34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Towards an Online Detection of Workload in Industrial Work Environments*

Matthias Schultze-Kraft[1], Manfred Gugler[2], Gabriel Curio[2] and Benjamin Blankertz[1]

Abstract— **Many work environments require a high level of continuous and sustained attention of the human operator. This is particularly demanding in monotonous tasks in stimulus-poor environments and can lead to performance drops and deficient production. Here, we explore a novel brain-based approach to deal with this problem: an online monitor of the operator's workload can be used to close the loop of interaction between the human and the machine. Parameters of a, say, manufacturing plant can be adapted to the momentary cognitive state of its operator and thereby enhance the working conditions as well as the production output. The proposed system has its roots in the vast literature of cognitive science in which neural correlates of concepts like mental workload have been studied extensively and in the methods for real-time analysis of brain signals from brain-computer interface research. Our workload monitor was developed in experiments under laboratory conditions with ten participants and subsequently evaluated with six participants during online operation in a real industrial work environment. Our results provide evidence for the potential applicability of the proposed workload monitor in real world environments.**

I. INTRODUCTION

Despite the constant progress of automatization in industrial plants, human operators are indispensable for the monitoring, inspection and correction of specific procedures within automatized processes. Operators are required to maximally adapt to the fast pace of the operating machines, while keeping work performance constantly high. Furthermore, contrary to machines, humans are prone to fatigue and decreased vigilance which leads to errors and thus to suboptimal work flow. A reliable assessment of the instantaneous mental workload of the operator can be used as an adaptive mechanism in human-machine interfaces. In a self-regulatory manner, states of high workload can be compensated by a decrease of working speed, while the detection of low levels of workload may result in a speed up of the operating machines.

Of particular interest for the detection of workload is the use of oscillatory activity in the electroencephalogram (EEG). The human brain is known to exhibit oscillatory activity in particular frequency bands, which have been linked to different functions of the brain [1]. Several studies have tried to make use of these EEG rhythms in order to assess workload during task engagement and show that changes in workload modulate the power of EEG activity in the $\theta-$ (4-8 Hz) and $\alpha-$ (8-12 Hz) frequency bands [2], [3]. The modulation of those quantities can then be used as an index for workload. Such indices can be used in order to achieve adaptive automation and therefore enhance performance in operator tasks [4]. Other studies have shown that the amplitude of event related potentials (ERP) such as the P300 decreases with the number of simultaneous tasks [5].

The goal of this study was to provide guidance for the development of a human-machine interface that autonomously self-regulates the work speed according to the continuous assessment of the operators workload. Another goal was the utilizability of such a system under real-life work environments, such as industrial manufacturing plants. Previous studies have shown that such non-laboratory environments entail numerous potential sources of noise and artifacts on the recorded EEG, thus making high demands on the workload detector system [6]. We therefore chose to approach the development of a detector system in two parts. In the first part of this study we performed experiments under laboratory conditions, mimicking the task demands under real industrial environments. This allowed us to extensively explore the technical and algorithmic possibilities for an EEG-based workload detection, simultaneously disregarding potential noise sources. In a second part, we tested the developed workload detector in an experimental facility of a manufacturing company, thereby fully stressing the system in its designated operation area.

II. METHODS

A. Experimental Setup

1) E1: Laboratory condition: Ten male subjects participated in the laboratory experiment (E1). The subjects were instructed to carry out a task (a catching game - inspired by the designated industrial work task, see next paragraph) on a 21-inch touch screen lying on a table in front of them. Objects randomly tagged with three out of four predefined colors (multiple colors in one object not allowed, the order of colors matters) were falling vertically with equal velocity from random positions at top of the screen, approaching the bottom of the screen. Using their fingers the subjects were able to tag (and untag) a bucket positioned at the bottom of the screen with three out of the four predefined colors and to move it horizontally along the bottom screen border. The task was to catch each falling object with the bucket before it reaches the bottom, ensuring that each time the bucket was tagged with the same colors and in the same

*This work was supported by the BMBF grants 01IB001A,B and 01GQ0850

[1]M. Schultze-Kraft and B. Blankertz are with the Bernstein Focus: Neurotechnology, Berlin, Germany, and the Neurotechnology Group, Berlin Institute of Technology, Berlin, Germany `schultze-kraft at tu-berlin.de`

[2]M. Gugler and G. Curio are with the Charité - University Medicine Berlin, Campus Benjamin Franklin, Berlin, Germany

