen and thus the flow of oxygenated blood in the activated region, while the levels of deoxygenated blood in said region decrease simultaneously. Because of the different optical properties of oxygenated and deoxygenated blood, fNIRS can make this process visible and give an indication of the changes in cortical activation (for a more thorough explanation see [17]). The majority of studies indicate increasing (pre-) frontal cortical oxygenation (more oxygenated, less deoxygenated blood) with increasing mental workload [5, 13, 14]. There is one problem, however. The literature also points towards increasing (pre-) frontal cortical activation with increasing mental fatigue [7, 23, 24]. This could indicate that frontal cortical activation, as captured by fNIRS, is an indication of general demand, but rather unspecific. Unfortunately, most studies are performed on frontal and prefrontal regions, so there are only few published findings on parietal activation to this date [19]. In sum, while fNIRS is an interesting addition to EEG, the results gained from the measurement cannot differentiate mental workload from other types and sources of demand. This is highly problematic for the development of assistance systems based on neurophysiology. Assistance for a fatigued pilot could look very different from assistance for an overloaded pilot, and a system unable to tell the difference may adapt in the wrong direction. Moreover, if both effects indeed cancel each other out, an overloaded and fatigued pilot may not be offered any assistance because the system could no longer detect “unwanted” patterns of activation.

Our Systematic Approach to Achieving Validity

What can we as researchers do to overcome this problem of validity? The simple answer: Be aware of the limitations of the methods we apply, and investigate their limits systematically. In order to find valid neurophysiological measures of mental workload, we did a series of studies, systematically focusing on mental workload [18] and mental fatigue [19] while controlling for the influence of the respective other. In the following, we detail our approach. The described experiments were approved by the ethics commission of the German Psychological Society (DGPs) and conducted in accordance with the declaration of Helsinki.

Internal Validity: Inducing a Concept while Controlling for Confounds

If one wants to disentangle the neurophysiological correlates of mental workload from those of other concepts, one needs to make sure to induce only mental workload and control for other influences. In our recent research [18], we took great care to induce mental workload in four levels by means of increasing the difficulty of an adapted n-back task that was tailored to the flight context. We controlled for influences of mental fatigue by keeping the duration of the task to approx. 45 minutes, by randomizing the difficulty levels and by means of statistical analyses. Moreover, we used self-report and performance measures to ensure we actually did induce four distinct levels of mental workload.

Convergent Validity: Combining and Comparing Measurement Methods

We performed simultaneous fNIRS and EEG measurements to investigate the convergent validity between the measurements. In order to do so, we used compatible devices and chose a montage in which EEG electrodes and fNIRS optodes covered the same areas. There is software available for such purposes, like the fNIRS Optodes’ Location Decider fOLD [37], and we highly recommend making use of such tools as well as documenting the exact measurement locations in order to foster replicability.

Sensitivity and Specificity: Testing Measures on Different Concepts

In order to see which neurophysiological changes were unique to mental workload and could not be mistaken for mental fatigue, we also needed to research mental fatigue. Thus, we conducted a subsequent experiment in which we induced mental fatigue and controlled for confounding with mental workload [19]. By comparing the results of both studies, we would be able to see which neurophysiological measures were sensitive to mental workload, i.e. would vary with increasing mental workload, and specific to it, i.e. would not vary with increasing mental fatigue. In order to make the results comparable we designed the second experiment to be as similar as possible to the first. we used the same measurement equipment and montage, the same laboratory and flight simulator and the same cognitive task. We induced mental fatigue by increasing time on task (90 minutes) and kept mental workload constant by applying only one moderate difficulty level derived from our first study (details on our methods can be found in [18, 19]). We also chose the same measures, and data processing and analysis steps we had used in our first study.

Our Results: As Valid as Can Be?

As our current aim is to highlight the methods rather than the results, we will only briefly discuss our results here and refer the interested reader to the individual publications [18–20] as well as a comparative in-depth discussion in [17]. In short, we analysed and compared the fNIRS data (oxygenated blood HbO, deoxygenated blood HbR) and EEG frequency bands (frontal theta, and parietal alpha and beta), as well as two commonly used mental workload indices (Task Load Index TLI [32]; Engagement Index EI [28]) [20] between both experiments.

