


Neural Correlates of Cognitive Load While Playing an Emergency Simulation Game: A Functional Near-Infrared Spectroscopy (fNIRS) Study

Natalia Sevchenko , Betti Schopp, Thomas Dresler, Ann-Christine Ehlis, Manuel Ninaus, Korbinian Moeller, and Peter Gerjets

Abstract—Functional near-infrared spectroscopy (fNIRS) provides reliable results for determining cognitive load based on averaged cortical blood flow during multiple repetitions of short cognitive tasks. At the same time, it remains unclear how to use this technique for assessing cognitive load during prolonged single-trial activity. In this study, we used a computer-based emergency simulation game for inducing different levels of cognitive load. We propose a novel approach to measure cognitive load using specific time slots, determined based on simulation log-data interpreted in light of Barrouillet’s time-based resource-sharing model. To validate this approach, we compared cortical activity in dorsolateral prefrontal cortex (DLPFC) and left inferior frontal gyrus (IFG)

regions measured at four specific time slots during a simulation. We found significant associations between cognitive load and neuronal activity within the DLPFC depending on the chosen time slot, whereas no such dependencies were found for the IFG. These results illustrate how knowledge of task structure could be used advantageously for the identification of cognitive load. Although requiring further investigation in terms of reliability and generalizability, the presented approach can be considered promising evidence that fNIRS might be suitable for more general reliable assessments of cognitive load during prolonged single-trial activities and for real-time adaptations in simulation-based learning environments.

Index Terms—Adaptivity, cognitive load, functional near-infrared spectroscopy (fNIRS), game, simulation.

Manuscript received 26 February 2021; revised 25 June 2021 and 17 December 2021; accepted 18 December 2021. Date of publication 13 January 2022; date of current version 15 December 2022. This work was supported by Daimler Trucks AG. (Corresponding author: Natalia Sevchenko.)

Natalia Sevchenko is with the Department of Psychology, University of Tuebingen, 72074 Tuebingen, Germany, and also with Daimler Trucks AG, 70771 Stuttgart, Germany (e-mail: n.sevchenko@gmail.com).

Betti Schopp is with the Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health, University Hospital Tuebingen, 72076 Tuebingen, Germany (e-mail: betti.schopp@med.uni-tuebingen.de).

Thomas Dresler is with LEAD Graduate School & Research Network, University of Tuebingen, 72072 Tuebingen, Germany, and also with the Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health, University Hospital Tuebingen, 72076 Tuebingen, Germany (e-mail: thomas.dresler@med.uni-tuebingen.de).

Ann-Christine Ehlis is with the Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health, University Hospital Tuebingen, 72076 Tuebingen, Germany, and also with LEAD Graduate School & Research Network, University of Tuebingen, 72072 Tuebingen, Germany (e-mail: ann-christine.ehlis@med.uni-tuebingen.de).

Manuel Ninaus is with the University of Innsbruck, 6020 Innsbruck, Austria, with the Leibniz Institut für Wissensmedien, University of Tuebingen, 72076 Tuebingen, Germany, with LEAD Graduate School & Research Network, University of Tuebingen, 72072 Tuebingen, Germany, and also with the Department of Psychology, Karl-Franzens-University Graz Institute of Psychology, 8010 Graz, Austria (e-mail: manuel.ninaus@uni-graz.at).

Korbinian Moeller is with LEAD Graduate School & Research Network, University of Tuebingen, 72072 Tuebingen, Germany, and also with the Centre for Mathematical Cognition School of Science Loughborough University, LE1-132U Loughborough U.K. (e-mail: k.moeller@lboro.ac.uk).

Peter Gerjets is with LEAD Graduate School & Research Network, University of Tuebingen, 72072 Tuebingen, Germany, and also with the Leibniz Institut für Wissensmedien, University of Tuebingen, 72076 Tuebingen, Germany (e-mail: p.gerjets@iwm-tuebingen.de).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Ethics committee of Leibniz-Instituts für Wissensmedien (IWM). The patients/participants provided their written informed consent to participate in this study.

This article has supplementary material provided by the authors and color versions of one or more figures available at <https://doi.org/10.1109/TG.2022.3142954>.

