Hybrid fNIRS-EEG based discrimination of 5 levels of memory load

Christian Herff[1], Ole Fortmann[1], Chun-Yu Tse[2], Xiaoqin Cheng[3],
Felix Putze[1], Dominic Heger[1] and Tanja Schultz[1]

***Abstract*— In this study, we show that both electroencephalograhy (EEG) and functional Near-Infrared Spectroscopy (fNIRS) can be used to discriminate between 5 levels of memory load. We induce memory load with the memory updating task, which is known to robustly generate memory load and allows us to define 5 different levels of load. Typical experiments only discriminate between low and high workload or up to a maximum of three classes. To the best of our knowledge, the memory updating task has not been used in combination with brain activity measurements before.**

Here, accuracies of up to 93% are achieved for the binary classification between very high and very low workload. On average, two levels of workload could be discriminated with 74% accuracy. Classification between the full five classes yielded 44% accuracy on average. Despite the fact that EEG results consistently outperformed the results obtained with fNIRS, we could show that the feature-level fusion of both modalities increased robustness of classification results. A reliable discrimination between different levels of memory load could be used to adapt user interfaces or present the right amount of information to a learner.

I. INTRODUCTION

The adaptation of user interfaces using physiological measurements has recently gained increased attention [1]. If inherently internal information like emotion, or, as in this study, the memory load, could be robustly determined, user interfaces could react accordingly and thus greatly increase efficiency and naturalness of human-computer interaction.

Brain activity measurements have been shown to allow the identification of these mental states more robustly than other physiological measurements like electrodermal activity, heart-rate related parameters and eye-gaze measurements [2], [3]. The most common modality for such, so called, passive Brain-Computer Interfaces (BCIs) [4] is the electroencephalogram (EEG), which has been used to robustly discriminate between high and low workload [5]. An alternative method for the measuring of brain activity is functional Near-Infrared Spectroscopy (fNIRS), which indirectly measures concentration changes of oxygenated and deoxygenated hemoglobin through the absorption of near-infrared light. While EEG has high temporal resolution, signals at each measurement location are a summation of various brain and non-brain sources, yielding a low spatial resolution. Hemodynamic changes measured by fNIRS take several seconds to be measurable and thus offer a poor temporal resolution, but can be localized to a very small area, giving high spatial resolution. These two modalities thus complement each other in spatial and temporal resolution. Besides classic BCI paradigms, fNIRS has been successfully used, among others, to classify emotion [6], the type of task a user is engaged in [7] and has been applied to more realistic settings such as workload monitoring of air traffic controllers [8]. See [9] for a review of fNIRS for brain imaging in general and [10] for passive BCIs specifically.

The combination of several brain measurement modalities is called hybrid BCI [11]. Hybrid BCIs have been used to identify the type of attention (visual or auditory) participants are engaged in [12] and have been shown to enhance performance in a motor imagery task over single modality interfaces [13].

Different levels of memory load have been evaluated with fNIRS [14], EEG [15] and the combination of both [16] employing the n-back paradigm, with a maximum of 3 levels of memory load. In this study, a larger number of levels was targeted. For this purpose, we chose a variation of the memory updating task. The memory updating task was first presented by Salthouse et al. [17] and adapted by Oberauer et al. [18]. In this task, memory load is induced with a number of digits that have to be remembered and constantly updated with arithmetic operations. By varying the number of digits, the memory demand can be adapted and thus, different levels of memory load induced. This procedure allows for a larger number of distinct load levels [19] than the regularly used n-back paradigm. To the best of our knowledge, the memory updating task has not been used in combination with brain activity measurements. Additionally to the novelty of the task, the number of memory load levels is usually far lower than the five classes investigated here.

II. MATERIAL AND METHODS

A. Experiment

A variation of the memory updating task [17], [18], [19] was used to induce 5 different levels of memory load. During a trial, a participant is shown a row of boxes on the screen. The number of boxes depends on the difficulty level and ranges between one and five in our study. At the beginning of each trial, the boxes are shown empty for one second. The initial digit is then shown for 2.5 seconds in each box. Afterward simple additions and subtractions (between -7 and $+7$) are displayed for 2.5 seconds in a randomly chosen box. Participants have to apply the displayed operation to the currently remembered digit and hence update their

[1]Christian Herff, Ole Fortmann, Felix Putze, Dominic Heger and Tanja Sschultz are with the Cognitive Systems Lab, Karlsruhe Institute of Technology, Karlsruhe, Germany christian.herff@kit.edu