An overview of our findings is shown in Figure 3. We found that frontal measures (i.e. cortical oxygenation, theta band power and TLI) were sensitive to changes in mental workload, and that by combining EEG and fNIRS data, we could differentiate all four induced levels of mental workload, more than with the separate measurements. However, the frontal EEG measures proved sensitive to increasing mental fatigue as well, thus lacking specificity to mental workload. fNIRS may prove a viable option to differentiate the concepts, but our results were somewhat inconclusive (for a discussion see [17]). Parietal activity (i.e. alpha and beta band power and EI) lacked sensitivity to changing mental workload altogether.

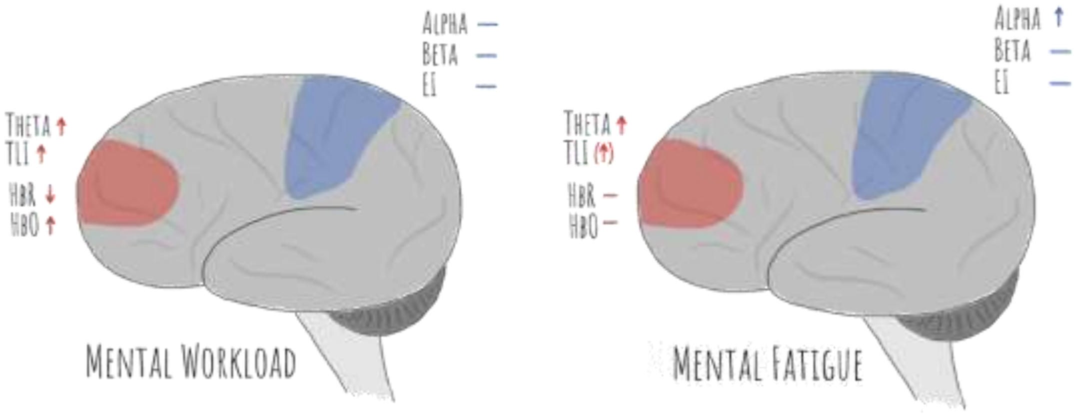

Figure 3. Schematic comparison of the results of our previous studies for mental workload and mental fatigue [18–20]. Red indicates frontal cortical areas, blue indicates parietal areas. Significant increases are marked ↑; significant decreases are marked ↓; no significant changes are marked – ; brackets indicate significant overall effect but no significant post-hoc tests. Figure taken from Hamann, 2023, p. 55 [17].

Conclusion

Even though we took great care to control for confounding, and could thus compare “pure” effects of mental workload and mental fatigue, it was quite difficult to tell the concepts apart. What is evident is that increasing frontal cortical activation indicates increasing demand, but cannot be used to explain its cause exactly, even under laboratory conditions. Unfortunately, in the cockpit it is unlikely to find pure mental workload. During long-haul flights, mental fatigue may well arise. Rostering and long shifts can lead to sleepiness on top of mental workload. And frequent task switching is inherent to piloting an aircraft. Thus, neurophysiological measurements may not be ideal for measuring mental workload “in the wild” and for differentiating cognitive concepts. We may be able to capture increasing or decreasing demand in a pilot, but not where exactly this demand originates from.

This may sound a little disappointing, but does not mean we should abandon neurophysiological measurements altogether. The fact alone that we can gain insights into a pilot’s current cognitive demand and monitor changes should be impressive enough. And maybe this ability is already sufficient for our purposes. In an aircraft, there is an abundance of other sources of information readily available. Covariates like the duration of the mission, weather, aircraft system health indications and even the pilots' inputs into the system could bridge the gap between cognitive demand and its origin [21]. Increasing frontal activation in combination with system failure messages may be a good indication of mental workload, while the same pattern of frontal activation after multiple uneventful hours of flight is likely due to mental fatigue. Such a multimodal approach and combination of the overall state of both pilot (via neurophysiology) and aircraft (via covariates) could help to achieve the vision of an adaptive assistance system that is tailored to the current needs of the pilot without relying too much on the power of one measurement. In the end, the important part is to be aware of the capabilities and limitations of our methods instead of just assuming validity, and to apply them carefully where they are suitable and useful.