Digital Object Identifier 10.1109/TG.2022.3142954

I. INTRODUCTION

DAILY life is becoming increasingly automated: autopilots, speed and lane assistants, etc. help us to deal with well-controlled and predictable settings. In contrast, however, time-critical situations, such as driving a car in heavy traffic or performing a complex surgery still require a qualified human operator, preferably trained for such challenges. As it is not trivial to set up an analog training environment for life- and time-critical emergencies, computer-aided simulations, and digital training scenarios can be used advantageously in this area [1]–[3]. Along with the potential to simulate dangerous time-critical situations, digital training scenarios also allow for the collection of individual data, which can be used to model cognitive or emotional states of the user [4] including cognitive load.

Cognitive load is considered to reflect “how hard the brain is working to meet task demands” [5]. According to the most influential theoretical view, the relationship between user performance and cognitive load is best described by Yerkes and Dodson [6] as an “inverted-U,” also closely related to the concept of “flow” proposed by Csikszentmihalyi [7], [8], meaning that performance usually decreases in cognitive over- and under-load conditions [e.g., 6, 9, 10]. Consequently, human–machine interaction such as any computer-based training, might be optimized using a monitoring system capable of detecting variations in cognitive load [11], which seems feasible in real-world training environments [12], [13].

The literature describes four main categories of techniques for assessing cognitive load [14]–[16]. First, subjective measures

are collected using self-reported questionnaires e.g., [17], [18]. They are inexpensive and reliable [19], but are not capable of tracking variations in cognitive load online over time. Second, performance-based approaches evaluate fluctuations in human performance, i.e., task outcome measures, and relate these to changes in psychological constructs. This category of measurement techniques appears intuitively to be the most obvious and direct, but it is often impossible to obtain the necessary data until the actual task has already been completed. Moreover, it is hard to determine whether observed variations in performance have occurred due to changes in cognitive load or some other factors [14]. Third, behavioral measurements evaluate differences in interaction behavior (e.g., speech patterns or mouse usage) with the training system [20]–[24]. These measures are usually unobtrusive and inexpensive, do not require additional equipment, and potentially allow for continuous monitoring of cognitive states. However, behavioral patterns can be influenced by factors unrelated to cognitive load (e.g., emotions or stress). Finally, physiological approaches are based on the evidence that changes in cognitive states are accompanied by physiological changes [25]–[30]. Their advantage is that they allow continuous recording of data and thus might be used for online adaptation of training environments. At the same time, monitoring of physiological signals often requires expensive equipment as well as sophisticated filtering and analysis procedures. Moreover, different types of physiological measures differ considerably in their obtrusiveness and practicability in real-life settings. While sensors for heart rate variability (HRV), electrodermal activity (EDA), or eye-tracking can be used in a comparably discreet way, electroencephalography (EEG) or functional magnetic resonance imaging (fMRI) are less practical or even impracticable in real-life situations because of their signal sensitivity and immobility [see: 31]. According to Brunken *et al.* [32] we can further categorize physiological measurements into indirect and direct methods, based on the type of relation between cognitive load and observed variables. From this perspective, measurements such as pupil dilation or HRV are only indirectly related to cognitive load, as they can be influenced by other factors such as emotional response or stress, whereas imaging techniques such as fMRI, EEG, and functional near-infrared spectroscopy (fNIRS) can be considered direct methods, as they usually assess cortical activation during task execution¹. This approach seems to be very promising, as it can be applied in real-time, independent of the training software. On the other hand, the major limitation of most neurophysiological and imaging techniques is that they are expensive, complex, and immobile, which significantly limits their use in ecologically valid studies. In this respect, fNIRS offers some advantages over other neuroimaging techniques and appears promising in this field.

¹To be perfectly precise, all these methods can also be considered as indirect, because they deduce cognitive activity through blood flow or electrical activation. However, as there is no more direct method of capturing it today, we retain this terminology below.

A. Functional Near-Infrared Spectroscopy

Near-infrared spectroscopy was first presented in the fundamental work of Jobsis [33]. It represents a noninvasive neuroimaging method for measuring cortical hemodynamics, which relies on the mechanism of neurovascular coupling and optical spectroscopy [34], for review see: [35] and [36]. Hereby, near-infrared light with strictly defined wavelengths in the range between 600 and 1000 nm is emitted through the scalp of the participant. It penetrates up to approx. 2 cm [37] deep into the tissue (depending on the distance between light source and detector) and thus reaches outer cortical gray matter, where the emitted light is partially absorbed by oxygenated (HbO₂) and deoxygenated (HbR) hemoglobin molecules that differ significantly in their absorption spectra. The residual reflection of the respective wavelength is received by a detector, with a usual inter-optode distance of about 3 cm [36], [37]. Based on the difference between emitted and detected light [see 38], the relative concentration of HbO₂ and HbR in the brain tissue underlying the mean distance between light transmitter and its detector is calculated, allowing inferences about local changes in blood flow.