[2]Chun-Yu Tse is with the Department of Psychology & Center for Cognition and Brain Studies, The Chinese University of Hong Kong, China

[3]Xiaoqin Cheng is with the Department of Psychology and LSI Neurobiology/Aging Programme, National University of Singapore, Singapore

remembered number. The number of operations is fixed for each difficulty level to keep the trials at a fixed length of 31s. After the trial, the recall phase starts, in which participants have 3 seconds per box to recall and type in the final result. This recall phase is ignored in EEG and fNIRS analysis. The recall phase is followed by either 15 or 25 seconds of pause, during which a fixation cross was displayed and the participants were asked to relax. We focus on the differences between memory updating levels and ignore the pause trials for this analysis.

Participants were asked to avoid unnecessary motion during trials. We recorded 10 trials per difficulty level and participant, resulting in a total of 50 trials per participant. Figure 1 illustrates the experiment procedure.

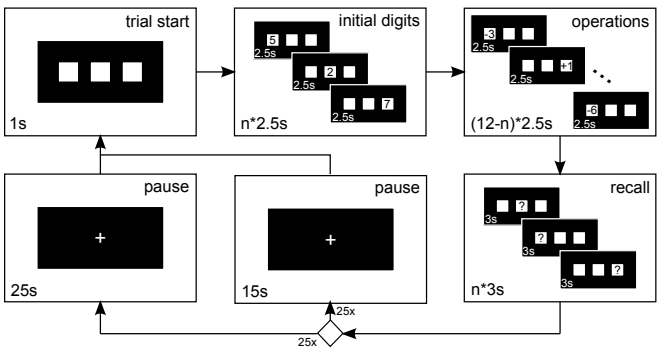


Fig. 1. Experimental design of the memory updating task. The difficulty level n influences the number of boxes per trial. Figure illustrates a 3-box example.

Trials are assigned a label corresponding to the number of boxes shown. In this study, we use the entire 31 seconds of trial data, but exclude pause and recall phases.

B. Data acquisition

During the experiment, we recorded fNIRS signals from 28 sources, emitting near-infrared light and 15 detector optodes located on the forehead to measure hemodynamic activity in the prefrontal cortex. We used a frequency-modulated oximeter (Imagent, ISS Inc.) measuring at two wavelengths (690nm and 830nm). Modulation frequency was set to 110 MHz and the sampling frequency was 19.5 Hz.

In parallel, EEG activity was recorded using three electrodes on the midline at positions Fz, Cz and Pz according to the international 10-20 system. These midline locations were chosen as this region has shown strong activations in previous studies investigating memory load [16]. Additionally, 4 electrodes were placed around the eyes to record electrooculography (EOG). Both modalities are recorded using an ANT amplifier (ANT, Netherlands) and sampled at 256 Hz. All electrodes were referenced to the nose. Ground electrode was placed on the forehead. Source and detector optodes, as well as electrodes are fixed to participants' heads with a rigid custom-made holder. Optode and electrode positions were digitized for each participant individually using an ANT Visor system. The resulting coordinates were used to calculate exact distances for each source and detector pair for the conversion process of raw optical densities to hemoglobin concentration changes.

Figure 2 illustrates optode and electrode positions.

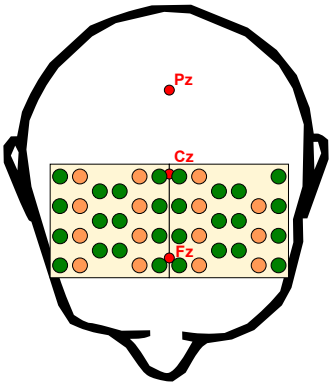


Fig. 2. Optode and electrode montage. Green circles represent sources, orange circles detectors and red circles indicate electrode position.

C. Participants

We recorded 10 healthy participants (3 female) with a mean age of 25.5 years (SD = 0.97). All participants gave written consent and had no history of neurological diseases. They had participated in EEG and fNIRS experiments previously but had no experience with the memory updating task.

D. EEG processing

To reduce eye movement artifacts in the EEG recordings, we applied the EOG regression methods proposed in [20]. We extracted the powerspectrum in 1 Hz wide bins between 4 and 25 Hz using Welch's method. These features were extracted for each of the three electrodes resulting in a 66-dimensional feature space for EEG data. We calculated the mean μ_{train} and standard deviation σ_{train} of each feature in the training set and normalized both training and test sets by subtracting μ_{train} and dividing by σ_{train} (z-normalization).