References

1. Billings, C. E. (1991). Toward a human-centered aircraft automation philosophy. *The International Journal of Aviation Psychology* **1**(4), 261–270. doi:10.1207/s15327108ijap0104_1.
2. Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews* **44**, 58–75. doi:10.1016/j.neubiorev.2012.10.003.
3. Boumann, H., Hamann, A., Biella, M., Carstengerdes, N., & Sammito, S. (2023). Suitability of physiological, self-report and behavioral measures for assessing mental workload in pilots. In D. Harris & W.-C. Li (Eds.), *Engineering Psychology and Cognitive Ergonomics* (Vol. 14017, pp. 3–20, Lecture Notes in Computer Science). Cham: Springer Nature Switzerland.
4. Brouwer, A.-M., Zander, T. O., van Erp, J. B. F., Korteling, J. E., & Bronkhorst, A. W. (2015). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Frontiers in Neuroscience* **9**, 136. doi:10.3389/fnins.2015.00136.
5. Causse, M., Chua, Z. K., Peysakhovich, V., Del Campo, N., & Matton, N. (2017). Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS. *Scientific Reports* **7**(1), 5222. doi:10.1038/s41598-017-05378-x.
6. Charbonnier, S., Roy, R. N., Bonnet, S., & Campagne, A. (2016). EEG index for control operators' mental fatigue monitoring using interactions between brain regions. *Expert Systems with Applications* **52**, 91–98. doi:10.1016/j.eswa.2016.01.013.
7. Chuang, C.-H., Cao, Z., King, J.-T., Wu, B.-S., Wang, Y.-K., & Lin, C.-T. (2018). Brain electrodynamic and hemodynamic signatures against fatigue during driving. *Frontiers in Neuroscience* **12**, 181. doi:10.3389/fnins.2018.00181.
8. CIEHF (2020). *The human dimension in tomorrow's aviation system: White Paper* .
9. Dehais, F., Duprès, A., Blum, S., Drougard, N., Scannella, S., Roy, R. N., et al. (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). doi:10.3390/s19061324.
10. Dekker, S., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology & Work* **6**(2), 79–86. doi:10.1007/s10111-003-0136-9.
11. Dussault, C., Jouanin, J.-C., & Guezennec, C.-Y. (2004). EEG and ECG changes during selected flight sequences. *Aviation, Space, and Environmental Medicine* **75**(10), 889–897.
12. Fitts, P. M. (Ed.) (1951). *Human engineering for an effective air-navigation and traffic-control system* . Washington, DC, USA: National Research Council.
13. Gateau, T., Ayaz, H., & Dehais, F. (2018). In silico vs. over the clouds: On-the-fly mental state estimation of aircraft pilots, using a functional near infrared spectroscopy based passive-BCI. *Frontiers in Human Neuroscience* **12**, 187. doi:10.3389/fnhum.2018.00187.
14. Geissler, C. F., Domes, G., & Frings, C. (2020). Shedding light on the frontal hemodynamics of spatial working memory using functional near-infrared spectroscopy. *Neuropsychologia* **146**, 107570. doi:10.1016/j.neuropsychologia.2020.107570.
15. Grandjean, E. (1979). Fatigue in industry. *British Journal of Industrial Medicine* **36**(3), 175–186. doi:10.1136/oem.36.3.175.