Neuronal activation in a particular brain area increases its metabolic needs and leads to an increase in local cerebral blood flow [39], [40]. This response is called “functional hyperemia” and is mediated by mechanisms of neurovascular coupling [41], resulting in an increase in HbO₂ with a simultaneous decrease in HbR in the respective area. Based on these observations, a temporal change in detected HbO₂ and HbR hemoglobin concentration allows conclusions to be drawn regarding changes in local brain activation. Therefore, depending on the measurement area, it seems possible to draw conclusions about which cognitive processes take place in the brain at a certain point of time. When attempting to evaluate changes in cortical activation related to, for instance, cognitive load, areas of the prefrontal cortex (PFC) are typically regarded [36].

B. Prefrontal Cortex

The PFC is part of the neocortex and is located in the anterior part of the frontal lobe. The integrative theory of prefrontal cortex function by Miller and Cohen [42] posits that the PFC generates special “bias” signal patterns that alter further stimulus processing either by blocking or amplifying desired processing paths and thus adapting reflectively produced behaviors to the current context. For example, if on the way to work (which is an automated behavior) we see someone lying in the street, we have the choice of interrupting the automated behavior or not, which would depend on our experience and the context, including whether we have enough time or whether we see that the person is already being helped, etc. This decision-making process seems associated with PFC involvement along with at least partial influences of working memory [43]. As such, the PFC seems an obvious location to place fNIRS optodes when measuring cognitive load. This placement also has an obvious technical advantage. The scalp in the forehead area and above the forehead is usually less hair coverage, which makes the measurement more reliable [44].

C. fNIRS: Strengths and Limitations

When comparing the temporal and spatial resolution of neuroimaging methods, fNIRS provides solid results in both domains [45], even when comparing short segments of data [46]. The great strength of the method is that, depending on the fNIRS device used, experimental tasks can be performed while sitting, standing, or even in motion, which makes the obtained results ecologically far more valid than results from experiments in which participants have to lie down, as during fMRI measurements. Furthermore, compared to EEG, this method is less susceptible to motion artifacts, electro-oculographic and facial electromyographic activity, as well as electrical environmental noise, which might be particularly problematic in neuronal measurements during human-machine interactions [see 47]. As such, fNIRS has been successfully used for investigating cognitive and emotional processes during gaming [e.g., 48, 49].

However, fNIRS is not completely free of artifacts and the recorded data must be pre-processed for analysis. The measurement method is sensitive to changing light conditions, which must be taken into account when planning a study. Moreover, the hair of the participant must be carefully pushed to the side at each fNIRS measurement optode, as it can strongly influence the quality of the signal [50], [51].

A solid body of evidence supports that fNIRS measures of PFC are sensitive to changes in cognitive load. Several experiments documented an increase of PFC activation with increasing difficulty of n-back [52]–[57] and *Stroop* tasks [58], [59]. These fundamental studies substantiated the usage of fNIRS for the measurement of cortical activation in laboratory settings. However, in contrast to a well-structured laboratory experiment, typical real-life settings contain a multitude of heterogeneous events (e.g., visual recognition, emotion, and memorization) that occur simultaneously and concurrently, making it difficult to impossible to assign cortical activation to specific events. That raises the question of whether and how technology can be used for more complex experiments with heterogeneous and more ecologically valid tasks, as they typically happen in everyday life. First such attempts (see some examples below) have already been made.

For example, Ayaz *et al.* [5] examined the hemodynamics during an air traffic management task in which participants had to handle 6, 12, or 18 air forces, respectively. The researchers found increased activity within the left PFC corresponding to defined levels of difficulty during short task sequences as compared to pretask resting periods.

Bruno *et al.* [60] also observed increased activity of bilateral dorsolateral PFC with increasing task difficulty using fNIRS for a simulated driving task. In another study by Unni *et al.* [61], an elegant design was used to integrate a standard n-back routine into a driving task using a virtual reality driving simulator. Task difficulty was adjusted by asking participants to adjust their speed according to the speed signal that appeared before n steps (i.e., in the 1-back condition participants had to adapt their speed to the last speed sign, while in the 4-back condition they had to maintain the speed prescribed by the fourth last sign). The

observed changes in HbR in bilateral inferior frontal areas and bilateral temporo-occipital areas were found to reflect cognitive load induced by the adapted n-back task.

