Towards User-Aware VR Learning Environments: Combining Brain-Computer Interfaces with Virtual Reality for Mental State Decoding

Katharina Lingelbach*
katharina.lingelbach@iao.fraunhofer.de
Fraunhofer Institute for Industrial Engineering IAO
Stuttgart, Germany

Daniel Diers
daniel.diers@iat.uni-stuttgart.de
Institute of Human Factors and Technology Management, University of Stuttgart
Stuttgart, Germany

Mathias Vukelić
mathias.vukelic@iao.fraunhofer.de
Fraunhofer Institute for Industrial Engineering IAO
Stuttgart, Germany

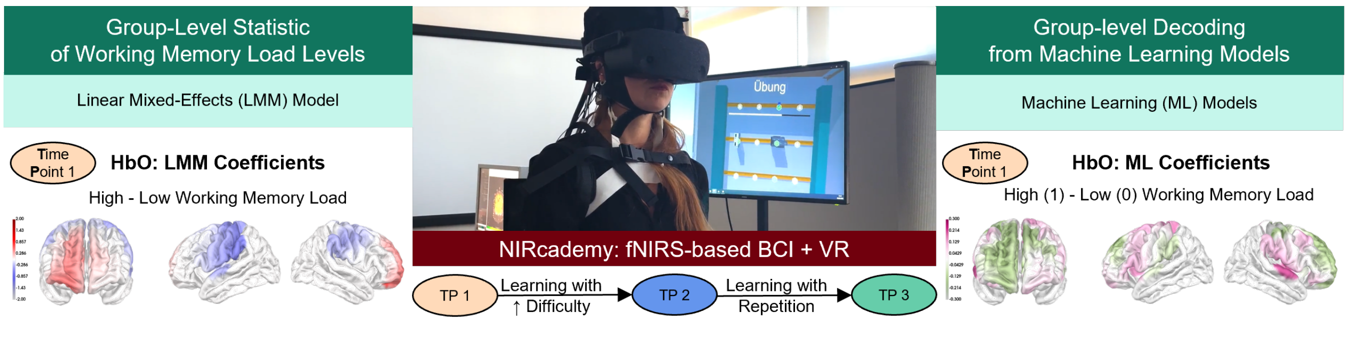

Figure 1: Investigating brain patterns of mental states across multiple virtual reality (VR) learning sessions. Two approaches are introduced, (1) left side: a group-level statistic with a linear mixed-effects (LMM) model and (2) right side: a decoding approach using machine learning (ML). In both approaches, model coefficients were visualized on the cortical surface of a 3D brain image to identify informative patterns distinguishing between low and high working memory load.

ABSTRACT

User-aware adaptive systems can greatly benefit from brain-computer interface (BCI) technologies. BCIs allow continuous monitoring of users' mental states and tailoring of the system to their individual skills and needs. We conducted a feasibility study integrating a BCI using functional near-infrared spectroscopy (fNIRS) into a virtual reality (VR) environment for realistic industrial learning scenarios. Using a fNIRS-based BCI allowed us to a) identify learning progress of individuals based on their working memory load across multiple learning sessions and b) investigate the underlying brain patterns. Our results showed a non-linear relationship between task difficulty and brain responses in the prefrontal cortex (PFC). Finally, we were able to draw four major conclusions regarding architecture components and vital research perspectives, to progress towards a vision of user-aware adaptive system design.

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/ 23/04.
https://doi.org/10.1145/3544549.3585716

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; **Virtual reality**; **Interactive systems and tools**; **Empirical studies in HCI**;

KEYWORDS

Brain-Computer Interface (BCI), Mental State Monitoring, Workload, Working Memory, Learning

ACM Reference Format:
Katharina Lingelbach, Daniel Diers, and Mathias Vukelić. 2023. Towards User-Aware VR Learning Environments: Combining Brain-Computer Interfaces with Virtual Reality for Mental State Decoding. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3544549.3585716

1 INTRODUCTION

Nowadays, with steadily changing skill requirements and the resulting need for re-skilling employees, suitable learning scenarios and environments for training are essential for companies to remain competitive and ensure job satisfaction. Concurrently, virtual reality (VR) has become a revolutionary key technology to create immersive and engaging learning experiences [42]. It not only allows exploring possible actions and visualizing consequences in a secure and cost-effective setting, but also enables for customization of the learning environment (e.g., content, speed, and/or format) to suit the particular skills and needs of the user [22, 42].
Despite the advances in innovative human-computer interaction (HCI) technologies towards taking into account contextual and

environmental settings [10, 25, 36, 39], there remains a lack of knowledge into user preferences, cognitive skills and abilities. To provide an optimal learning experience, user-aware adaptive systems need to be developed that also incorporate the user's current mental state [61, 62]. Insights from educational psychology suggest that an optimal fit between task requirements and learning environment with the user's competencies and needs leads to high self-motivation, engagement, and, therefore, better learning performance [50, 64].

Technologies that promote users in achieving their goals hold potential for significantly improving personalized learning. One key challenge for the development of such user-aware adaptive systems lies in the continuous and robust monitoring of mental states. To unobtrusively monitor a user's mental state, easy-to-use wearable electroencephalography (EEG), or functional near-infrared spectroscopy (fNIRS) devices appear to be well suited [5, 14, 38, 49, 60]. In recent years, there have been significant strides in gaining insight into the brain functions in cognitive processes during driving [19, 45], learning [53, 63], and other outdoor activities [8, 43, 48, 68]. The introduction of passive brain-computer interfaces (BCI) [66] as a new concept for HCI laid the groundwork for combining neurotechnologies with context-aware systems, thus, facilitating the development of user-aware adaptive systems [5, 10, 61]. Passive BCI technologies are continuously monitoring brain processes in order to a) infer the user's current cognitive or emotional states [5, 10, 32] and b) maintain a conceptual representation of the user. EEG is currently the most commonly used technology for integrating passive BCI into mobile VR environments [28, 38, 49]. Only few attempts have been made to utilize fNIRS [21, 47], which is an optical brain imaging technique measuring metabolic changes in the concentrations of local oxygenated (HbO) and deoxygenated hemoglobin (HbR; [15]). When compared to EEG, fNIRS has the advantage of being less susceptible to movement-induced artifacts [21, 40, 47, 54] and providing a high spatial resolution, which allows for the localization of brain regions involved in cognitive functions [52, 57].

