A Task-Independent Workload Classifier for Neuroadaptive Technology: Preliminary Data

Laurens R. Krol[1], Sarah-Christin Freytag[1,2], Markus Fleck[1], Klaus Gramann[1], Thorsten O. Zander[1]

1) Chair of Biological Psychology and Neuroergonomics
Technische Universität Berlin
Berlin, Germany

2) Chair of Human-Machine Systems
Technische Universität Berlin
Berlin, Germany

***Abstract*—Passive brain-computer interfacing allows computer systems direct access to aspects of their user's cognition. In essence, a computer system can gain information about its user without this user needing to explicitly communicate it. Based on this information, human-computer interaction can be made more symmetrical, solving an age-old but still fundamental problem of present-day interaction techniques. For practical real-world application of this technology, it is important that cognitive states can be identified accurately and efficiently. Here we present preliminary data demonstrating it is possible to calibrate a task-independent classifier to identify when a user is under heavy workload across different activities. We used different types of mental arithmetic and even a semantic task. Task-independent classification is an important step towards real-world practical application of this technology.**

Keywords—workload, passive brain-computer interfaces, neuroadaptive technology, universal classifier

I. Introduction

Already in 1990, Edward Tufte recognized that the vast amount of information computers can process and the relatively slow and cumbersome methods to program and control them constitute a *communication bottleneck* and a source of potential error [1]. Today's computers are considerably more powerful, and, despite efforts towards the development of more natural interaction techniques, human-computer interaction (HCI) remains vastly *asymmetrical*. Beyond the explicit communication bottleneck, this is due to a) an information asymmetry: the human operator potentially has access to any and all details concerning the machine's operational state, whereas the machine only has access to the instructions it receives from the operator, supposed to encode their exact intentions; and b) while the human user is capable of dealing with and working around errors and inconsistencies in the communication, the machine is not [2]. Fischer [3] argues that human-computer interaction, just as human-human interaction, can be improved by an increased understanding between the communication partners. This requires that both agents in the interaction have an appropriate amount of understandable information about the other. For the machine to "understand" its user, it would require a more detailed model of its user, fed by more than merely behavioral data and sequential commands. In order to make the machine a true "team player" [4] with independent agency rather than a mere servant, it would need to know about its human partner's underlying intentions, interpretations, opinions...

The communication bottleneck inherent in common present-day interaction techniques prevents users from efficiently communicating such information themselves. However, novel approaches to HCI based on physiological measurements are able to provide input to the machine without placing additional burdens on the users [5]. In particular, *passive brain-computer interfacing* (pBCI; [6]) has our attention as it involves real-time access to the seat of human cognition—the brain. pBCI uses traditional BCI methodology to unobtrusively detect covert aspects of user state, reflecting the user's cognitive and/or affective state. For BCI, brain activity is recorded, usually noninvasively through e.g. electroencephalography (EEG), and automatically interpreted using signal processing and machine learning techniques [7]. Rather than enabling direct, explicit control via the central nervous system by means of coupling voluntary "thought patterns" to specific machine responses, *passive* BCI interprets ongoing measurements of brain activity in order to assess and evaluate the user's ongoing or transient state. As such, passive BCI can provide a continuous stream of *implicit input* coming from the user, but not consciously communicated by them [8,9].

Online assessments of user state can be co-registered against other available information pertaining to current parameters of the interaction. By correlating subjective cognitive/affective user states with external events, a machine can learn how this user interprets these events. In this way, a rich and detailed user model can be generated and continuously updated without placing additional demand on the user. If specific system states consistently lead to an increased assessment of stress levels, overtime, the machine can learn to avoid such states by instead steering towards states that were previously associated with lower levels. Such *neuroadaptive technology* based on implicit input may in certain cases be controlled entirely without conscious interaction.