These examples indicate that fNIRS can be a useful method for conducting ecologically valid realistic experiments. However, although these studies investigated realistic tasks, they mainly used methods that are only applicable in laboratory settings. In all of these studies, hemodynamic cortical activation was measured during short periods of high cognitive load, and data were aggregated over a large number of equivalent repetitions. Thus, these results indicate usability of fNIRS technology while measuring cognitive load, but does not answer the question of whether fNIRS can also be used to measure cognitive load online during long-term heterogeneous tasks that are not repeated several times, as it happens often in real-life. And because such tasks usually involve different mental activities, which can overlap in time and be executed in different sequences depending on participants' strategy, it is hard to tell exactly what specific timeframes should be used when evaluating cognitive load. Another common approach as described by Gerjets *et al.* [12] consists of using machine learning methods based on behavioral and (neuro-) physiological data, which seems to be quite feasible for real-live adaptation, leaving open the question of how to generalize the model, because such a data-driven approach cannot be easily generalized to different subjects and situations. Against this background, one might wonder whether an appropriate theoretical approach might help to resolve these issues. For measurement of cognitive load during the simulation of time-critical emergencies, a time-based resource-sharing (TBRS) model by Barrouillet *et al.* [62] might provide a suitable theoretical framework.

D. Time-Based Resource-Sharing Model of Cognitive Load

Barrouillet *et al.* [62] emphasized that, in addition to task complexity, cognitive load is strongly dependent on the time available for the task at hand, which is particularly relevant for time-critical situations. According to the TBRS model cognitive load depends on the proportion of time, “during which attention is captured in such a way that the storage of information is disturbed,” which is not trivial to determine. Nevertheless, in their recent paper, Sevchenko *et al.* [63] have shown that this model can be successfully applied to predicting cognitive load in time-critical serious games. Thereby, they used the TBRS model to define a behavioral metric based on the ratio of specific time intervals. Relying on log data, they found that the game flow can be divided into so-called action blocks, i.e., time-segments of dense actions (burst) followed by some waiting time (idle). When assuming that during the burst periods participants are operating at their cognitive limit, which appears plausible under time-critical emergency situations, their cognitive load should be comparably high (i.e., at their personal maximum), and thus, cognitive load of an action block can be estimated as the relation of the duration of the burst phase to the total duration of the inspected action block (burst + idle). This way, behavioral metric temporal action density decay (TADD) was proposed and validated in 47 participants. Interestingly, it was discovered that

TADD reflecting this relation within the first action block (initial TADD) was a more valid measure of cognitive load as compared with averaged TADD sequence calculated over a whole level. This means that a simple behavioral metric at the beginning of a level significantly predicted the success of the whole level. The fact that initial TADD was based on data collected very early on during the game process makes it potentially useful for real-time adaptation systems.

This apparently provides us with a measurement foundation that 1) is theory-driven and can thus be generalized to any time-critical resource management situation; 2) is based on relatively short time intervals (burst and idle); 3) can be calculated from the initial phase of training and is thus suitable for real-time adaptation. However, which time slots should be used when applying this analytical approach using fNIRS methodology? Idle time intervals can become too short (or even equal to zero), e.g., if the player fails to act fast enough and the first occupied emergency personnel finishes his task before the player has completed all ongoing task assignments. This makes idle time intervals unqualified for fNIRS analysis. On the other hand, burst periods seem sufficiently long and comparable, because during the first burst interval all participants are completing nearly the same tasks. However, also because all participants are supposed to operate at their limit under time-critical conditions, we can expect similar neuronal activity during this phase. In this article, we aimed at exploring these options and evaluating whether the direct observation of cortical hemodynamics during the initial burst phase and the idle phase directly afterward can provide additional insights into the nature and assessment of cognitive load.

E. Present Study

In this study, we pursue two main questions. We are interested in whether fNIRS technology is feasible for cognitive load detection in realistic time-critical emergency situations realized with a game; and if so, which time intervals should be used, considering the TBRs model [62] and previous results. In particular, we evaluate cortical activation in the PFC region for specified time slots addressing whether the corresponding cortical activation is related 1) to the task difficulty, 2) to the achieved performance, 3) and to the subjective perception of cognitive load. To allow for variance in cognitive load, we use a computer-based time-critical emergency simulation game, which requires management of time-critical situations and realizes different levels of difficulty.

