

A New Method for Mental Workload Registration

Thea Radüntz^(✉)

Unit. 3.4 ‘Mental Health and Cognitive Capacity’,
Federal Institute for Occupational Safety and Health,
Nöldnerstr. 40/42, 10000 Berlin, Germany
`raduentz.thea@baua.bund.de`

Abstract. Complex and highly automated systems impose high demands on employees with respect to cognitive capacity and the ability to cope with workload. Objectively registering mental workload at workplaces with high cognitive demands would enable prevention of over- and underload. Although urgently needed, such technical measurement is currently unfeasible. Hence, the goal of this work is the establishment of precisely such an objective method.

In this article we briefly present a new method for registering mental workload by means of the electroencephalogram (EEG). Based on so called Dual Frequency Head Maps (DFHM) every 5s we obtain an index of mental state ranging between the classes low, moderate, and high workload.

Finally, we present results from a sample set of 54 people during the execution of the cognitive tasks 0-back, stroop test and AOSPAN in a laboratory setting. We compare them with our expectations based on the knowledge of task requirements on the executive functions as well as with further workload relevant biosignal data, performance data, and the NASA-TLX as a subjective questionnaire method. By this we gain proof of the integrity of the new method.

Keywords: Mental workload · Electroencephalogram (EEG) · Signal processing · Pattern recognition

1 Introduction

The computerization of the modern working world is evolving ever more rapidly and aims to facilitate our life. However there is a growing consensus surrounding the negative consequences of inappropriate workload on employee health and on the safety of persons. The consequences may arise from the inability to cope with increasing demands imposed on an individuals cognitive capacity and hence due to high mental workload [4, 5, 8]. On the other hand the proliferation of automation can also be linked to monotonous tasks that reduce employees arousal and induce underload [1, 2, 6, 9, 10].

Although the relation between workload and performance is well studied and there has been years of research targeting the registering of mental workload, there is no generally accepted, reliable, objective, and continuous method for this. Such a method would allow for defining the individual optimal workload range in which task solving is most efficient. Our aim is to develop a method that relies on features from brain activity, the center of human information processing. Neuronal brain state monitoring can then be used for ergonomic evaluation and improvement of human-machine systems and hence contribute to the optimization of workload.

This article describes the development of a continuous method for neuronal mental workload registration. Cognitive tasks were conducted in a laboratory setting aiming to identify EEG features indicative for mental workload. For our task battery we selected tasks reflecting executive functions [7]. Such functions are responsible for everyday actions demanding non-schema-based processing and requiring attentional control. Hence, the EEG features obtained should allow for the development of a generalized method that can measure mental workload independently from tasks.

2 Methods

2.1 Procedure and Subjects

The investigation took place in the shielded lab of the Federal Institute of Occupational Safety and Health. Materials, procedures and the sample set have already been described in [13]. To recapitulate, the sample consists of 54 people between 34 and 62 years of age and shows high variability in respect to the cognitive capacity and hence to the expected mental workload. The experiment was fully carried out with each subject in a single day and consisted of a training phase where the subjects were familiarized with the tasks and the main experiment.

We used several workload measuring methods. Each of these has its own pros and cons. Hence, the collection of additional workload indexing data was aimed to consolidate the development of our method by giving us the opportunity to control possible subject-dependent confounders but also further information in case of doubts at the end.

2.2 Tasks

Different cognitive task requirements were realized through the implementation of a task battery with the E-Prime application suite. Nine tasks of diverse complexity and difficulty inducing different levels of mental workload are included in our test battery [11, 13].

In this paper we concentrate on the analysis and evaluation of three tasks: 0-back as the easiest one, stroop test as an inhibition task, and AOSPAN as a demanding dual task (see Figs. 1, 2, 3). The latter is a translated version of the AOSPAN task developed by [14]. The analysis of rest measurements serves as a reference point measurement.

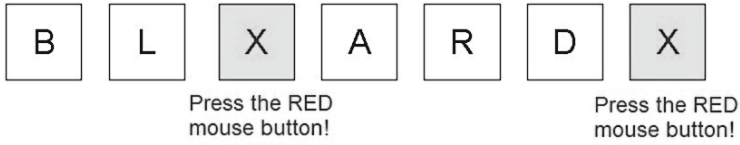


Fig. 1. 0-back task: Press the mouse button if the presented letter is ‘X’.