For pBCI-based neuroadaptive technology to be efficiently used in everyday or professional life, some shortcomings shared by all BCI applications must still be overcome [10]. For example, based on electroencephalography recordings (EEG), the required hardware needs to be made more comfortable to

This is the author's final preprint version of the manuscript published as: Krol, L. R., Freytag, S.-C., Fleck, M., Gramann, K., & Zander, T. O. (2016). A task-independent workload classifier for neuroadaptive technology: Preliminary data. In 2016 IEEE International Conference on Systems Man and Cybernetics (SMC) (pp. 003171–003174). doi: 10.1109/SMC.2016.7844722

wear for longer periods of time, and easier to apply. Because brain activity is heavily context-dependent and changes over time, it is currently still necessary to calibrate separate classifiers for separate tasks and separate sessions, i.e. even for the same person, the machine must re-learn their activity patterns in every new context. This is commonly done by applying supervised machine learning methods to EEG recordings representing known user states [11], "teaching" a classifier to recognize those user states in new data. In this conference contribution, we present work done towards generalizing a classifier, calibrated in a short amount of time, across different tasks.

A much researched cognitive state is the state of high task- or workload. Different levels of load can have a large influence on human wellbeing and performance in almost all tasks (e.g. [12], [13]), making it an important state to be aware of especially in safety-critical environments, but it can also serve as a meaningful indicator in educational [14] or leisure contexts (also see [14] for a review). An advantage of task load in the context of neuroadaptive technology is that it is reflected in the oscillatory power of specific frequency bands. In particular, an increase frontal theta power and a simultaneous decrease in parietal alpha power have been consistently found to accompany increased load levels [15]. Band power estimates can be obtained continuously from ongoing EEG and may thus provide a continuous measure of load, e.g. feeding into a user model that controls adaptive automation [16].

Still, a classifier calibrated to evaluate online levels of workload induced by one task may not necessarily be able to accurately evaluate workload induced, in the same person, by another task. (But see e.g. [17] for a classifier calibrated on the same task in different contexts.) Here, we present preliminary data showing the generalizability of a classifier trained on a mental arithmetic task towards two other mathematical and semantic tasks. Calibrated offline on a subtraction task, in online conditions the classifier could accurately classify subtraction, multiplication, and word finding tasks from the ongoing EEG.

The experimental paradigm furthermore exhibited an example of neuroadaptive behavior. During low load conditions, as assessed automatically during online blocks, visual distraction elements would be shown on the screen to keep participants from getting bored. These would again disappear once levels of high load were detected, in order not to distract the participants from their task. As such, the continuous evaluation of user load was used as implicit input to control the level of distraction that may be needed to keep the participants engaged during the experiment.

II. Methods

A. Participants

Six participants aged 24-27 participated in this study. None had a history of neurological disease and all had normal or corrected-to-normal vision. All gave written consent to participate in this study, which falls under a general-coverage approval of the local ethics committee. Participants were given EUR 10,- per hour or course credit for their participation.

B. Experimental Set-Up and Procedure

Participants were seated in a padded chair approximately one meter away from a computer display. EEG was recorded continuously using 64 active Ag/AgCl electrodes mounted according to the extended 10-20 system on an elastic cap. The signal was sampled at 5000 Hz and amplified using BrainAmp DC amplifiers (Brain Products GmbH, Gilching, Germany). All electrodes were referenced to FCz and the ground electrode was placed at position AFz.

After having read the instructions and given time to acquaint themselves with the task, participants first performed a calibration block of 40 trials: 20 with high load, 20 with low load. During high load, participants were presented with an equation of the form $a - b$, instructing them to count backwards from a in steps of b (i.e. a modified Brown-Peterson distraction technique). a was any integer between 200 and 1200; b ranged from 6 to 19, excluding 10 and 15. During low load conditions, the absence of such an equation instructed participants to relax, with eyes open, calling to mind a specific a freely chosen but consistent scene from memory to focus attention inwards. Both high and low load trials could or could not (50% chance) be accompanied by visual *sparkles*. These were 10 small sparkles wandering smoothly over the screen in random walks governed by perlin noise, providing visual distraction. In low load conditions without sparkles, a crosshair was shown. This 50% chance of sparkles was chosen so as to evenly balance the sparkles between classes, preventing the eye movements they likely induce from being class-specific during calibration. A self-paced break was implemented after every 10 trials. Each trial lasted 10 seconds, for a total of 200 seconds of EEG data per task load class. Based on this data, a classifier was calibrated as described below.