In this work, we are introducing a new experimental paradigm in VR that enables us to study users' learning progress in a realistic industrial task. The goal is to explore the potential of utilizing fNIRS signals to detect changes in working memory load across multiple sessions, and gain insights which help customizing fNIRS-based passive BCI for user-adaptive learning systems. We contribute to the existing research by (1) presenting an **architecture called NIRcademy for integrating fNIRS-based passive BCI into a VR learning environment** that enables training for realistic industrial scenarios; and (2) investigating **informative brain activation patterns** with fNIRS that can be used to decode different working memory levels across multiple learning sessions.

2 RELATED WORK

Various cognitive processes are involved during learning, qualifying them as insightful candidates to be monitored. Workload or visuospatial working memory load, that is the ability to process, store, and update information on visual properties and current location of an object [34], is one of the most intensively studied cognitive processes in HCI [5, 40, 47, 55] and of particular interest for learning [40].

In the past, mobile EEG was predominately used to monitor working memory load [2, 12, 55] as well as further cognitive processes in VR scenarios [28, 32]. For instance, for stress detection in offshore environments [1], stress and workload recognition among crew members in maritime simulations [30], and in construction safety training [24]. However, fNIRS is gaining more and more importance for everyday world applications (e.g., [9, 44, 56, 58, 60]). In fNIRS studies, various areas of the prefrontal cortex, especially dorsolateral and ventrolateral parts, are identified as key regions for working memory load [3, 31]. When working memory load is experienced, an increase of the local HbO and a decrease of the HbR concentration [15] can be observed in both regions [21, 35]. Current research utilizing fNIRS-based BCI in VR [21, 47] focuses primarily on the decoding of mental states within a single session. Furthermore, there are only few studies integrating fNIRS-based BCI in naturalistic VR-HCI applications (e.g., in industrial high-risk shutdown maintenance training [51]). Investigation into modulations of the brain activation patterns underlying user's learning progress among multiple sessions is scarce. Additionally, major components necessary for the development of user-aware BCI-based VR learning environments including online decoding of mental states and design concepts to adapt VR learning environments are rarely addressed in the existing literature.

Our feasibility study aims to close some of these gaps by a) introducing a new experimental paradigm for industrial learning scenarios in VR, and b) investigating visuospatial working memory load, state decoding, and the modulation of brain activation patterns across multiple sessions.

3 NIRCADEMY - A NEW EXPERIMENTAL PARADIGM USING FNIRS IN INDUSTRIAL VR LEARNING SCENARIOS

NIRcademy is based on an existing VR learning environment called VRcademy and extends it by integrating fNIRS-based BCI. Using pre-existing 3D data, VRcademy allows developers and end-users to create step-by-step VR training lessons for various industrial tasks (e.g., assembly or disassembly of machines; [6]). Here, we designed a learning scenario from an electrical engineering context. Participants had to learn the order and positioning of electrical components to be installed in a switch cabinet. This naturalistic task, thus, required visuospatial working memory and allowed systematic scaling of task difficulty. Each learning unit was divided into four phases (see Figure 2):

(1) A **baseline phase** lasting 30 sec in which participants fixated a cross-hair in the center of their visual field. Within these 30 seconds, fNIRS HbO and HbR concentrations recovered to resting state levels.
(2) A **training phase** in which participants received step-by-step visual instructions on where to put which components in the switch cabinet. The target component and position were highlighted in each trial. In the first time point, participants had no time limits to memorize each electrical component and follow the instruction. In the second and third time point, time was limited to 15 sec per component in order to further increase task difficulty and avoid ceiling effects in the test phase.

(3) A **memorize phase** lasting 30 sec in which participants fixated a cross-hair in the center of their visual field and were instructed to mentally repeat the learned sequence. The length of the memorize phase was set to be similar to the baseline phase in order to guarantee the recovery of fNIRS HbO and HbR concentrations to resting state levels, and facilitate comparison between a pre- and post-learning phase (this analysis is beyond the scope of this work).
(4) A **test phase** in which participants had to perform the installation of each component within 7 sec based on the previously learned and memorized sequence.

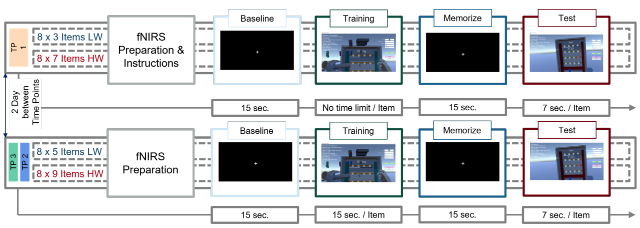

Figure 2: Experimental procedure of the feasibility study with four different task phases and three learning sessions (time point (TP) 1 - 3). Items: to-be-installed electrical components. During the guided training phase, target electrical components and positions were highlighted. LW: low working memory load (blue), HW: high working memory load (red).

To induce different working memory load levels, we manipulated task difficulty by varying the number of components to be memorized across learning sessions. In the first session, participants had to memorize three electrical components in the low and seven components in the high working memory load condition. In the second and third sessions, participants had to memorize five components in the low and nine components in the high load condition (Figure 2). The choice of memorizable items was based on preliminary tests aiming to identify two distinguishable difficulty levels [26, 35, 37] while avoiding under- and overload [35]. Every session comprised 8 learning units per condition and lasted approximately 90 min. After each learning unit, participants rated the perceived effort and frustration using an adapted version of the NASA Task Load Index [18] as well as perceived stress with a slider ranging from 0 (low) to 100 (high).

3.1 Participants

11 volunteers (five female, mean age = 27.9, SD = 3.02, range = 24–34 years) participated in the NIRcademy experiment with nine of the participants completing all three sessions (see Figure 2). They were screened using an online survey for sufficient knowledge of German, intact color vision, and no self-reported drug additions as well as mental, neurological, or cardiovascular diseases. All participants received monetary compensation for participation and signed an informed consent. The study was approved by the local ethics committee ($ID : 827/2020BO1$).