E. fNIRS processing

We restricted our analysis to the DC component of the measured signal and converted the optical densities to changes in oxygenated and deoxygenated hemoglobin (HbO and HbR) using the HomER package [21]. As HbO and HbR are strongly correlated [22] with responses more pronounced in HbO, we limit our analysis to the HbO signals. As each of the 15 detectors measures the light intensities from each of the 28 sources, a total of 420 channels with different source-detector distances was recorded. To restrict this amount to information bearing channels, and exclude channels with too large source-detector distances, we only consider channels showing a clear pulse artifact, which is expected in clean fNIRS signals. Channels with pulse were identified automatically, if a peak was detected in the log-powerspectrum p frequency band 0.8 to 1.7 Hz:

$$\max\{p(f)|f \in [0.8, 1.7]\} - \text{mean}\{p(f)|f \in [0.8, 1.7]\} > 0.5$$

The log-powerspectrum was extracted with Welch's method and yielded 0.2 Hz wide frequency bins. This procedure reduced the amount of channels from 420 to 13-47 depending on the participant. The data was then lowpass-filtered to attenuate motion artifacts and heart-beat with an elliptic IIR filter with filter-order 6 and cut-off frequency of 0.5 Hz. As a feature, the slope of a line fitted to the data of a trial was extract for each channel using a least-squares approach. The slope has been shown to capture relevant information in single trial fNIRS data in previous studies [14]. This resulted in 13 to 47 features for the fNIRS data. The fNIRS slope features were z-normalized, as described in Section II-D.

F. Evaluation

To evaluate how well 5 classes of memory load can be discriminated, we first investigated the classification of load levels in binary conditions. Each pairing of memory load levels was evaluated for fNIRS and EEG data, resulting in a total of 10 tasks. To assess the possibility of multi-class discrimination, we evaluated the 5-class condition, as well.

Evaluation was done participant-dependently using 10-fold cross-validation. We trained a regularized LDA on the training data, with an optimal shrinkage parameter determined with the analytic methods proposed in [23]. Regularized classifiers have been shown to yield good results in high-dimensional features spaces with small amounts of training samples, as is the case in BCIs with EEG [24] and fNIRS [25], [26]. Multi-class classification was performed using a one-vs-one approach, resulting in 10 classifiers and deciding via majority vote. In addition to the analysis of fNIRS and EEG independently, we evaluated the feature-level fusion of both modalities by combining the 66-dimensional EEG feature space with the 13 to 47 dimensional feature space from fNIRS, resulting in 79 to 113 dimensions for the fused feature space. We denote the fused feature space as FUSION in later section.

III. RESULTS

In the binary classification conditions, high accuracies are achieved with both EEG and fNIRS. As expected, highest accuracies were achieved when discriminating between one-box and 5-box conditions yielding 71% correct on average for fNIRS, 90% for EEG and 93% for FUSION. Accuracies for EEG and FUSION never drop below 58%, while fNIRS only achieves chance accuracies for the discrimination between 1 and 2 box conditions. Classification of one memory load level against another works consistently better for EEG than for fNIRS with a mean over all conditions of 60% accuracy for fNIRS and 74% for EEG. Feature-level FUSION yielded a mean of 74% as well, albeit more results significantly better than chance could be achieved using the feature-level FUSION. Testing for each participant and condition (total of 100 = 10 participants x 10 conditions) individually whether the 10 folds yielded results significantly larger than chance, we found that 26% are significantly better than chance (one-sided t-test, $p < 0.05$) for fNIRS, 61% for EEG and 69% for FUSION.

Results are strongly correlated with the distance between load levels (fNIRS $r = 0.38$, $p < 0.001$; EEG $r = 0.64$, $p < 0.001$; Fusion $r = 0.62$, $p < 0.001$)) meaning that the further the levels are apart, the better the classification. All classification results for the binary conditions can be found in Figure 3 (a).

In addition to the binary conditions, we evaluated how well five levels of memory load could be discriminated 3 (b). Discrimination works significantly better than chance for fNIRS (29%), EEG (42%) and the FUSION approach (44%). Figure 4 illustrates the confusion matrix for the FUSION approach. It is clearly visible that confusions occurred most often with adjacent load levels and that confusions with farther away levels are very rare.