The calibrated classifier was applied online in a second, online block, repeating the previous block. The only difference was that the visual sparkles were now controlled online by the classifier output: a detected high load resulted in a reduction of sparkles, and a detected low load in an increase. The number of sparkles on the screen varied linearly between 0 and 15, representing the online evaluated mean task load in the previous 2 seconds. To this end, the classifier was applied to the last second of data from the ongoing stream at a frequency of 10 Hz. Only classifier output representing an 80% certainty or more for one or the other class was included in the calculation of the mean

A third and final online block consisted of 208 trials in blocks of 26: 13 high, and 13 low load alternately in two new tasks. Instead of the subtraction equation, per block the display showed either a multiplication task, or a scrambled word to be recovered. The multiplication equations consisted of one number between 6 and 19, and one between 21 and 79. The words were randomly scrambled 5- or 7-letter German words (nouns, verbs, adjectives, and adverbs). Participants pressed the enter button when they had found the solution, ending the trial.

Two pilot participants, 1 and 2, did not perform the word task and instead only performed 40 trials of multiplication.

This is the author's final preprint version of the manuscript published as: Krol, L. R., Freytag, S.-C., Fleck, M., Gramann, K., & Zander, T. O. (2016). A task-independent workload classifier for neuroadaptive technology: Preliminary data. In 2016 IEEE International Conference on Systems Man and Cybernetics (SMC) (pp. 003171–003174). doi: 10.1109/SMC.2016.7844722

C. Classification and Analysis

Individual classifiers were calibrated on the data from the first block. The data was divided into consecutive 1-second epochs of high versus low condition data. Filter bank common spatial patterns (FBCSP; [18]) was applied to extract features describing the power in approximately the theta (4-7 Hz) and alpha (8-13) bands using three patterns per band. Linear discriminant analysis (LDA) was used to separate the classes, using a 5-fold nested cross validation with margins of 5. The reported offline accuracy estimations were calculated using the same cross validation scheme.

The calibrated classifier was then applied to 1-second epochs taken from the subtraction, multiplication, and word data, respectively and binary classification was made between "high" and "low" task load. Accuracies reflect the percentage of correct classifications.

Calibration and classification was done using the open-source MATLAB-based toolbox BCILAB (version 1.2) [19].

TABLE I. CLASSIFIER ACCURACY

	Offline accuracy (%)			*Online accuracy (%)*		
Partic.	*TP*	*TN*	*Acc.*	*Subtr.*	*Multi.*	*Words*
1	81	76	78	72	57	
2	60	70	65	72	64	
3	81	85	83	83	89	85
4	53	65	59	56	76	72
5	68	67	67	58	53	57
6	70	69	69	65	75	92
Mean	**69**	**72**	**70**	**68**	**69**	**76**
St.dev.	11	7	09	10	13	15

Table 1. True positive (TP), true negative (TN), and overall accuracy (Acc.) measures for the offline cross-validated accuracy estimations, and online accuracy measures for the three tasks: subtraction (Subtr.), multiplication (Multi.), and words.