978-1-4577-1787-1/12/$26.00 ©2012 IEEE

order as the catched object. Catching with wrong colors was considered an error, as well as letting an object hit the bottom of the screen. The falling speed of the objects was constant throughout the experiment, however the interval between the occurrence of each object varied in two different conditions. In the low workload condition (L) the interval between each object was constant and chosen such that subjects were able to accomplish the task with low error rate and reporting the task as demanding but not stressful. In the high workload condition (H) the intervals were shorter and varied randomly, resulting in a reported increased sense of stress and in higher error rates. Each subject performed four blocks of 24 minutes each, each consisting of 16 sub-blocks of 90 seconds each of alternating L and H conditions (the background color shading in the top panel of Fig. 2 illustrates the structure of one block). In order to mimic the conditions of an industrial work place, during the whole experiment a closed loop recording of a real acoustic scenery at an industrial work environment was played through speakers at high volume. The touch screen task was implemented in the open source framework Pyff [7].

2) E2: Industrial environment: Five male subjects and one female subject participated in the real-life experiment (E2). The experiments took place in a test facility of a manufacturing plant, where an automated system transports glass flasks on a conveyor belt. The subjects were standing by to the conveyor system, next to them was a table with a monitor and four plastic bowls, each containing small chips of a certain color. They were instructed to take each flask arriving on the conveyor belt and fill it with three chips of the three colors as displayed on the monitor, put the filled flask back on the conveyor belt behind them, and repeat this procedure with the next incoming flask. An automated system checked if the flask was filled correctly or not filled at all. The experiment consisted of four blocks of 24 minutes each. The first two blocks consisted of 12 sub-blocks of 120 seconds each of alternating L and H conditions. As with the laboratory experiment, in the low workload condition L the speed of the conveyor belt was low (30% of the maximum conveyor speed) such that subjects were able to accomplish the task without making any errors, while in the high workload condition H the speed of the conveyor belt was considerably higher (90% of the maximum conveyor speed), inducing subjectively perceived stress and an error rate above zero. In the last two 24 minute blocks the speed of the conveyor system was changed in either of two ways: 1) Each time a significant positive or negative change of the workload was detected by the now calibrated workload detector system (see II-D) the speed was decreased or increased by 20%, respectively, or 2) when no workload change was detected for a predefined duration of 240 seconds, sudden increases of speed by 40% were automatically enforced in order to test the validity of self-regulation.

B. Data Acquisition

Both in the laboratory and the main experiment, EEG data was recorded at 1000 Hz using BrainAmp amplifiers and 64-channel actiCAP (E1) or 64-channel easyCAP (E2).

C. Data Analysis

1) Class discrimination: After recording, EEG data were downsampled to 200 Hz and epoched into segments of 3 seconds length. Subsequently channels and epochs containing artifacts were removed. The data was subsequently bandpass filtered in the $\delta-$ (1-3 Hz), $\theta-$ (4-7 Hz), $\alpha-$ (8-14 Hz) and $\beta-,\gamma-$ (15-35 Hz) frequency range and the signed r^2 value of class discriminability between conditions L and H was computed for each channel.

2) Feature Extraction and Classification: For feature extraction, EEG data were downsampled, epoched and bandpass filtered as described in the last paragraph. In a next step, using an automated Common Spatial Patterns (CSP) approach [8] optimal spatial filters with respect to class discriminability were computed for each of the four frequency bands. Finally, for each frequency band the logarithm of the mean band power of the two classes were concatenated and served as features for the calibration of a linear discriminant analysis (LDA) classifier with shrinkage of the covariance matrix [9]. In E1 we trained the classifier with the data from three of the four 24-minute blocks and tested on the remaining block. In E2 we trained the classifier on one of the two calibration blocks and tested on the other.

D. The Workload Detector

In order to extract an indicator of the momentary workload of the subject we chose the following approach: We first define a time window T over which we expect notable modulations of the workload to occur. From initial exploratory analysis we chose $T = 45$ sec. For any time point $t > 2T$ during the experiment we then perform a Wilcoxon rank-sum test of the classifier output samples in the windows $[t-2T \quad t-T]$ and $[t-T \quad t]$. If the test's null hypothesis can be rejected with $p < 0.01$ (or $p < 0.05$, alternatively) we assume that those two classifier output samples come from distributions with unequal medians and therefore identify a significant change of workload at time point t. For convenience, we workload detector returns the negative logarithm of the p-value, multiplied by the sign of the difference of samples' medians.