II. METHODS

The study was carried out as part of a larger project that included several other physiological measures such as cardiac measurements, galvanic skin response, and eye-tracking. The aim of the study was to investigate whether the fNIRS methodology would be feasible for detecting cognitive load in realistic environments.

A. Participants

In this study, we present data of 27 volunteers (18 females, 9 males) aged between 20 and 49 years ($M = 25.9$; $SD = 7.2$). Data of further 20 participants were excluded from the present

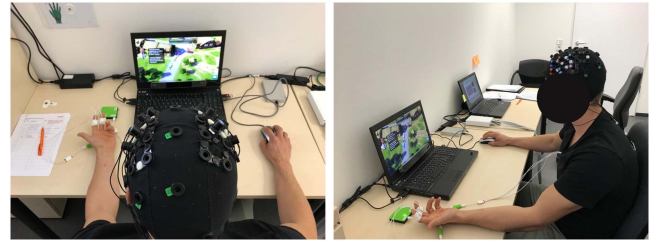


Fig. 1. Experimental setup.

analysis due to the poor quality of the fNIRS recording. All participants were right-handed, spoke fluent German, were recruited via an online database, and were compensated monetarily for their time expenditure. They reported no neurological, mental, or cardiovascular disorders, and did not take any psychotropic medication. The local ethics committee approved the study and written informed consent was obtained from all participants prior to the experiment.

B. Task

All participants played an adapted version of a game-based simulation of different emergency situations [Emergency: 64], in which they had to coordinate emergency personnel, such as paramedics, emergency doctors, and firefighters, as well as auxiliary items such as ambulances, fire trucks, and fire truck ladders to rescue victims and extinguish fires.

After getting familiar with the games' simulation routine by playing a tutorial and a training scenario, participants were confronted with two experimental scenarios: Fire and Train Crash. Each scenario included three levels of difficulty: easy, medium, and hard. These were defined by varying the number of tasks to be performed and the number of personnel to be coordinated within a given period of time. Because increased task density would require not only more actions but also better planning, coordination, and prioritization, we expected this to generate a concomitant increase in cognitive load. The time pressure was imposed by setting time limits for the levels as well as time bars that showed how fast a victim would die if not helped [for detailed summary see: 63].

C. Experimental Setup and Design

The experiment was implemented in a quiet room under constant light conditions. The Emergency game simulation was presented on a 16" notebook driven at a screen resolution rate of 1920×1080 using a conventional computer mouse as the only interaction tool. Data for cortical hemodynamics were acquired on an additional notebook as can be seen in Fig. 1, using a portable NIRSPORT-2 device [65]. The simulation started with an introductory training sequence, which was followed by two experimental scenarios: Fire and Train Crash. At the end of each of the three levels per scenario, NASA-TLX scores were collected. Each participant executed the defined sequence of levels only once, which lasted about one hour including the training phase.

D. Measured Variables

For estimating cognitive load we used a continuous multichannel recording of cortical hemodynamics during time slots that were derived from participants' behavior as described below. These measurements were subsequently associated with level difficulty, performance data, and subjective estimation of cognitive load. Additional data such as age and sex were acquired prior to the experiment using a self-report survey.

1) *Level Difficulty and Performance Data*: For each level a real difficulty score was defined as the percentage of participants, who failed to complete the level, this means, could not extinguish all the fires and transport all injured persons to the hospital within the defined time limit. Individuals' performance per level was represented by the binary indicator of whether the level was completed successfully or not.

2) *NASA-TLX*: Subjective estimation of cognitive load was acquired using subscales of the multidimensional NASA-TLX [18] questionnaire, which consists of six items/dimensions rated on a 21-level scale (0 to 100 points with steps of 5). These dimensions correspond to various theories distinguishing between physical, mental, and emotional facets of operators' load [66]. In the current study, we used the subscales addressing the mental facet, i.e., mental demand, temporal demand, and effort, which represents a common procedure when investigating specific facets of workload [67], [68].

3) *Hemodynamic Cortical Data*: Hemodynamic cortical data were analyzed based on burst and idle time periods related to the initial TADD metric proposed by Sevchenko *et al.* [63]. According to this approach, the time series of participants' actions during each level was divided into so-called action blocks, consisting of active (burst) and waiting (idle) periods.