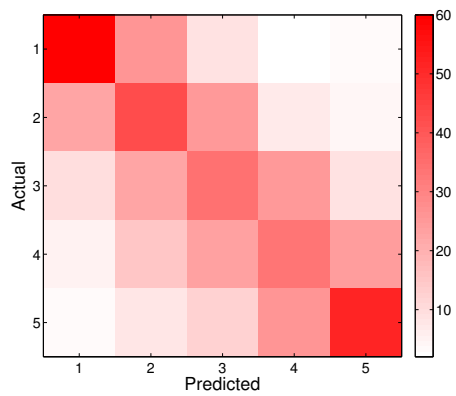


Fig. 4. Confusion matrix in the five class experiment using the fused feature space.

IV. CONCLUSIONS

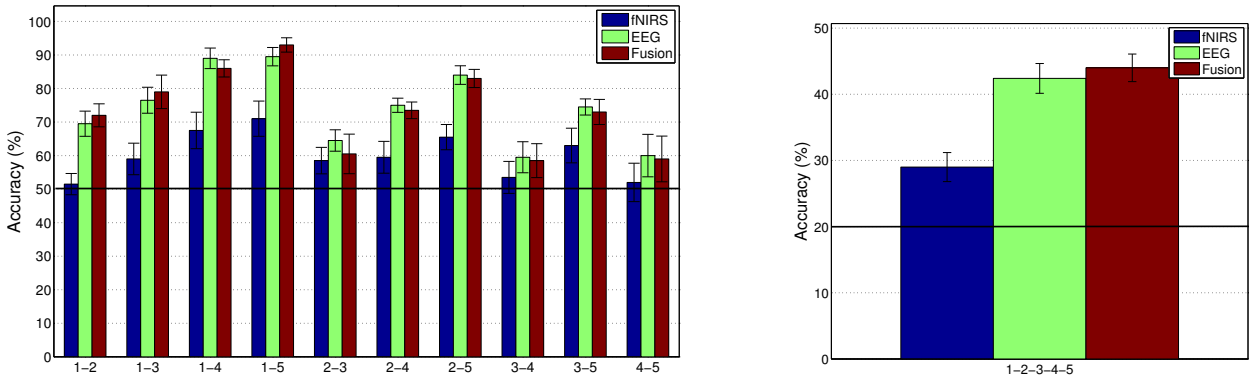
In this paper, we have shown that 5 levels of induced memory load can be robustly discriminated using fNIRS and EEG. This shows the potential for both modalities to be used for more than just the identification of high and low levels of workload or memory load.

A combination of both modalities increased classification results significantly when discriminating between 5 levels of memory load and increased the robustness in binary classification conditions. Even though the combination of both modalities increases robustness of classification, it has to be decided whether this increase justifies the usage of an additional sensor.

Our study was conducted in a lab environment and does thus not face challenges that would be posed by a real-life scenario. Further evaluations are needed to determine whether the window length of 31 seconds can be reduced while maintaining high accuracies to meet the requirements of real-life application.

ACKNOWLEDGMENT

The authors would like to thank Professor Trevor Penney of the National University of Singapore for productive co-operation, providing access to the fNIRS system and useful discussion.



(a) (b)

Fig. 3. Mean classification results for the 10 binary conditions (a) and the five-class condition (b). Whiskers indicate standard errors. Solid lines denotes naive classification rate.

REFERENCES

[1] S. H. Fairclough, "Fundamentals of physiological computing," *Interacting with computers*, vol. 21, no. 1, pp. 133–145, 2009.

[2] J. Frey, C. Mühl, F. Lotte, M. Hachet, *et al.*, "Review of the use of electroencephalography as an evaluation method for human-computer interaction," in *PhyCS 2014-International Conference on Physiological Computing Systems*, 2014.

[3] M. A. Hogervorst, A.-M. Brouwer, and J. B. Van Erp, "Combining and comparing eeg, peripheral physiology and eye-related measures for the assessment of mental workload," *Frontiers in neuroscience*, vol. 8, 2014.

[4] T. O. Zander and C. Kothe, "Towards passive braincomputer interfaces: applying braincomputer interface technology to humanmachine systems in general," *Journal of Neural Engineering*, vol. 8, no. 2, p. 025005, 2011.

[5] C. Kothe and S. Makeig, "Estimation of task workload from eeg data: New and current tools and perspectives," in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 6547–6551.

[6] D. Heger, R. Mutter, C. Herff, F. Putze, and T. Schultz, "Continuous recognition of affective states by functional near infrared spectroscopy signals," in *Affective Computing and Intelligent Interaction*. Springer, 2013, pp. 436–446.