Fig. 2. Stroop task as an inhibition task: Differently colored words appear on the screen one at a time. Press the mouse button (yellow, green, red, blue) that matches the font color, ignoring the meaning of the word. Try to work quickly and accurately. (Color figure online)

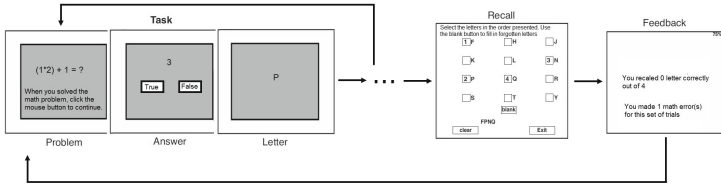


Fig. 3. AOSPAN as dual task (image adapted from [14]): memorize a set of letters in the order presented while simultaneously solving math problems. Trials consist of 3 sets of each set size, with the set sizes ranging from 3–7.

2.3 Subjective Ratings

Paired comparisons of the workload sources were conducted after each task during the training phase as the first part of the computerized version of the NASA-TLX questionnaire method [3]. Subjects were asked to rate the workload sources in 15 pairwise comparisons of NASA-TLX’s six workload dimensions: mental demand, physical demand, temporal demand, performance, effort, frustration.

As part of the main experiment, following each task we then conducted the second part of the NASA-TLX, the ratings of its subscales. Subjects were asked to rate the task for each of the six workload dimensions within a 100-point range with 5-point steps. They indicated their rating by clicking on a 5-point step box with an optical mouse.

2.4 Physiological Measures

In the main experiment during the execution of the tasks we registered the electroencephalogram (EEG), as well as further biosignal data (i.e. heart rate, blood pressure).

EEG. The EEG was captured by 25 electrodes placed at positions according to the 10–20-system and recorded with reference to Cz and at a sample rate of 500 Hz. For signal recording we used an amplifier from BrainProducts GmbH and their BrainRecorder software.

The recorded EEG signal is filtered with a bandpass filter (order 100) between 0.5 and 40 Hz. Subsequently, independent component analysis (ICA) is applied to the signal and the calculated independent components are visually inspected and classified as either an artifact or signal component. The signal components are projected back onto the scalp channels. The artifact-corrected EEG signal is transformed to average reference and cut into segments of 10 s length, overlapping by 5 s. Subsequently, the workload relevant frequency bands (θ : 4–8 Hz, α : 8–12 Hz) are computed over the segments using the Fast Fourier Transformation (FFT).

The newly developed method of dual frequency head maps (DFHM) is based on our analysis of the EEG spectra demonstrating an increase of the frontal theta band power and a decrease of the parietal alpha band power with increasing task difficulty level. Subsequently, labelling of the DFHM based on expert knowledge and classifier training is performed and workload is individually classified in the range of low load, moderate load, and high load [12]. The DFHM are computed for each EEG segment. Hence, the algorithm computes in an interval of 5 s a new workload index value. At the end, we calculate for each person and task three percentage values for the portion of the segments of each sector (LLS: low load segments, MLS: moderate load segments, HLS: high load segments).

Cardiovascular parameters. Blood pressure was recorded continuously by the FMS Finometer Pro device. A finger cuff was placed around the subject’s finger and systolic and diastolic blood pressure as well as the heart rate were detected automatically. The recorded data was processed in the time domain.

2.5 Performance

We concentrated on the analysis of the individual accuracy rates for all three tasks. For AOSPAN, correct responses include the number of sets in which the letters are recalled in correct serial order and correct math problem solving.

2.6 Statistical Analysis

Six ANOVAs were carried out utilizing repeated measures design, one within-subject factor (portion of LLS, portion of HLS, systolic BP, HR, accuracy rate or NASA-TLX). For the factors portion of LLS and HLS, systolic BP, and HR we had five levels (the three tasks and the two rest measurements) while for the factors accuracy rates and NASA-TLX we had only three levels (the three tasks). Differences between the levels were examined and tested with a post-hoc test (Bonferroni).