III. RESULTS

We present our results as follows: We first report results from the spectral analysis of EEG data with respect to class discriminability and contamination with motion artifacts, subsequently present offline results of the workload detector and eventually show results from the online application of the detector.

A. Spectral Analysis of EEG Data

Previous studies report that different levels of mental workload are reflected in the modulation of EEG power in the $\theta-$ and $\alpha-$ frequency bands [2], [3], [6]. A corresponding analysis of our data showed only very moderate discriminability power between conditions L and H for

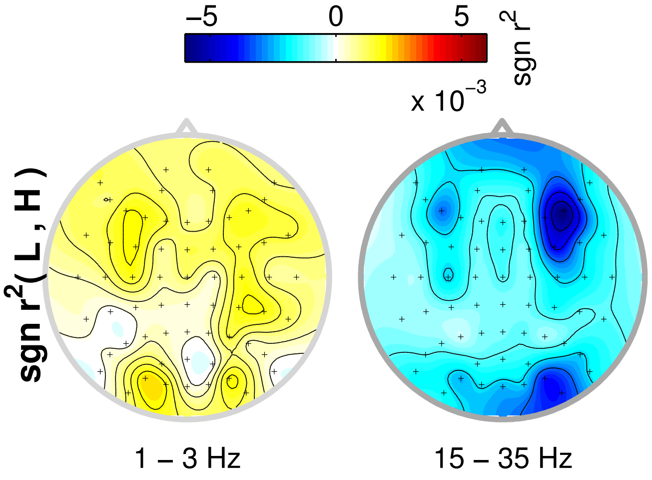

Fig. 1. Scalp topographies of signed r^2 values of the band power of the $\delta-$ and $\beta-,\gamma-$ frequency bands, indicating the discriminability power of both bands between conditions L and H. The data shown are the grand average over all subjects in E1, corrected for multiple comparisons (Bonferroni).

those frequency bands (not shown). However, as Fig. 1 shows, further analysis revealed that the $\delta-$ and $\beta-,\gamma-$ frequency band power yield a higher discriminability of the two workload conditions. The spatial distribution of the r^2 values furthermore suggests that they are not a result of artifacts.

We furthermore point out that EEG power in the $40-45$ Hz band was significantly increased in E2 compared to E1, thus corroborating the assumption that EEG data in E2 was substantially contaminated with muscle activity due to extensive motion of subjects during the task.

B. Workload Detector: Offline Validation

As presented in the methods section, for our workload detection approach we extract features from temporally and spatially filtered EEG data and use these to train a classifier. Fig. 2 shows the output of the classifier for two subjects from E1 and one subject from E2. For subjects 'gaa' and 'laj' we observe a clear task-related modulation of the classifier output, while for subject 'icb' this modulation is rather limited. We further see that the classifier output is not instantaneous with respect to the task condition but rather lags up to 20 seconds.

The red lines in Fig. 2 show the the workload detector output, computed as described in II-D. The time points at which the detector output crosses the $\alpha = 0.01$ from below or $\alpha = -0.01$ from above are considered time points of significant changes of workload. For subjects 'gaa' and 'laj' the the modulation of the classifier output is strong enough to allow a detection of each transition from either the L- to the H-condition and vice versa (with the exception of the very last transition in subject 'laj'). For subject 'icb', we observe that the classifier output modulations are weaker, the detector misses three transitions and produces one false positive (at approx. $t = 920$ sec).