3.2 Hardware and Software Components

Hardware: The integration of a mobile fNIRS system into VR requires some considerations: (1) Infrared light emitted by the VR equipment can interfere with the infrared signal used by fNIRS. (2) there is a trade-off between a well-covered fNIRS sensor placement and stable head-mounted display (HMD) attachment. Preparatory testing for potential interference revealed that the first and second generation Lighthouse base stations commonly found with Valve and HTC Vive ^{TM} HMDs can affect the fNIRS signal quality. Furthermore, Oculus HMDs were evaluated as less suitable due to infrared light-based controller tracking. We finally decided to choose the Windows Mixed Reality device HP Reverb G1 with controllers using visible light for positional tracking. The HP Reverb G1 has a per-eye resolution of 2048 x 2048 pixels, refresh rate of 90 Hz, and smaller head strap compared to later models allowing for a good fNIRS sensor coverage. Since the HP Reverb G1 still features an infrared LED and sensor, we disabled it and configured the HMD without it. Brain activity was recorded using the NIRSport2 System and Aurora software at a sampling rate of 5.8 Hz (NIRx Medical Technologies, LLC). The placement of the sensors was chosen to optimally cover lateral parts of prefrontal cortex which is associated with working memory [3, 31] (Figure 3).

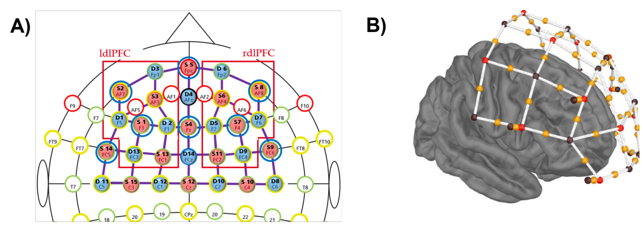

Figure 3: A) Sensor placement including 15 sources (red circles), 14 detectors (blue circles), and 8 short channels (blue ring around source) covering the prefrontal cortex. rdlPFC: right dorsolateral prefrontal cortex; ldlPFC: left dorsolateral prefrontal cortex. B) 3D illustration of the sensor placement.

Software: The VR learning environment was implemented using the Unity game engine. Analogous to the chosen HMD, the program ran at 90 Hz. The experimental software architecture consisted of three interconnected programs synchronized via Lab Streaming Layer [29]:

(1) the experimental paradigm NIRcademy,
(2) the acquisition software Aurora, and
(3) a Python script to orchestrate data streams, save recorded data, and provide an environment for online state decoding.

Additionally, after completion of each trial, the NIRcademy program sent behavioral data as well as meta information to the Python recording software.

3.3 Signal Processing and Machine Learning

Recorded fNIRS signals were analyzed offline using MNE-Python [17] and MNE-NIRS [33]. Following guidelines by Yücel et al. [65], signals were first converted to optical density [17, 33]. Next, bad channels were rejected using the scalp-coupling index as a quality measure for each channel. A threshold of below 0.5 was applied to exclude bad channels from further analyses and a temporal derivative distribution repair was performed to account for baseline shifts and spike artifacts [16]. To obtain HbO and HbR concentration changes (μM), optical density was transformed using the modified Beer-Lambert law [33, 65] and a partial path-length factor of 6 [17]. A fourth-order zero-phase infinite impulse response (IIR) Butterworth band-pass filter was applied with cutoff frequencies of 0.05 and 0.7 Hz and transition bandwidth of 0.02 and 0.2 Hz.

From a theoretical point of view, HbO and HbR concentration changes are proposed to be highly correlated [15, 65], motivating researchers to confine analyses to one chromophore, mostly HbO [59]. We also excluded HbR signals to analyze HbO in depth as a first step.

Group-level Statistic of Working Memory Load Levels: For the comparison of working memory load levels during the test phase, we calculated forward models [20] from evoked hemodynamic signals using generalized linear models (GLM) [59, 65]. For the hemodynamic response function estimation, we used a predefined canonical statistical parametric map hemodynamic response function generated from a linear combination of two Gamma functions [33, 59]. A third-order polynomial drift as well as signals from short channels were included as GLM regressors in order to separate systemic signals from task-related brain activity [65]. Thus, individual estimates of evoked hemodynamic brain activity were obtained for each experimental condition, channel, and participant. In the next step, the contrast *high* − *low* load was calculated, and group-level statistics using a linear mixed-effects model (LMM) were performed [33] using the R packages lme4 [4].

For the LMM statistics, missing values and outliers exceeding 1.5 of the interquartile range were removed and coefficients z-standardized. The LMM included fixed effects of condition, source-detector-pairs, and chromophore (HbO or HbR) while accounting for participant as random effect [33]. Computed estimates were determined as significant based on the Bonferroni-corrected confidence interval using the number of channels to account for multiple comparisons. Significant Bonferroni-corrected coefficients were projected on the cortical surface of a 3D brain image.

Group-level Decoding from Machine Learning Models: We used a linear multivariate classifier and calculate backward models [20] to discriminate the different load levels. Linear models like a linear discriminant analysis (LDA) allow for an interpretation of model coefficients and comparison to the forward model [20]. The feature set comprised the average HbO concentration changes per channel from epochs of 4 sec during the test phase. We rejected epochs exceeding 80 μM in their peak-to-peak value. We performed the decoding using on average within one participant $TP1$ $M = 75.00$, $(SD = 15.00)$ epochs in the first session, $TP2$ $M = 108.00$, $(SD = 13.97)$ epochs in the second session, and $TP3$ $M = 111.56$, $(SD = 1.33)$ epochs in the third session. 2651 epochs could be extracted from the signals in total (aggregated across participants and sessions).

A LDA was applied participant-wise in a nested 5-fold cross-validation with 20 repetitions (implemented using [41]). Features were z-standardized and the hyperparameter *solver* was optimized using a grid search. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC-ROC). Average performance and its 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ confidence interval were estimated using non-parametric bootstrapping (5000 iterations) over the cross-validation folds of one participant and session. The average performance and confidence intervals were then compared to an empirical chance level (i.e, a dummy classifier, [41]). A dummy classifier provides an empirical chance level by generating predictions according to the class distribution in the training set. The dummy classifier was trained participant- and session-wise. However, its mean and CIs were estimated via bootstrapping using the cross-validation folds of all participants and sessions. The classifier's feature coefficients were extracted and averaged among participants, then projected onto a 3D brain image's cortical surface for comparison to the LMM coefficients.