16. Grissmann, S., Faller, J., Scharinger, C., Spüler, M., & Gerjets, P. (2017). Electroencephalography based analysis of working memory load and affective valence in an N-back task with emotional stimuli. *Frontiers in Human Neuroscience* **11**, 616. doi:10.3389/fnhum.2017.00616.
17. Hamann, A. (2023). *A systematic investigation of EEG and fNIRS measures for the assessment of mental workload in the cockpit*. Dissertation. Technische Universität Dresden, Dresden, Germany. https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-869293.
18. Hamann, A., & Carstengerdes, N. (2022). Investigating mental workload-induced changes in cortical oxygenation and frontal theta activity during simulated flights. *Scientific Reports* **12**(1), 6449. doi:10.1038/s41598-022-10044-y.
19. Hamann, A., & Carstengerdes, N. (2023). Assessing the development of mental fatigue during simulated flights with concurrent EEG-fNIRS measurement. *Scientific Reports* **13**(1), 4738. doi:10.1038/s41598-023-31264-w.
20. Hamann, A., & Carstengerdes, N. (2023). Don't Think Twice, It's All Right? – An examination of commonly used EEG indices and their sensitivity to mental workload. In D. Harris & W.-C. Li (Eds.), *Engineering Psychology and Cognitive Ergonomics* (Vol. 14017, pp. 65–78, Lecture Notes in Computer Science). Cham: Springer Nature Switzerland.
21. Hinss, M. F., Brock, A. M., & Roy, R. N. (2022). Cognitive effects of prolonged continuous human-machine interaction: The case for mental state-based adaptive interfaces. *Frontiers in Neuroergonomics* **3**. doi:10.3389/fnrgo.2022.935092.
22. Martins, A. P. G. (2016). A review of important cognitive concepts in aviation. *Aviation* **20**(2), 65–84. doi:10.3846/16487788.2016.1196559.
23. Nguyen, T., Ahn, S., Jang, H., Jun, S. C., & Kim, J. G. (2017). Utilization of a combined EEG/NIRS system to predict driver drowsiness. *Scientific Reports* **7**, 43933. doi:10.1038/srep43933.
24. Nogueira, M. G., Silvestrin, M., Barreto, C. S. F., Sato, J. R., Mesquita, R. C., Biazoli, C., et al. (2022). Differences in brain activity between fast and slow responses on psychomotor vigilance task: an fNIRS study. *Brain Imaging and Behavior*, 1–12. doi:10.1007/s11682-021-00611-8.
25. O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* . New York: John Wiley & Sons.
26. Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., & Barnes, M. (1992). *Theory and design of adaptive automation in aviation systems: Progress report* . Warminster, PA, USA.
27. Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making* **2**(2), 140–160. doi:10.1518/155534308X284417.
28. Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology* **40**(1-2), 187–195. doi:10.1016/0301-0511(95)05116-3.
29. Puma, S., Matton, N., Paubel, P.-V., Raufaste, É., & El-Yagoubi, R. (2018). Using theta and alpha band power to assess cognitive workload in multitasking environments. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology* **123**, 111–120. doi:10.1016/j.ijpsycho.2017.10.004.
30. Roy, R. N., Bonnet, S., Charbonnier, S., & Campagne, A. (2013). Mental fatigue and working memory load estimation: Interaction and implications for EEG-based passive BCI. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* **2013**, 6607–6610. doi:10.1109/EMBC.2013.6611070.
31. Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance*: *Theory and Applications* (1st ed., pp. 37–63, Human Factors in Transportation). London: CRC Press.
32. Smith, M. E., Gevins, A., Brown, H., Karnik, A., & Du, R. (2001). Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction. *Human Factors* **43**(3), 366–380. doi:10.1518/001872001775898287.

33. Tran, Y., Craig, A., Craig, R., Chai, R., & Nguyen, H. (2020). The influence of mental fatigue on brain activity: Evidence from a systematic review with meta-analyses. *Psychophysiology* **57**(5), e13554. doi:10.1111/psyp.13554.
34. Trejo, L. J., Knuth, K., Prado, R., Rosipal, R., Kubitz, K., Kochavi, R. L., et al. (2007). EEG-based estimation of mental fatigue: Convergent evidence for a three-state model. In D. D. Schmorrow & L. M. Reeves (Eds.), *International Conference on Foundations of Augmented Cognition, Beijing, China* (Vol. 4565, pp. 201–211, Lecture Notes in Computer Science, Vol. 4565). Berlin: Springer. doi:10.1007/978-3-540-73216-7_23.
35. van Weelden, E., Alimardani, M., Wiltshire, T. J., & Louwerse, M. M. (2022). Aviation and neurophysiology: A systematic review. *Applied Ergonomics* **105**, 103838. doi:10.1016/j.apergo.2022.103838.
36. Wascher, E., Rasch, B., Sänger, J., Hoffmann, S., Schneider, D., Rinkenauer, G., et al. (2014). Frontal theta activity reflects distinct aspects of mental fatigue. *Biological Psychology* **96**, 57–65. doi:10.1016/j.biopsycho.2013.11.010.
37. Zimeo Morais, G. A., Balardin, J. B., & Sato, J. R. (2018). fNIRS Optodes' Location Decider (fOLD): a toolbox for probe arrangement guided by brain regions-of-interest. *Scientific Reports* **8**(1), 3341. doi:10.1038/s41598-018-21716-z.