Fig. 3 shows for all subjects of both experiments the

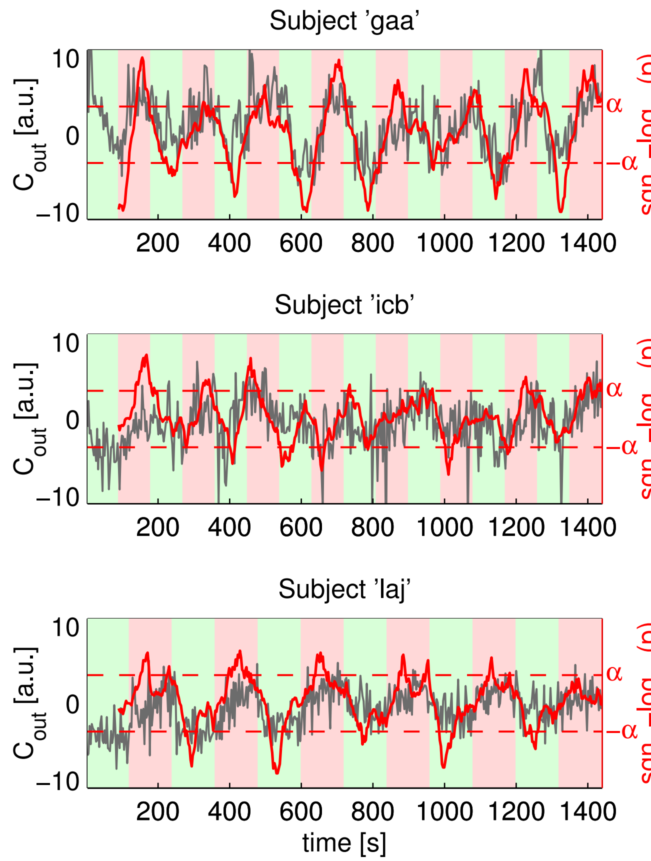

Fig. 2. Classifier output and detection of significant changes in workload for subjects 'gaa' and 'icb' (E1, block 4, training on blocks 1-3) and 'laj' (E2, block 2, training on block 1). Shaded background colors indicate the sub-blocks of conditions L (green) and H (red). The classifier output at each 3-sec epoch is shown in dark gray. Shown in solid red is the output of the workload detector, dashed red lines indicate the significance level $\alpha = 0.01$.

signed r^2 value calculated as the correlation between the classifier output and the class membership. Given that subject 'icb' shows the second worst r^2 value, the C_{out} modulations and workload detector output in Fig. 2 are relatively sound. Furthermore, the comparison of the group average of the r^2 values in E1 (0.25 ± 0.075) and E2 (0.4 ± 0.1) suggests that the performance of the workload detector did not suffer from increased artifact contamination of the EEG in the industrial environment (E2). One might speculate about the reasons of the increase of average in E2 compared to E1. It could have psychological reasons related to the more immersive environment in the real working condition. But it could also have a more technical reason, like the smaller percentage of transition periods in E2 due to the increased duration of L and H blocks.

C. Workload Detector: Online Self-regulation

In E2, after the classifier of the detector had been calibrated using the features extracted from the data of the first two blocks, in the two subsequent blocks we aimed at testing the efficiency of our workload detector as a self-regulating system in an everyday work task. Fig. 4 shows

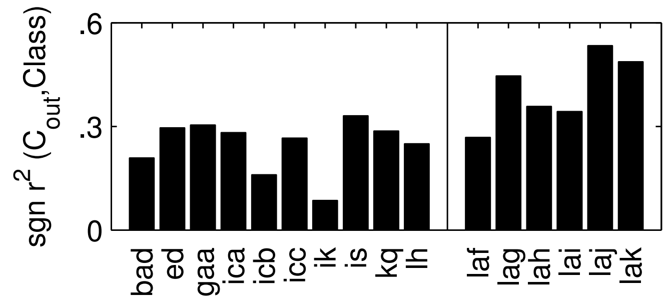

Fig. 3. Signed r^2 values for all subjects in E1 (left group) and E2 (right group), computed as the correlation between the classifier output values of the fourth (E1) or second (E2) block and their respective class membership ($\{-1\}$ for condition L, $\{1\}$ for condition H).

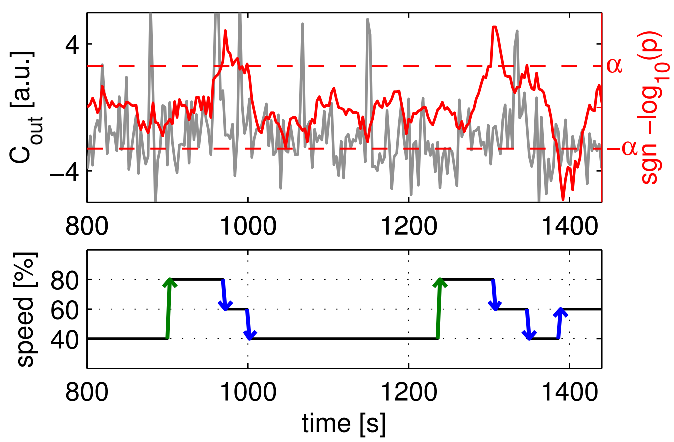

Fig. 4. Self-regulation of task difficulty via detection of changes in workload for subject 'laf'. Top: Classifier output (gray), workload detector output (solid red) and detector significance level $\alpha = 0.05$ (dashed red). Bottom: Relative speed of the conveyor belt (w.r.t. maximum speed) at each time point. Green arrows indicate the time points of externally induced increases of speed, blue arrows indicate time points of classifier-based self-regulation of speed.