3 Results

3.1 Subjective Ratings and Performance

Subjective ratings. Figure 4(a) shows the average workload index for the selected tasks 0-back, stroop test, and AOSPAN as representatives of a low, a moderate and a high workload tasks. Workload means changed significantly during the experiment (Greenhouse-Geisser $F(1.94; 102.61) = 92.00, p < 0.001$). Post-hoc analysis revealed significant changes of the subjectively rated mean workload index between all tasks.

Performance. Figure 4(b) shows the average accuracy rates for the selected tasks 0-back, stroop test, and AOSPAN. Accuracy rate means changed significantly during the experiment (Greenhouse-Geisser $F(1.12; 59.21) = 377.15, p < 0.001$). Post-hoc analysis revealed significant changes of the mean accuracy rates between all tasks.

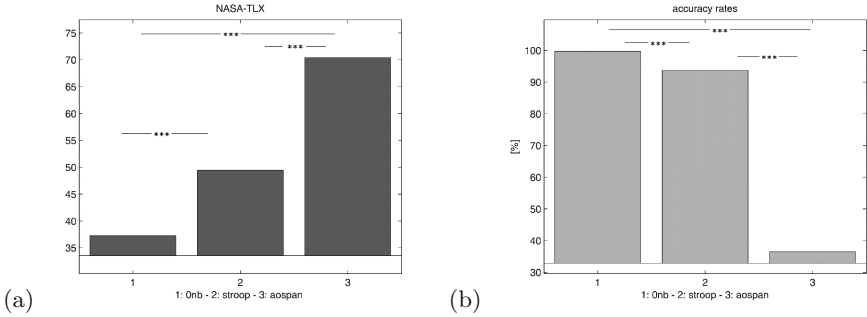


Fig. 4. (a) NASA-TLX computed for 0-back, stroop test, and AOSPAN over 54 subjects. (b) Accuracy rates computed for 0-back, stroop test, and AOSPAN over 54 subjects.

3.2 Physiological Measures

EEG. Analysis of the classified EEG segments demonstrates a proportion increase of the high load segments and a proportion decrease of the low load segments with increasing task difficulty level. Means of LLS and HLS changed significantly during the experiment (Greenhouse-Geisser $F(2.99; 158.02) = 98.51, p < 0.001$; Greenhouse-Geisser $F(2.53; 134.07) = 64.26, p < 0.001$). Results obtained from the assessment of the EEG segments are presented in Fig. 5.

Post-hoc analysis of the proportion of HLS showed that the means were significantly larger as more difficult the tasks were. Significant differences were identified between all tasks as well as between the tasks and the rest measurements. No significant differences could be found between the rest measurement at the begin and at the end of the experiment.

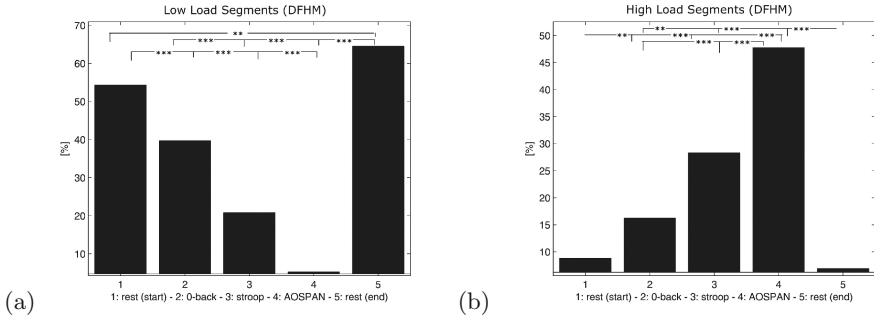


Fig. 5. EEG - proportion of LLS (a) and HLS (b) computed for 0-back, stroop task, and AOSPAN over 54 subjects.

The proportion of LLS revealed significant changes between all measurements.

Cardiovascular Parameters. Both systolic BP and HR differed between the measurements significantly (Greenhouse-Geisser $F(3.33; 176.22) = 31.42, p < 0.001$; Greenhouse-Geisser $F(3.53; 187.22) = 25.13, p < 0.001$).

HR during the rest measurement at the beginning and at the end were, according to post-hoc analysis, lower than during all three tasks. Furthermore, significant changes in HR could be found between the easy 0-back task and the difficult AOSPAN task. No significant differences could be found between the easiest task 0-back and the stroop task, between the stroop task and the most demanding AOSPAN task nor among the two rest measurements.