TABLE II. ONLINE OUTPUT AND REACTION TIMES

	Mean continuous classifier output						
	Subtraction		*Multiplication*		*Words*	*Mean RT (s)*	
Partic.	*High*	*Low*	*High*	*Low*	*High*	*Multi.*	*Words*
1	1.63	1.29	1.57	1.49		4.97	
2	1.70	1.41	1.53	1.36		5.39	
3	1.74	1.28	1.80	1.55	1.79	5.95	5.12
4	1.67	1.54	1.71	1.57	1.72	5.99	5.41
5	1.41	1.34	1.48	1.48	1.51	6.37	4.42
6	1.52	1.30	1.73	1.63	1.83	7.02	7.29
Mean	**1.61**	**1.36**	**1.64**	**1.51**	**1.71**	**5.95**	**5.56**
St.dev	0.12	0.10	0.13	0.09	0.14	0.72	1.23
t-Test	p = 0.004		p = 0.008		p=0.022		

Table 2. Mean continuous classifier output (scaled between 2 = 100%"high" and 1 = 100% "low") for the three online conditions, and mean reaction times for multiplication (Multi.) and words conditions. t-Test results compare high versus low conditions in the indicated condition. Since multiplication and words were done in one block, both "high" measures were tested against the shared mean "low".

III. RESULTS

Table 1 lists all classification accuracies. The mean estimated offline classification accuracy for the subtraction task over all six participants comes to 70%±9. Mean online accuracies of this classifier are 68%±10 for subtraction data, 69%±13 for multiplication data, and 76%±15 for word data. Chance level is at 50%.

Table 2 lists the mean of *all* classification values produced during the online blocks determining the amount of sparkles, including the values ignored during the online feedback. Values could vary between 1 and 2, with 1.5 representing complete uncertainty between the classes. One-tailed paired-samples t-tests between high and low conditions were significant for all tasks at $\alpha < 0.025$. During online subtraction, the high condition was significantly different from the low condition ($t(6) = 5.15$, $p = 0.004$), as well as for multiplication ($t(6) = 4.19$, $p = 0.008$), and words ($t(6) = 3.22$, $p = 0.022$).

Also listed are the mean reaction times indicating when the participants had solved the multiplication and word tasks. The mean reaction times were 5.9 and 5.6 seconds, respectively.

IV. DISCUSSION

For all but participant number 4, the classifier was able to separate the offline subtraction classes with acceptable accuracy, and for all but number 5 the classifier calibrated on subtraction performed well—in most cases even better—on the other two tasks: multiplication and word finding. Even for number 4, where the classifier was less able to detect the subtraction task offline or online, the multiplication and word tasks were detected with surprising accuracy.

Although visual distraction was balanced between classes during classification, it could be noted that the different tasks themselves may elicit different eye movements. For the subtraction task, only the initial condition was given on the screen: the subsequent steps were all represented internally by the participant. For the words, however, it is likely participants would have continued to look at the stimulus to be reminded of the available letters. If indeed the subtraction task evokes no systematic eye movements, then so too the classifier ought to be independent of eye movement artefacts in the EEG. Visual inspection of the FBCSP topographies indeed revealed no profuse influence of eye movements.

It should also be noted that the low-workload condition was the same for all three high-workload tasks. Subsequent steps will include additional low-workload conditions (e.g. different difficulty levels of the tasks), and will additionally investigate classification accuracies between different high-workload conditions. This will shed further light on the neurophysiological similarities and differences between the different conditions, and the ability to calibrate both task-specific and task-independent workload classifiers.

Having said that, this preliminary data does support to the idea of developing a *task-independent workload classifier* that can be quickly calibrated (i.e. in under 7 minutes, but we have reason to believe that this is not the minimum) and applied to a number of tasks that it was not explicitly trained on.

A task-independent, generalized workload classifier would continue to work reliably even when the human switches tasks, greatly enhancing their applicability in modern working environments.

When properly controlled for confounding variables, successful task-independent application of a classifier points to a general validity of the underlying construct. Although workload remains a construct of varied scientific definitions, it is also a term in common parlance with an intuitive meaning. Perhaps a data-driven approach across different tasks that are intuitively understood to induce "workload" could point to neurophysiological commonalities, i.e. features identifiable by a single classifier, and help elucidate the construct.