During the burst period, participants manage their emergency personnel, and after the last available personnel are assigned a task, the idle interval occurs and lasts until the first personnel is available again. In this study, we analyzed hemodynamic data obtained during four subsequent time slots. The first time slot (initial burst) starts with the first user action and its duration corresponds to the duration of the initial burst phase, which varied between participants. The three subsequent time slots start with the end of the corresponding initial burst and last 20 s each (t0–t20: starting directly after the initial burst; t20–t40: starting 20 s after the end of the initial burst; t40–t60: starting 40 s after the end of the initial burst).

Because we assumed that all participants would be working at their cognitive limit during the burst phase, we expected no effects of task difficulty, performance and subjective cognitive load on neuronal activation during this time. After the burst phase is over, participants who experienced high cognitive load were supposed to have only a very short idle phase if any. Consequently, during the 20 s following the initial burst, participants who experience low cognitive load should still be in the initial idle phase, whereas for participants with high cognitive load, the next burst should already start. Assuming that hemodynamic cortical activation during the initial burst phase differs from the activation during the initial idle phase, we expected to find effects

on cortical activation during this time. As time progresses, we expected more heterogeneity in the data, this means, depending on their strategies more and more participants would start with a new burst phase while others were still or again in an idle phase. Therefore, we expected smaller or even no significant effects in the later stages of the level.

E. FNIRS Imaging Procedure

Participants' PFC hemodynamic was recorded using a portable NIRSPORT-2 device [65], which works with two wavelengths (750 and 859 nm) and provides a time series of relative HbO₂ and HbR concentration changes [38]. We used eight light emitters and eight detectors, resulting in 20 channels, which were placed into electrode caps CUCMS-56/58 [69] and adjusted to the Cz and Fpz positions according to the 10–20 system [70]. The probeset covered the dorsolateral PFC and left inferior frontal gyrus (IFG), the latter being part of the PFC that is involved in speech processing [71], including the production of the inner dialogue [72]. The distance between a light emitter and its corresponding detector (i.e., the interoptode distance, the middle of which corresponds to a measurement channel), was approx. 3 cm. Data recording was performed at a sampling rate of 7.8125 Hz. The data were preprocessed with MATLAB version 2017a as described in Section II-E.

F. Data Analyses

Data analyses were carried out following three main steps.

1) *FNIRS Data Preprocessing*: First, data were corrected for high amplitude movement artifacts by the TDDR correction [73], followed by bandpass filtering (0.01–0.1 Hz) and a correlation-based signal improvement (CBSI) [74]. After that, visual data inspection for outlier-channels was conducted, and on average 2.2 channels per participant were interpolated with their surrounding channels. A following PCA-based Gaussian kernel filter was used for global signal correction [75] and data were z-standardized. Then, we determined the area under the curve (AUC) for this signal by calculating its integral. An important final step involved time normalization of the AUC (nAUC) to control for interindividual differences in the duration of the initial burst phases between participants ($M = 37.9$, $SD = 14.1$, see appendix: Fig. 3). Although no significant association was found between the duration of the initial burst phase and performance (linear mixed-effect analysis: $\chi^2(1) = 0.89$, $p = 0.34$, conditional $R^2 = 0.07$, marginal $R^2 < 0.001$, $\beta = -2.32$; see appendix: Fig. 4), we aimed at minimizing potential bias due to the duration of the initial burst phase by standardizing AUC individually for each participant.

2) *ROI Definition*: After preprocessing was completed, we combined 20 channels into the following six regions of interest (ROI), illustrated by Fig. 2: DLPFC left (DLPFC-L), DLPFC right (DLPFC-R), and IFG (IFG). ROI definition was carried out on a data-driven basis combined with theoretical assumptions about the different sections of the PFC and their respective functions. Thereby, we visually inspected for all channels a timeframe from 5–30 s after the initial burst phase had started.

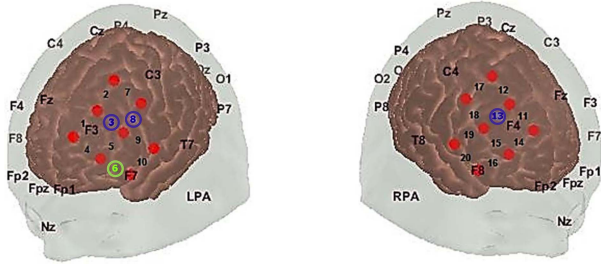


Fig. 2. Anatomical allocation of the channels to the brain regions and ROI definition. DLPFC left: channels 3, 8; DLPFC right: channels 13; IFG: 6. The figure was generated by the MATLAB-based AtlasViewer [86] with manually added channel numbers.