Systolic BP means were significantly larger during the AOSPAN task than in 0-back and stroop tasks. Additionally, they were significantly larger during the three task measurements and the rest measurements. No significant changes

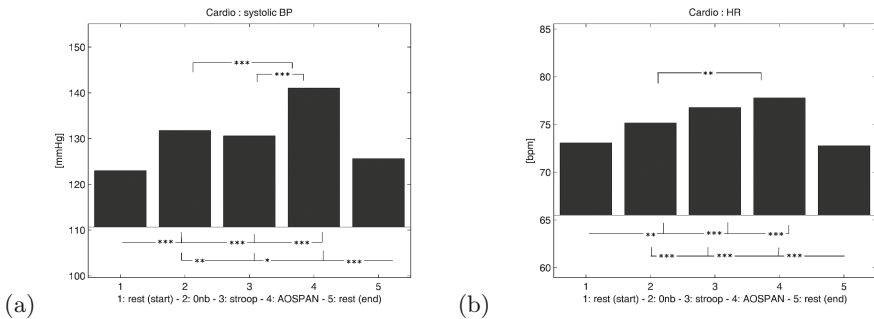


Fig. 6. Systolic BP (a) and HR (b) computed for 0-back, stroop task, and AOSPAN over 54 subjects.

could be found between the two rest measurements nor between the 0-back and the stroop task.

Results of systolic BP and HR are presented in Fig. 6(a) and (b).

4 Discussion

The registration of mental workload by means of the EEG is the central issue addressed by this paper. We induced different levels of mental workload on the basis of a task battery but for the sake of convenience, we concentrated here on the 0-back, stroop and AOSPAN tasks. Cognitive requirements of the first task are quite low and the task can be assumed to be an easy task. The stroop task is more demanding due to higher requirements on the ability for inhibition. It can be classified as a moderate to difficult task but not as challenging as the AOSPAN task. The AOSPAN task demands memory control while dealing with distraction due to the math problem solving. It is a dual-task with high workload requirements.

Subjective ratings derived from the NASA-TLX questionnaire as well as performance data demonstrate significant workload differences between all three tasks. These results emphasize our assumption of gradual workload differences between the tasks. Cardiovascular parameter indicate significant differences between the rest measurements at the begin and the end of the experiment and the three tasks. They also show significant differences between the demanding AOSPAN task and the easy task. However, no significant differences could be observed among the rest measurements nor between 0-back and stroop task. Although there exists a positive tendency, it does not reach the significance level.

The EEG and the frequently observed variability of the θ - and α -band according to attention, fatigue and mental workload, constitute the theoretical background for the new method of the DFHM. The obtained index can be used for neuronal mental state monitoring and ranges between low, moderate, and high workload. Results analyzing the proportions of the HLS and LLS are in concordance with the results expected based on difficulty levels resulting from the requirements of the tasks. The most demanding AOSPAN task contains significantly less segments of low load than the other tasks and the rest measurements. The stroop task includes less LLS than the 0-back task and the rest measurements, while the 0-back task includes less than the rest measurements. Furthermore, the rest measurement at the beginning has less LLS than the rest measurement at the end indicating that the workload at the beginning is a bit higher than at the end of the experiment. All these differences were found to be significant.

In respect of the HLS, the AOSPAN task again shows substantially higher values than all other measurements. Considering also its small proportion of LLS, AOSPAN is a high mental workload task. Stroop task includes significantly higher proportions of HLS than 0-back and the rest measurements. Finally, 0-back task comprises less segments of high load than the rest measurements. Interestingly, there is no significant difference of HLS's proportion among the

rest measurements although there is tendency indicating that the activation of the subjects at the beginning is a bit higher than at the end, similar to the findings from the analysis of the LLS's proportion. Hence, we can assume that the index is able to distinguish between very small gradual differences, in particular when both HLS and LLS are considered simultaneously.

To sum up, our results from the new DFHM method for measuring mental workload are solidly in line both with the accuracy rates and the subjective ratings. Furthermore, they are in concordance with our expectations that result from the known task requirements.