4 RESULTS

Group-level Statistic of Working Memory Load Levels: Results from the group-level statistic (Figure 4, A) revealed a strong involvement of right frontopolar regions as well as the dorsolateral prefrontal cortex in the high compared to the low load condition during the first session (Figure 4, A, upper row, left). Additionally, HbO concentration was reduced in the left and partially right motor cortex. Especially in areas associated with right hand and facial muscle activity. Interestingly, strong involvement of the right prefrontal cortex observed during the first session was not found during the second and third session (Figure 4, A, middle and lower row, left). There were only marginal differences in the involvement of the dorsolateral prefrontal cortex between the two load conditions in the two latter session (Figure 4, A, middle and lower row). In the second session, posterior parts of the right dorsolateral prefrontal cortex including areas of the premotor cortex were even less involved during high compared to low working memory, as reflected in lower local HbO concentration during the high load condition (Figure 4, A, middle row, right). Although task difficulty was kept stable in the second and third session, we observed a modulation in the brain activation patterns across sessions (Figure 4, A, middle and lower row). In the second session, high working memory load increased HbO concentration in speech-related regions of the left hemisphere, including the Broca area (Figure 4, A, middle row, middle). For the third session we found the opposite, that is decreased activation during high load in these areas (Figure 4, A, lower row, middle). Hand and facial-related regions of the right motor cortex were more activated during high compared to low load in the third session (Figure 4, A, lower row, right).

Group-level Decoding from Machine Learning Models: In the backward models, we observed decreased LDA coefficients in right frontopolar regions and the dorsolateral prefrontal cortex signifying that high HbO concentration was attributed with Class 0 (green), that is low load, during the first session (Figure 4, B, upper row, right). Recruitment of posterior parts of the left dorsolateral prefrontal cortex, bilateral anterior parts of the premotor cortex, as well as the right Broca area, was informative for predicting Class 1 (pink), that is high load (Figure 4, B, upper row, middle and right). In the second session, activation of the right dorsolateral prefrontal cortex, motor cortex as well as left ventrolateral prefrontal cortex increased probability of predicting low load (Figure 4, B, middle row, green). Here, we observed a strong influence of the left motor cortex indexing hand regions (Figure 4, B, middle row, middle). In the third session, recruitment of left frontopolar regions was associated with high load, while recruitment of right ventrolateral and frontopolar regions as well as areas related to mouth and jaw muscles was predictive for low working memory load (Figure 4, B, lower row).

The average decoding performance was for most participants and learning sessions significantly above the empirical chance level estimated at 54.93% (95%*CI* [54.35; 55.52]; Figure 4, C). In the first session, we observed best classification results with the highest performance for Participant 9 at 72.95% (95%*CI* [71.05; 74.83]) and an average performance across participants of 64.77% (95%*CI* [64.14; 65.42]). Working memory load could not be reliably monitored for Participants 1 and 6. In the second session, we observed an overall drop in classification performance. Although still significant, average performance was only at 58.62% (95%*CI* [58.0; 59.22]). For Participants 3 and 4, the classifier failed to reliably predict the current state. In the final learning session, performance increased slightly with an average of 60.47% (95%*CI* [59.92; 61.02]) and below chance level predictions for Participants 1 and 9.

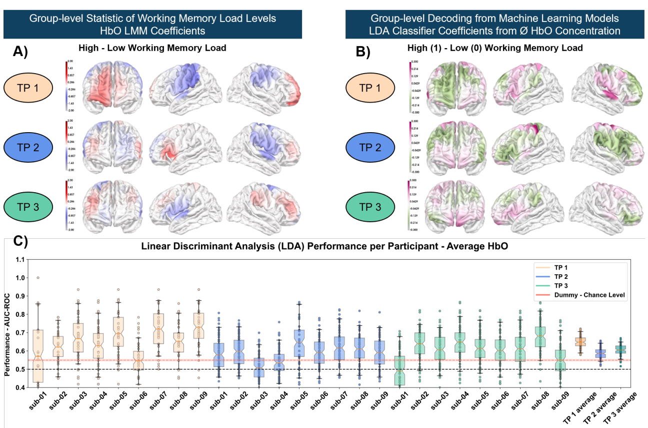

Figure 4: Results per VR learning session of the group-level A) statistic using a linear mixed-effects model (LMM) and B) decoding from machine learning models. Significant LMM and machine learning coefficients were visualized from three perspectives. Left) rostral perspective showing the brain from the front, middle) lateral perspective showing the left hemisphere, and right) showing the right hemisphere. Class 0: low working memory load; Class 1: high working memory load. TP: Time point. C) LDA classifier performance per participant as well as grand averages. Notches of the box plots: upper and lower boundary of the mean's 95 % confidence interval (CI). Box: 50% of the distribution. Whiskers: 5th and 95th quantile of the distribution. Solid orange line: bootstrapped mean. Dashed black line: theoretical chance level of 0.5. Dashed red line: empirical chance level.

Behavioral and Subjective Measures: Behavioral performance (i.e., response accuracy among one block in percentage) and subjective measures (NASA TLX effort, frustration, and stress) were analyzed using non-parametric bootstrapping and confidence intervals [11] of the participant-wise contrast *high - low* working memory load (Figure 5). There were clear behavioral differences in all three sessions with significantly lower response accuracy during high load (first session: −16.77%, 95%*CI* [−27.39; −7.77]; second session: −10.93%, 95%*CI* [−18.55; −4.03]; third session: −7.05%, 95%*CI* [−12.14; −2.14]; Figure 5, B). Regarding the subjective measures, we observed only in the last session a difference for the contrast *high - low*, with high working memory load associated with increased effort (7.01, 95%*CI* [0.67; 12.81]), frustration (8.03, 95%*CI* [1.20; 14.36]), and stress (11.21, 95%*CI* [1.99; 19.86]; Figure 5, B).

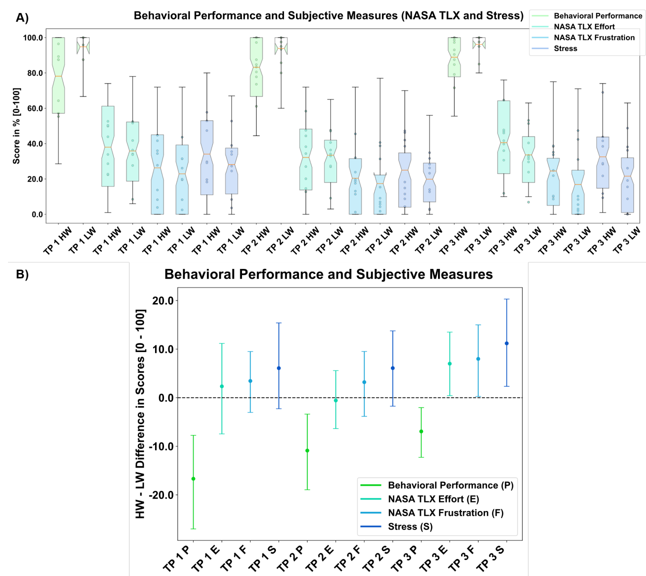

Figure 5: Results of the Behavioral Performance, NASA TLX and Stress Scales. A) Bootstrapped mean and its 95% confidence interval of the two conditions per session and scale. Dots represent single scores of the participants. The solid orange line represents the bootstrapped mean. Notches of the box plots represent the upper and lower boundary of the mean's 95 % confidence interval, the box visualizes 50% of the distribution and whiskers the 5th and 95th quantile of the distribution. B) Bootstrapped mean and its 95% confidence interval for the *high - low* contrast of the behavioral performance (P), NASA TLX effort (E), NASA TLX frustration (F), and stress (S) per learning session.