Channels were combined based on the anatomical topology, theoretical knowledge, and hemodynamic activation shown during this period.

3) *Statistical Analyses*: We employed linear mixed-effect analyses using statistical Software R [76] with the lme4 package [77]. The assumptions of homoscedasticity and normality were verified by visual inspection of residual plots. The p-values were obtained by likelihood ratio tests of the full model (with the effect of the investigated parameter) tested against a reduced model (without its effect). In case of a significant result, further model analyses were applied using the report package of Makowski *et al.* [78]. Standardized parameters were obtained by fitting a model on a standardized version of the dataset.

III. RESULTS

In this study, we examined hemodynamic cortical activation within specified time windows during a time-critical emergency simulation. Thereby, we aimed to evaluate whether these observations can be used to assess cognitive load induced by the simulation design. In particular, for four defined time periods (initial burst and following t_0 – t_{20} , t_{20} – t_{40} , and t_{40} – t_{60}), we examined the impact of real difficulty of the level (percentage of participants who failed to complete all tasks within the specified time limit) and the participants' performance per level (binary indicator of whether the level was completed successfully or not) on the hemodynamic cortical activation level (nAUC: time-normalized integral of the CBSI signal) in a specified ROI. Furthermore, we evaluated the impact of nAUC on subjective ratings of cognitive load for the whole simulation level, acquired using NASA-TLX questionnaire. Detailed statistics are provided in Appendix.

1) *Initial Burst*: We found a significantly positive effect of real difficulty on hemodynamic cortical activation in DLPFC-L and DLPFC-R, meaning that higher levels of real difficulty were associated with higher hemodynamic activity. In contrast, no significant association was found between real difficulty and hemodynamic cortical activation within the IFG ROI (Appendix: Table II). Performance was significantly negatively associated with hemodynamic cortical activation in the DLPFC-R, meaning that higher level of performance was associated with lower hemodynamic activity in this ROI (Appendix: Table III). No significant associations between perceived cognitive load and

hemodynamic cortical activation in terms of mental demand and effort were found (see Appendix: Tables IV and VI), whereas time demand was significantly positively associated with neuronal activation within DLPFC-R, implying that participants exhibiting higher neuronal activation also perceived higher time pressure (see Appendix: Table V).

2) t_0 – t_{20} : We found no significant effects within any of the ROIs regarding real difficulty and performance (see Appendix: Tables VII and VIII). Regarding perceived cognitive load, we found a significant negative effect of hemodynamic cortical activity only within DLPFC-L on the perceived mental demand (see Table IX), implying that higher hemodynamic cortical activation directly after the burst phase was associated with the experience of lower mental demand. Likewise, subjective time demand was significantly negatively associated with hemodynamic cortical activation in DLPFC-L ROI, i. e., the increased neuronal activity was related to the experience of less time pressure, whereas no significant effects for other ROIs were found (see Table X). Perceived effort was significantly associated with hemodynamic cortical activation within both ROIs related to DLPFC, again no effect of IFG was observed (see Table XI). This effect was negative, meaning that the increased neuronal activity was related to the experience of less effort.

3) t_{20} – t_{40} : We found a significant negative association between real difficulty and cortical hemodynamics only within DLPFC-R, meaning that higher real difficulty was associated with lower hemodynamic activation. No significant effects were found for other ROIs (see Appendix: Table XII). Regarding performance and subjective cognitive load, we found no significant effects within any of the ROIs (Appendix: Tables XIII, XIV, XV, and XVI).

4) t_{40} – t_{60} : We found no significant effects for this time period (see Appendix: Tables XVII, XVIII, XIX, XX, and XXI).

IV. DISCUSSION

Determining cognitive load seems crucial for the development of training environments for life- and time-critical emergencies. Here, the use of fNIRS methodology may provide reliable results in this respect not only in the laboratory but also in ecologically more valid experiments, using averaged data of repeated measures of relatively short cognitive tasks. At the same time, it remains unclear how to use fNIRS in prolonged single-trial activity involving heterogeneous tasks. In particular, because averaging over a too long period of time may well obscure potential differences in hemodynamic cortical activation, the question arises as to which time windows may be used sensibly to obtain reliable data on experienced cognitive load.

This study describes an attempt to solve this challenge by following a theory-driven top-down approach for investigating realistic time-critical emergency situations from the perspective of resource management.