Acknowledgments. We would like to thank Dr Sergei Schapkin, Dr Patrick Gajewski, and Prof Michael Falkenstein for selection of the battery's tasks. We would like to thank Mr Ludger Blanke for technical support during the timing tests for the tasks. In addition, we would like to thank Ms Xenija Weißbecker-Klaus, Mr Robert Sonnenberg, Dr Sergei Schapkin, and Ms Marion Freyer for general task testing and for conducting the laboratory experiments. Furthermore, we would like to thank Ms Marion Exner for daily operational support and our student assistant Jon Scouten for proofreading.

More information about the project where our EEG data were acquired can be found under the following link: <http://www.baua.de/de/Forschung/Forschungsprojekte/f2312.html?nn=2799254>.

References

1. Debitz, U., Gruber, H., Richter, G.: Psychische Gesundheit am Arbeitsplatz. Teil 2: Erkennen, Beurteilen und Verhüten von Fehlbeanspruchungen, vol. 2, 3rd edn. InfoMediaVerlag (2003)
2. Hacker, W., Richter, P.: Psychische Fehlbeanspruchung. Psychische Ermüdung, Monotonie, Stigmung und Stress. (Spezielle Arbeits- und Ingenieurpsychologie in Einzeldarstellungen), vol. 2, 2nd edn. Springer, Heidelberg (1984)
3. Hart, S.G., Staveland, L.E.: Development of the NASA TLX: results of empirical and theoretical research. In: Hancock, P., Meshkati, N. (eds.) Human Mental Workload, pp. 139–183. North Holland, Amsterdam (1988)
4. Kompier, M.A.J., Kristensen, T.S.: Organisational work stress interventions in a theoretical, methodological and practical context. In: Dunham, J. (ed.) Stress in the Workplace: Past, Present and Future, pp. 164–190. Whurr Publishers, London (2001)
5. Landsbergis, P.A., Cahill, J., Schnall, P.: The changing organisation of work and the safety and health of working people: a commentary. *J. Occup. Environ. Med.* **45**(1), 61–72 (2003)
6. May, J.F., Baldwin, C.L.: Driver fatigue: the importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research Part F* **12**, pp. 218–224 (2008). Accessed 03 Nov 2011
7. Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., Wager, T.D.: The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cogn. Psychol.* **41**(1), 49–100 (2000). http://psych.colorado.edu/tor/Papers/Unity-Diversity_Exec_Functions.pdf. Accessed 17 Mar 2014

8. Niosh, N.: The changing organization of work and the safety and health of working people. Technical report 2002-116, National Institute for Occupational Safety and Health (NIOSH) (2002)
9. Parasuraman, R., Molloy, R., Singh, I.L.: Performance consequences of automation induced complacency. *Int. J. Aviat. Psychol.* **3**, 1–23 (1993)
10. Parasuraman, R., Mouloua, M., Molloy, R.: Monitoring automation failures in human machine systems. In: Mouloua, M., Parasuraman, R. (eds.) *Human Performance in Automated Systems: Current Research and Trends*, pp. 45–49. Earlbaum, Hillsdale (1994)
11. Radüntz, T.: Neuronal mental workload registration during execution of cognitive tasks. In: Harris, D. (ed.) *EPCE 2014. LNCS*, vol. 8532, pp. 59–69. Springer, Heidelberg (2014)
12. Radüntz, T.: Kontinuierliche Bewertung psychischer Beanspruchung an informationsintensiven Arbeitsplätzen auf Basis des Elektroenzephalogramms. Ph.D. thesis, Department of Computer Science, Humboldt-Universität zu Berlin, Berlin (2016). <http://edoc.hu-berlin.de/docviews/abstract.php?id=42402>
13. Radüntz, T., Freude, G.: Towards a continuous method for mental workload registration. In: Harris, D. (ed.) *EPCE 2015. LNCS*, vol. 9174, pp. 176–187. Springer, Heidelberg (2015). http://dx.doi.org/10.1007/978-3-319-20373-7_17
14. Unsworth, N., Heitz, R.P., Schrock, J.C., Engle, R.W.: An automated version of the operation span task. *Behav. Res. Methods* **37**(3), 498–505 (2005). http://webs.wofford.edu/boppkl/coursefiles/Thesis/articles/UnsworthHeitz_2005_autoospan.pdf. Accessed 02 Apr 2014