5 DISCUSSION

5.1 Brain Activation Patterns and Learning Across Multiple Learning Sessions

By investigating brain activation patterns and their dynamics across multiple learning sessions in VR, we gained insights that are of great importance for the development of fNIRS-based mental state monitoring.

A major finding is the complex non-linear relationship between task difficulty level and hemodynamic response in the prefrontal cortex (PFC), which is in line with previous studies [13, 35]. In the first session, where overall task difficulty was rather low, right dorsolateral prefrontal cortex activation was strongly related to working memory load. This area distinguished the two working memory load conditions not only in the group-level statistics (Figure 4), but also machine learning decoding (Figure 4). Referring to the attenuation hypothesis, participants could perform both load levels well without approaching any state of overload possibly decreasing prefrontal cortex activity [13, 35].

Due to our longitudinal approach, we could investigate effects of workload transitions by increasing the overall task difficulty. The nonlinear relationship between PFC activation and load level caused effects of load manipulation to diminish. Although steps between difficulty levels were the same as in the first session, we saw less difference in the prefrontal cortex involvement between high and low load in the second and third session (Figure 4, A and B, middle and lower row). According to McKendrick and Harwood [35] suggesting a curvilinear relationship, we likely encountered a depression (low load with 3 items) to peak (high load with 7 items) phase in the first session (Figure 4, A and B, upper row). Thereby, we induced the largest difference in the hemodynamic response of the prefrontal cortex. By shifting the difficulty level (onset) in the second learning session, this difference dissolved revealing similar dorsolateral prefrontal cortex hemodynamic responses in both load levels. Interestingly, we observed an increase in classification performance and slightly stronger right dorsolateral prefrontal cortex involvement during high load from the second to the third session in the group-level statistic (Figure 4, A, lower row). Hence, the relationship between prefrontal cortex response and load level appears to be dynamically modulated by repetition (learning effects). However, the increasing difference between the two load conditions in the third session was not predictive enough to be used by the machine learning classifier in the decoding (Figure 4, B, lower row).

Our results indicated that hemodynamic responses in hand- and face-related areas of the motor cortex (Figure 4, A and B, upper row) as well as language-related areas (Figure 4, A and B, middle and lower row) were partially used to discriminate load levels. Based on the latter observation, the potential of recognizing mnemonic strategies (e.g., abstract-spatial or categorical-linguistic) and using this knowledge could be explored by assessing the associated learning success and presenting it explicitly to the user.

When comparing the forward and backward modelling results, the decoding models used smaller regions and more distributed patterns for the classification. While activation in some areas was interpreted consistently across both approaches (e.g., the right dorsolateral and premotor cortex in the second session; Figure 4, A and B, middle row, right), activation in other areas was interpreted oppositely (e.g., the right dorsolateral cortex in the first session; Figure 4, A and B, upper row, left). A possible explanation is that the approaches differed in the modelling of inter-individual differences. While decoding models were learned participant-wise, one group-level statistic model was estimated for all participants including inter-individual differences as a random-effect factor. The patterns of classifier coefficients might differ from the here observed when a) visualized participant-wise instead of averaged over participants, or b) one model would be trained for all participants. Since the sample size of the feasibility study was rather small, including further participants could increase consistency between the two approaches and generalizability of the identified patterns.

The potential of using brain activation patterns to decipher the current workload is particularly evident when the statistical results of subjective measures are considered. Retrospective questioning of participants did not significantly differentiate conditions until the third session, and, therefore, was not sensitive enough to detect subtle differences in the first two sessions. Although behavioral performance differed significantly between conditions in all three sessions and is, thus, suitable for detecting the current load levels it has the disadvantage that a number of errors must first occur before mental states can be decoded and the system adjusted, accordingly.

5.2 Towards User-Aware Adaptive Systems for Personalized Learning

In this work, we have introduced NIRcademy, which is a novel VR environment for applied learning scenarios allowing for continuous mental state decoding with fNIRS-based BCI. The design of robust user models including individual skills and needs as well as the current mental state is one key challenge in the development of user-aware adaptive systems. Our results revealed informative brain patterns for working memory load detection. They, further, emphasized the importance for machine learning features able to capture non-linear relationships between conditions at critical transition points (i.e., when high load evolves into overload). These comprehensive insights help to accurately predict mental states with machine learning [7, 27, 59]. In the future, we aim to replicate the findings using HbR signals, transfer our findings into a real-time passive BCI analysis for continuous brain monitoring and extend NIRcademy to further learning scenarios. Attractive use case scenarios are, among others, trainings for industrial safety-critical infrastructures (e.g., [51]), robot-assisted language learning [46], multi-agent adaptive human-robot collaborations [23], and robot-assisted medical interventions [67].

Four conclusions can be drawn from our study: 1) NIRcademy reveals great potential for future user-aware adaptive learning systems due to its modular hardware and software components, which is important for flexible user-tailored adaptation steps. In addition, adaptations in learning and training scenarios are less time-critical compared to a real-world high-risk HCI scenarios which facilitates the implementation of online state monitoring and relaxes requirements for communication speed and reactivity. 2) More systematic studies are required to investigate and evaluate the effectiveness and user acceptance of fNIRS-based BCI and VR technologies when designing new learning scenarios. This includes adaptation strategies for the learning environment (manipulating task difficulty via speed, assistive mode options, or complexity) and communication between components (frequency of adaptions and feedback, bi-directional adjustment options enabling users to reverse/chance adaption steps, as well as transparency reports regarding the learned user model). 3) Current mobile fNIRS systems are not yet optimally designed for sustained use in everyday applications [60]. They require further hardware development to guarantee comfortable, painless, and effortless usage (e.g., by integrating cushioned sensors directly into the HMD or a textile-like cap). 4) Longitudinal studies are inevitable to design reliable computational user models driven by real-time fNIRS-based BCI technologies. Thereby, meta-information (e.g., gender, age, or individual skills and preference) as well as additional information sources (e.g., behavioral and subjective measures) can be further integrated.