an exemplary time segment of the last block of subject 'laf'. At two time points the speed of the conveyor system was automatically increased by four speed levels (green arrows). While the hereby induced changes of workload are not clearly discernable by eye in the classifier output, they were detected by the detector. In both situations, this lead to a subsequent two-fold down-regulation of the speed back to the speed level prior to the enforced change. The very last down-regulation obviously resulted in a drop-off of the workload which was counteracted by a subsequent up-regulation of the speed. In three out of six subjects we frequently observed such self-regulation epochs that eventually ended in a medial speed level (data not shown).

IV. DISCUSSION

Previous studies have aimed at using EEG in order to develop a system that can predict or detect the mental workload of humans [2], [3], and use such detecting systems to achieve adaptive automation and enhance performance in operator tasks [4], [6]. However, the laboratory conditions of most experiments prohibit to draw conclusions as to whether such systems can cope with the difficulties arising from signal contamination due to sources of noise and due to extensive body motion in everyday work situations. In this study we presented an approach for the detection of significant workload changes of subjects operating machines under industrial environments. In a preliminary laboratory study we developed a workload detection system based on a linear classifier that is trained with spectral features extracted from spatially filtered EEG data. After successfully testing the system offline, we conducted experiments in a test facility of a manufacturing plant. Not only did this expose the system to various sources of noise from numerous adjacent industrial machines. The experimental task also required the subjects to constantly move their arms, heads and torsos, thus perpetually contaminating the EEG with motion artifacts. Despite exposure to such motion artifacts and other corrupting sources of noise, we show that our detector system is very robust, yielding workload detection results that are in the same range of goodness as those from the laboratory experiment. We find that our approach provides a means of utilizability as a self-regulation system in industrial work environments, potentially serving as a tool for the optimization of work flow [5]. Ultimately, this goes without saying that the development of any such system must embrace and deal with ethical debates concerning the risk of labor exploitation for the sake of economic optimization.

ACKNOWLEDGMENT

We thank Mr. Walter, Mr. Hoffmann, Ms. Gast and Mr. Welz of Siemens AG for the close cooperation in this project and for the opportunity to perform experiments in the Smart Automation research facility.

REFERENCES

[1] G. Buzsáki and A. Draguhn, Neuronal Oscillations in Cortical Networks. Science 2004;304:5679

[2] A. Holm, K. Lukander, J. Korpela, M. Sallinen, and K.M.I. Müller, Estimating Brain Load from the EEG. TheScientificWorldJOURNAL 2009;9:639-51

[3] A. Gevins and M.E. Smith, Neurophysiological Measures of Cognitive Workload During Human-Computer Interaction. Theoretical Issues in Ergonomics Science 2003;4(1):113-31

[4] G. Wilson and C. Russell, Performance Enhancement in an Uninhabited Air Vehicle Task Using Psychophysiologically Determined Adaptive Aiding. Human Factors: The Journal of the Human Factors and Ergonomics Society 2007;49:1005

[5] R. Parasuraman and G. Wilson, Putting the Brain to Work: Neuroergonomics Past, Present, and Future. Human Factors 2008;50(3):468-74

[6] J. Kohlmorgen, G. Dornhege, M. Braun, B. Blankertz, K.-R. Müller, G. Curio, K. Hagemann, A. Bruns, M. Schrauf, and W.E. Kincses, Improving human performance in a real operating environment through real-time mental workload detection. In: Toward Brain-Computer Interfacing (2007) MIT Press, Cambridge, MA, pp. 409-422

[7] B. Venthur, S. Scholler, J. Williamson, S. Dähne, M. S. Treder, M. T. Kramarek, K.-R. Müller and B. Blankertz, Pyff - A Pythonic Framework for Feedback Applications and Stimulus Presentation in Neuroscience. Frontiers in Neuroscience 2010;4:179

[8] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe and K.-R. Müller, Optimizing Spatial Filters for Robust EEG Single-Trial Analysis. IEEE Signal Process Mag 2008;25(1):41-56

[9] B. Blankertz, S. Lemm, M.S.Treder, S. Haufe, and K.-R Müller. Single-trial analysis and classiffcation of ERP components - a tutorial. Neuroimage 2011;56(2):814-25