[7] C. Herff, D. Heger, F. Putze, J. Hennrich, O. Fortmann, and T. Schultz, "Classification of mental tasks in the prefrontal cortex using fnirs," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, 2013, pp. 2160–2163.

[8] H. Ayaz, B. Willems, B. Bunce, P. A. Shewokis, K. Izzetoglu, S. Hah, A. Deshmukh, and B. Onaral, "Cognitive workload assessment of air traffic controllers using optical brain imaging sensors," *Advances in understanding human performance: neuroergonomics, human factors design, and special populations*, pp. 21–31, 2010.

[9] M. Ferrari and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fnirs) development and fields of application," *Neuroimage*, vol. 63, no. 2, pp. 921–935, 2012.

[10] M. Strait and M. Scheutz, "What we can and cannot (yet) do with functional near infrared spectroscopy," *Frontiers in Neuroscience*, vol. 8, no. 117, 2014.

[11] G. Pfurtscheller, B. Z. Allison, C. Brunner, G. Bauernfeind, T. Solis-Escalante, R. Scherer, T. O. Zander, G. Mueller-Putz, C. Neuper, and N. Birbaumer, "The hybrid bci," *Frontiers in neuroscience*, vol. 4, 2010.

[12] F. Putze, S. Hesslinger, C.-Y. Tse, Y. Huang, C. Herff, C. Guan, and T. Schultz, "Hybrid fnirs-eeg based classification of auditory and visual perception processes," *Frontiers in Neuroscience*, vol. 8, no. 373, 2014.

[13] S. Fazli, J. Mehnert, J. Steinbrink, G. Curio, A. Villringer, K.-R. Müller, and B. Blankertz, "Enhanced performance by a hybrid nirs–eeg brain computer interface," *Neuroimage*, vol. 59, no. 1, pp. 519–529, 2012.

[14] C. Herff, D. Heger, O. Fortmann, J. Hennrich, F. Putze, and T. Schultz, "Mental workload during n-back task - quantified in the prefrontal cortex using fnirs," *Frontiers in Human Neuroscience*, vol. 7, no. 935, 2014.

[15] A.-M. Brouwer, M. A. Hogervorst, J. B. Van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld, "Estimating workload using eeg spectral power and erps in the n-back task," *Journal of Neural Engineering*, vol. 9, no. 4, p. 045008, 2012.

[16] E. B. Coffey, A.-M. Brouwer, and J. B. van Erp, "Measuring workload using a combination of electroencephalography and near infrared spectroscopy," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 56, no. 1. SAGE Publications, 2012, pp. 1822–1826.

[17] T. A. Salthouse, R. L. Babcock, and R. J. Shaw, "Effects of adult age on structural and operational capacities in working memory." *Psychology and aging*, vol. 6, no. 1, p. 118, 1991.

[18] K. Oberauer, H.-M. Süß, R. Schulze, O. Wilhelm, and W. Wittmann, "Working memory capacityfacets of a cognitive ability construct," *Personality and Individual Differences*, vol. 29, no. 6, pp. 1017–1045, 2000.

[19] S. Lewandowsky, K. Oberauer, L.-X. Yang, and U. K. Ecker, "A working memory test battery for matlab," *Behavior Research Methods*, vol. 42, no. 2, pp. 571–585, 2010.

[20] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of eog artifacts in eeg recordings," *Clinical neurophysiology*, vol. 118, no. 1, pp. 98–104, 2007.

[21] T. J. Huppert, S. G. Diamond, M. A. Franceschini, and D. A. Boas, "Homer: a review of time-series analysis methods for near-infrared spectroscopy of the brain," *Applied optics*, vol. 48, no. 10, pp. D280–D298, 2009.

[22] X. Cui, S. Bray, and A. Reiss, "Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics," *NeuroImage*, vol. 49, no. 4, pp. 3039–46, Feb. 2010.

[23] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of multivariate analysis*, vol. 88, no. 2, pp. 365–411, 2004.

[24] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of erp componentsa tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814–825, 2011.

[25] G. Bauernfeind, D. Steyrl, C. Brunner, and G. R. Muller-Putz, "Single trial classification of fnirs-based brain-computer interface mental arithmetic data: A comparison between different classifiers," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2014, pp. 2004–2007.

[26] D. Heger, C. Herff, and T. Schultz, "Combining feature extraction and classification for fnirs bcis by regularized least squares optimization," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2014, pp. 2012–2015.