Through our complimentary and diligent interdisciplinary approach, we push the envelope of research in user-aware neuro-adaptive VR environments to create efficient and effective real-world industrial learning applications of the future.

ACKNOWLEDGMENTS

The research was supported by the Ministry of Economic Affairs, Labour, and Tourism Baden-Wuerttemberg within the project »KI-Fortschrittszentrum Lernende Systeme und Kognitive Robotik« and Federal Ministry of Economic Affairs and Climate Action [BMWi, NIRcademy, ZF4509902BA9].

REFERENCES

[1] Rumaisa Abu Hasan, Shahida Sulaiman, Nur Nabila Ashykin, Mohd Nasir Abdullah, Yasir Hafeez, and Syed Saad Azhar Ali. 2021. Workplace mental state monitoring during VR-based training for offshore environment. *Sensors* 21, 14 (2021), 4885.

[2] P Aricò, G Borghini, G Di Flumeri, A Colosimo, S Pozzi, and F Babiloni. 2016. A passive brain–computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks. *Progress in brain research* 228 (2016), 295–328.

[3] Hasan Ayaz, Patricia A Shewokis, Scott Bunce, Kurtulus Izzetoglu, Ben Willems, and Banu Onaral. 2012. Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 1 (2012), 36–47.

[4] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01

[5] Benjamin Blankertz, Laura Acqualagna, Sven Dähne, Stefan Haufe, Matthias Schultze-Kraft, Irene Sturm, Marija Ušćumlic, Markus A Wenzel, Gabriel Curio, and Klaus-Robert Müller. 2016. The Berlin brain-computer interface: progress beyond communication and control. *Frontiers in neuroscience* 10 (2016), 530.

[6] Matthias Bues, Tobias Schultze, and Benjamin Wingert. 2018. Konzeption und Implementierung einer VR-Lernumgebung für technische Dienstleistungen. In *Digitalisierung in der Aus- und Weiterbildung: Virtual und Augmented Reality für Industrie 4.0*, Oliver Thomas, Dirk Metzger, and Helmut Niegemann (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 113–123. https://doi.org/10.1007/978-3-662-56551-3_8

[7] Thomas Carlson, Erin Goddard, David M Kaplan, Colin Klein, and J Brendan Ritchie. 2018. Ghosts in machine learning for cognitive neuroscience: Moving from data to theory. *NeuroImage* 180 (2018), 88–100.

[8] Guy Cheron, Géraldine Petit, Julian Cheron, Axelle Leroy, Anita Cebolla, Carlos Cevallos, Mathieu Petieau, Thomas Hoellinger, David Zarka, Anne-Marie Clarinval, et al. 2016. Brain oscillations in sport: toward EEG biomarkers of performance. *Frontiers in psychology* 7 (2016), 246.

[9] Jong-Kwan Choi, Jae-Myoung Kim, Gunpil Hwang, Jaehyeok Yang, Min-Gyu Choi, and Hyeon-Min Bae. 2016. Time-divided spread-spectrum code-based 400 fW-detectable multichannel fNIRS IC for portable functional brain imaging. *IEEE Journal of Solid-State Circuits* 51, 2 (2016), 484–495.

[10] Caterina Cinel, Davide Valeriani, and Riccardo Poli. 2019. Neurotechnologies for human cognitive augmentation: current state of the art and future prospects. *Frontiers in human neuroscience* 13 (2019), 13.

[11] Geoff Cumming and Sue Finch. 2005. Inference by eye: confidence intervals and how to read pictures of data. *American psychologist* 60, 2 (2005), 170.

[12] Frédéric Dehais, Alban Duprès, Sarah Blum, Nicolas Drougard, Sébastien Scannella, Raphaëlle N Roy, and Fabien Lotte. 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 (2019), 1324.

[13] Gautier Durantin, J-F Gagnon, Sébastien Tremblay, and Frédéric Dehais. 2014. Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural brain research* 259 (2014), 16–23.

[14] Stephen H Fairclough. 2009. Fundamentals of physiological computing. *Interacting with computers* 21, 1-2 (2009), 133–145.

[15] Marco Ferrari and Valentina Quaresima. 2012. A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *Neuroimage* 63, 2 (2012), 921–935.

[16] Frank A Fishburn, Ruth S Ludlum, Chandan J Vaidya, and Andrei V Medvedev. 2019. Temporal derivative distribution repair (TDDR): a motion correction method for fNIRS. *Neuroimage* 184 (2019), 171–179.

[17] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen. 2013. MEG and EEG Data Analysis with MNE-Python. *Frontiers in Neuroscience* 7, 267 (2013), 1–13. https://doi.org/10.3389/fnins.2013.00267

[18] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, Amsterdam, Netherlands, 139–183.

[19] Stefan Haufe, Jeong-Woo Kim, Il-Hwa Kim, Andreas Sonnleitner, Michael Schrauf, Gabriel Curio, and Benjamin Blankertz. 2014. Electrophysiology-based detection of emergency braking intention in real-world driving. *Journal of neural engineering* 11, 5 (2014), 056011.

[20] Stefan Haufe, Frank Meinecke, Kai Görgen, Sven Dähne, John-Dylan Haynes, Benjamin Blankertz, and Felix Bießmann. 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87 (2014), 96–110.

[21] Christian Herff, Dominic Heger, Ole Fortmann, Johannes Hennrich, Felix Putze, and Tanja Schultz. 2014. Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS. *Frontiers in human neuroscience* 7 (2014), 935.

[22] Valentin Holzwarth, Johannes Schneider, Joshua Handali, Joy Gisler, Christian Hirt, Andreas Kunz, and Jan vom Brocke. 2021. Towards estimating affective states in virtual reality based on behavioral data. *Virtual reality* 25, 4 (2021), 1139–1152.