Further investigations will include a larger variety of tasks and an evaluation of the role of the neuroadaptively controlled sparkles in maintaining the participants' engagement. We will also look into the generalizability of this classifier between-subjects, i.e. take further steps towards a *universal workload classifier*, both task- and person-independent.

REFERENCES

[1] Tufte, E. R. (1990). Envisioning Information. Cheshire, CT: Graphics Press.

[2] Suchman, L. A. (1987). Plans and situated actions: the problem of human-machine communication. Cambridge, UK: Cambridge University Press.

[3] Fischer, G. (2001). "User modeling in human-computer interaction." User Modeling and User-Adapted Interaction, 11(1-2), 65-86.

[4] Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). "Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity." IEEE Intelligent Systems, 19(6), 91–95.

[5] Bulling, A., & Zander, T. O. (2014). "Cognition-aware computing. IEEE Pervasive Computing", 13(3), 80-83.

[6] 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), 025005.

[7] Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). "Brain-Computer Interfaces for Communication and Control." Clinical Neurophysiology, 113(6), 767–791.

[8] Zander, T. O., Brönstrup, J., Lorenz, R., & Krol, L. R. (2014). "Towards BCI-based Implicit Control in Human-Computer Interaction." In S. H. Fairclough & K. Gilleade (Eds.), Advances in Physiological Computing (pp. 67–90). Berlin, Germany: Springer.

[9] Rötting, M., Zander, T. O., Trösterer, S., & Dzaack, J. (2009). "Implicit interaction in multimodal human-machine systems." In M. C. Schlick (Ed.), Industrial engineering and ergonomics: Visions, concepts, methods and tools (pp. 523–536). Berlin, Heidelberg: Springer.

[10] Zander, T. O. (2012). Utilizing Brain-Computer Interfaces for Human-Machine Systems (Doctoral dissertation, Technische Universität Berlin, Berlin, Germany).

[11] Müller, K.-R., Anderson, C. W., & Birch, G. E. (2003). "Linear and nonlinear methods for brain-computer interfaces." IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11(2), 165–169.

[12] Hockey, G. R. J. (1997). "Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework." Biological Psychology, 45(1-3), 73–93.

[13] Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2016). Engineering psychology and human performance. New York, NY: Routledge.

[14] Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., & Zander, T. O.(2014). "Cognitive state monitoring and the design of adaptive instruction in digital environments: lessons learned from cognitive workload assessment using a passive brain-computer interface approach." Frontiers in Neuroscience, 8, 385.

[15] Klimesch, W. (1999). "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis." Brain Research Reviews, 29(2), 169–195.

[16] Byrne, E. A., & Parasuraman, R. (1996). "Psychophysiology and adaptive automation." Biological Psychology, 42(3), 249–268. (Psychophysiology of Workload)

[17] Mühl, C., Jeunet, C., & Lotte, F. (2015). "EEG-based workload estimation across affective contexts." In A.-M. Brouwer, , T. O. Zander, & J. B. F. van Erp (Eds.), Using neurophysiological signals that reflect cognitive or af f ective state (pp. 227–241). Frontiers Media SA.

[18] Ang, K. K., Chin, Z. Y., Zhang, H., & Guan, C. (2008, June). "Filter bank common spatial pattern (FBCSP) in brain-computer interface." In IEEE International Joint Conference on Neural Networks (IJCNN), 2008 (pp. 2390–2397).

[19] Kothe, C., & Makeig, S. (2013). "BCILAB: a platform for brain-computer interface development." Journal of Neural Engineering, 10(5), 056014.

This is the author's final preprint version of the manuscript published as: Krol, L. R., Freytag, S.-C., Fleck, M., Gramann, K., & Zander, T. O. (2016). A task-independent workload classifier for neuroadaptive technology: Preliminary data. In 2016 IEEE International Conference on Systems Man and Cybernetics (SMC) (pp. 003171–003174). doi: 10.1109/SMC.2016.7844722