[23] Alicia Howell-Munson, Emily Doherty, Peter Gavriel, Claire Nicolas, Adam Norton, Rodica Neamtu, Holly Yanco, Yi-Ning Wu, and Erin T Solovey. 2022. Towards Brain Metrics for Improving Multi-Agent Adaptive Human-Robot Collaboration: A Preliminary Study. In *2022 Symposium on Human-Computer Interaction for Work*. Association for Computing Machinery, New York, NY, United States, 1–10.

[24] Dongjin Huang, Xianglong Wang, Jinhua Liu, Jinyao Li, and Wen Tang. 2022. Virtual reality safety training using deep EEG-net and physiology data. *The visual computer* 38, 4 (2022), 1195–1207.

[25] Nikolaos Kaklanis, Pradipta Biswas, Yehya Mohamad, Mari Feli Gonzalez, Matthias Peissner, Patrick Langdon, Dimitrios Tzovaras, and Christopher Jung. 2016. Towards standardisation of user models for simulation and adaptation purposes. *Universal Access in the Information Society* 15 (2016), 21–48.

[26] Roy PC Kessels, Martine JE Van Zandvoort, Albert Postma, L Jaap Kappelle, and Edward HF De Haan. 2000. The Corsi block-tapping task: standardization and normative data. *Applied neuropsychology* 7, 4 (2000), 252–258.

[27] Jean-Rémi King, Laura Gwilliams, Chris Holdgraf, Jona Sassenhagen, Alexandre Barachant, Denis Alexander Engemann, Eric Larson, and Alexandre Gramfort. 2017. Encoding and Decoding Neuronal Dynamics: Methodological Framework to Uncover the Algorithms of Cognition. *HAL open science* hal-01848442 (2017), 1–19.

[28] Marius Klug, Sein Jeung, Anna Wunderlich, Lukas Gehrke, Janna Protzak, Zakaria Djebbara, Andreas Argubi-Wollesen, Bettina Wollesen, and Klaus Gramann. 2022. The BeMoBIL Pipeline for automated analyses of multimodal mobile brain and body imaging data. *bioRxiv* 1, 1 (2022), 2022–09.

[29] Christian Kothe, David Medine, Chadwick Boulay, Matthew Grivich, and Tristan Stenner. 2019. *Labstreaminglayer*. Labstreaminglayer. Retrieved January 04, 2023 from https://labstreaminglayer.readthedocs.io/index.html

[30] Wei Lun Lim, Yisi Liu, Salem Chandrasekaran Harihara Subramaniam, Serene Hui Ping Liew, Gopala Krishnan, Olga Sourina, Dimitrios Konovessis, Hock Eng Ang, and Lipo Wang. 2018. EEG-based mental workload and stress monitoring of crew members in maritime virtual simulator. In *Transactions on Computational Science XXXII*. Springer, Berlin, Heidelberg, 15–28.

[31] Tania Llana, Cristina Fernandez-Baizan, Magdalena Mendez-Lopez, Camino Fidalgo, and Marta Mendez. 2022. Functional near-infrared spectroscopy in the neuropsychological assessment of spatial memory: A systematic review. *Acta Psychologica* 224 (2022), 103525.

[32] Fabien Lotte, Josef Faller, Christoph Guger, Yann Renard, Gert Pfurtscheller, Anatole Lécuyer, and Robert Leeb. 2012. Combining BCI with virtual reality: towards new applications and improved BCI. In *Towards practical brain-computer interfaces*. Springer, Berlin/Heidelberg, Germany, 197–220.

[33] Robert Luke, Eric D Larson, Maureen J Shader, Hamish Innes-Brown, Lindsey Van Yper, Adrian KC Lee, Paul F Sowman, and David McAlpine. 2021. Analysis methods for measuring passive auditory fNIRS responses generated by a block-design paradigm. *Neurophotonics* 8, 2 (2021), 025008.

[34] J McAfoose and BT Baune. 2009. Exploring visual–spatial working memory: A critical review of concepts and models. *Neuropsychology review* 19 (2009), 130–142.

[35] Ryan McKendrick and Amanda Harwood. 2019. Cognitive workload and workload transitions elicit curvilinear hemodynamics during spatial working memory. *Frontiers in Human Neuroscience* 13 (2019), 405.

[36] Kim Mens, Rafael Capilla, Nicolás Cardozo, and Bruno Dumas. 2016. A taxonomy of context-aware software variability approaches. In *Companion Proceedings of the 15th International Conference on Modularity*. Association for Computing Machinery, New York, NY, United States, 119–124.

[37] George A Miller. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review* 63, 2 (1956), 81.

[38] Guiomar Niso, Elena Romero, Jeremy T Moreau, Alvaro Araujo, and Laurens R Krol. 2023. Wireless EEG: A survey of systems and studies. *NeuroImage* 269 (2023), 119774.

[39] François Osiurak, Jordan Navarro, and Emanuelle Reynaud. 2018. How our cognition shapes and is shaped by technology: A common framework for understanding human tool-use interactions in the past, present, and future. *Frontiers in psychology* 9 (2018), 293.

[40] Evan M Peck, Daniel Afergan, Beste F Yuksel, Francine Lalooses, and Robert JK Jacob. 2014. Using fNIRS to measure mental workload in the real world. In *Advances in physiological computing*. Springer, Berlin/Heidelberg, Germany, 117–139.

[41] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* 12 (2011), 2825–2830.

[42] Stéphanie Philippe, Alexis D Souchet, Petros Lameras, Panagiotis Petridis, Julien Caporal, Gildas Coldeboeuf, and Hadrien Duzan. 2020. Multimodal teaching, learning and training in virtual reality: a review and case study. *Virtual Reality & Intelligent Hardware* 2, 5 (2020), 421–442.

[43] Maria Piñeyro Salvidegoitia, Nadine Jacobsen, Anna-Katharina R Bauer, Benjamin Griffiths, Simon Hanslmayr, and Stefan Debener. 2019. Out and about: Subsequent memory effect captured in a natural outdoor environment with smartphone EEG. *Psychophysiology* 56, 5 (2019), e13331.

[44] Paola Pinti, Clarisse Aichelburg, Sam Gilbert, Antonia Hamilton, Joy Hirsch, Paul Burgess, and Ilias Tachtsidis. 2018. A review on the use of wearable functional near-infrared spectroscopy in naturalistic environments. *Japanese Psychological Research* 60, 4 (2018), 347–373.

[45] Kathrin Pollmann, Oilver Stefani, Amelie Bengsch, Matthias Peissner, and Mathias Vukelić. 2019. How to work in the car of the future? A neuroergonomical study assessing concentration, performance and workload based on subjective, behavioral and neurophysiological insights. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, United States, 1–14.

[46] Jos Prinsen, Ethel Pruss, Anita Vrins, Caterina Ceccato, and Maryam Alimardani. 2022. A Passive Brain-Computer Interface for Monitoring Engagement during Robot-Assisted Language Learning. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, IEEE, Prague, Czech Republic, 1967–1972.

[47] Felix Putze, Christian Herff, Christoph Tremmel, Tanja Schultz, and Dean J Krusienski. 2019. Decoding mental workload in virtual environments: a fNIRS study using an immersive n-back task. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, IEEE, Berlin, Germany, 3103–3106.

[48] Julian E Reiser, Edmund Wascher, Gerhard Rinkenauer, and Stefan Arnau. 2021. Cognitive-motor interference in the wild: Assessing the effects of movement complexity on task switching using mobile EEG. *European Journal of Neuroscience* 54, 12 (2021), 8175–8195.

[49] Raphaëlle N Roy, Marcel F Hinss, Ludovic Darmet, Simon Ladouce, Emilie S Jahanpour, Bertille Somon, Xiaoqi Xu, Nicolas Drougard, Frédéric Dehais, and Fabien Lotte. 2022. Retrospective on the First Passive Brain-Computer Interface Competition on Cross-Session Workload Estimation. *Frontiers in Neuroergonomics* 3 (2022), 1–8.

[50] Richard M Ryan and Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1 (2000), 68.

[51] Yangming Shi, Yibo Zhu, Ranjana K Mehta, and Jing Du. 2020. A neurophysiological approach to assess training outcome under stress: A virtual reality experiment of industrial shutdown maintenance using Functional Near-Infrared Spectroscopy (fNIRS). *Advanced Engineering Informatics* 46 (2020), 101153.

[52] Mojtaba Soltanlou, Maria A Sitnikova, Hans-Christoph Nuerk, and Thomas Dresler. 2018. Applications of functional near-infrared spectroscopy (fNIRS) in studying cognitive development: The case of mathematics and language. *Frontiers in psychology* 9 (2018), 277.

[53] Martin Spüler, Tanja Krumpe, Carina Walter, Christian Scharinger, Wolfgang Rosenstiel, and Peter Gerjets. 2017. Brain-computer interfaces for educational applications. *Informational environments: Effects of use, effective designs* 1 (2017), 177–201.

[54] Megan Strait and Matthias Scheutz. 2014. What we can and cannot (yet) do with functional near infrared spectroscopy. *Frontiers in neuroscience* 8 (2014), 117.

[55] Christoph Tremmel, Christian Herff, Tetsuya Sato, Krzysztof Rechowicz, Yusuke Yamani, and Dean J Krusienski. 2019. Estimating cognitive workload in an interactive virtual reality environment using EEG. *Frontiers in human neuroscience* 13 (2019), 401.

[56] Anirudh Unni, Klas Ihme, Meike Jipp, and Jochem W Rieger. 2017. Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: A realistic driving simulator study. *Frontiers in human neuroscience* 11 (2017), 167.

[57] Ross E Vanderwert and Charles A Nelson. 2014. The use of near-infrared spectroscopy in the study of typical and atypical development. *Neuroimage* 85 (2014), 264–271.

[58] Kevin J Verdière, Raphaëlle N Roy, and Frédéric Dehais. 2018. Detecting pilot's engagement using fNIRS connectivity features in an automated vs. manual landing scenario. *Frontiers in human neuroscience* 12 (2018), 6.

[59] Alexander von Lühmann, Antonio Ortega-Martinez, David A Boas, and Meryem Ayşe Yücel. 2020. Using the general linear model to improve performance in fNIRS single trial analysis and classification: a perspective. *Frontiers in human neuroscience* 14 (2020), 1–17.

[60] Alexander von Lühmann, Yilei Zheng, Antonio Ortega-Martinez, Swathi Kiran, David C Somers, Alice Cronin-Golomb, Louis N Awad, Terry D Ellis, David A Boas, and Meryem A Yücel. 2021. Toward Neuroscience of the Everyday World (NEW) using functional near-infrared spectroscopy. *Current opinion in biomedical engineering* 18 (2021), 100272.

[61] Mathias Vukelić. 2021. Connecting Brain and Machine: The Mind Is the Next Frontier. *Clinical Neurotechnology meets Artificial Intelligence: Philosophical, Ethical, Legal and Social Implications* 1 (2021), 215–226.

[62] Mathias Vukelić, Katharina Lingelbach, Kathrin Pollmann, and Matthias Peissner. 2020. Oscillatory EEG Signatures of Affective Processes during Interaction with Adaptive Computer Systems. *Brain Sciences* 11, 1 (2020), 35.

[63] Carina Walter, Wolfgang Rosenstiel, Martin Bogdan, Peter Gerjets, and Martin Spüler. 2017. Online EEG-based workload adaptation of an arithmetic learning environment. *Frontiers in human neuroscience* 11 (2017), 286.

[64] Mark S Young, Karel A Brookhuis, Christopher D Wickens, and Peter A Hancock. 2015. State of science: mental workload in ergonomics. *Ergonomics* 58, 1 (2015), 1–17.

[65] Meryem A Yücel, Alexander v Lühmann, Felix Scholkmann, Judit Gervain, Ippeita Dan, Hasan Ayaz, David Boas, Robert J Cooper, Joseph Culver, Clare E Elwell, et al. 2021. Best practices for fNIRS publications. *Neurophotonics* 8, 1 (2021), 012101.

[66] Thorsten O Zander and Christian Kothe. 2011. Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of neural engineering* 8, 2 (2011), 025005.

[67] Thorsten O Zander, Kunal Shetty, Romy Lorenz, Daniel R Leff, Laurens R Krol, Ara W Darzi, Klaus Gramann, and Guang-Zhong Yang. 2017. Automated task load detection with electroencephalography: towards passive brain–computer interfacing in robotic surgery. *Journal of Medical Robotics Research* 2, 01 (2017), 1750003.

[68] Rob Zink, Borbála Hunyadi, Sabine Van Huffel, and Maarten De Vos. 2016. Mobile EEG on the bike: disentangling attentional and physical contributions to auditory attention tasks. *Journal of neural engineering* 13, 4 (2016), 046017.