In accordance with our expectations and despite the fact that simulation levels varied significantly in terms of real difficulty and subjective cognitive load, only a few or no significant effects on hemodynamic cortical activation were observed during later

time intervals: t20–t40 and t40–t60 s after the end of the initial burst. These results substantiate that it is not always possible to determine differences in cognitive load by comparing random time intervals over different difficulty levels over a long and complex realistic task. Because a realistic cognitive task usually consists of heterogeneous subtasks, it cannot be guaranteed that all operators will perform these tasks in the same order and using the same strategy. This leads to a lot of “noise” in neuronal activity when different mental tasks are performed continuously or even overlapping at different times by different participants making it almost impossible to pinpoint influences of specific variables.

This problem might be solved by searching for comparable time slots using the global knowledge about the nature of the performed task. As hypothesized, we found a significant association between hemodynamic cortical activation within different areas of DLPFC immediately after the initial burst phase (t0–t20) and the subjective assessment of cognitive load of the entire level (assessed by NASA-TLX). This association seems robust, as it appeared for all investigated subscales (mental demand, time demand, and effort). Surprisingly, participants who showed stronger hemodynamic cortical activation within this time frame perceived the entire level as less demanding for them and vice versa. This pattern of results provides an interesting insight into the nature of the idle phase following the initial burst phase. While the simulation log captured the similarities in the way that participants acted, when managing initial resources during the first burst (i.e., allocating emergency personnel to their tasks) and thus may have relied on similar cognitive processes, there is no information about their cognitive engagement during the idle phase, because no logable behavior was performed in this time period. In principle, in this phase, they could either wait passively for the next burst or use this time for active monitoring and planning, which might result in better performance later on and thus to a reduction in the subjectively experienced cognitive load [79]. However, as no significant association between neuronal activity during this time slot and final performance was detected, another explanation seems more likely. It is conceivable that cognitively challenged participants tended to use the idle pause to relax as they might get tired of maintaining attention, which may have led to reduced hemodynamic cortical activation [80]. In contrast, more successful players might use this phase to monitor the situation and plan ahead their next steps. Nevertheless, answering this question seems to require further investigation.

Because we assumed that during the initial burst phase all participants would operate at their cognitive maximum, we expected no differences in hemodynamic cortical activation between participants and real difficulty at this time. Surprisingly, we found a considerable positive association between hemodynamic cortical activation within wide areas of the DLPFC ROI and real difficulty of the level as well as with perceived time pressure during the initial burst phase. In contrast, negative associations with the actual performance were found within the right DLPFC. First of all, these results may be interpreted as substantiation of the “inverted-U” shaped association of cognitive load and performance [6], [7], which assumes that the cognitive load should be kept within a certain range to obtain the best results.

Furthermore, the initial burst phase, which takes place at the very beginning of the level, appears nevertheless to be well suited to determine the degree of task difficulty and could therefore be used for real-time adaptation of training simulations. At the same time, however, further investigation is needed to determine whether this effect might persist under more stressful conditions. It is conceivable that our preliminary assumptions were correct, but the induced time pressure was not sufficient to make all participants work at their cognitive limit. In this case, a burst time slot would be suitable for measuring cognitive load in situations with low to medium time pressure, while for measuring cognitive load under high time pressure other options need to be identified.

All above-mentioned results refer to neuronal activity within broader ranges of the DLPFC. In this study, we found no association of IFG activation and participants’ cognitive load. This may be due to the nature of the experimental task, which did not require difficult calculations or other cognitive manipulations necessitating an inner dialogue.

A. Methodological Strengths and Constraints

Analytic approaches for simulated training environments are often based on data-driven probabilistic evaluations [81]–[83]. However, these seem insufficient for modeling cognitive user states, which can be “directly” [32] assessed via neurophysiological methods [for review see: 84]. Compared to other neuroimaging techniques, the use of the fNIRS methodology seems advantageous due to its relatively low cost, high mobility, robustness against various artifacts, and reliability when used on averaged data, whereas single-user single-trial evaluations still remain challenging [36], [85]. This study takes a step in this direction by proposing a theory-based approach to the choice of time frames suitable for assessing cognitive load in realistic single-trial time-critical emergency situations related to resource management. Unlike data-driven approaches such as machine learning, this approach might be generalized to similar environments, which however raises further questions. As mentioned above, we have to investigate whether a finer clustering of training scenarios is needed for the selection of the appropriate metrics, brain regions, and time frames for data collection. Also, in this constellation, fNIRS cannot be used as a stand-alone methodology, because it needs log data from the simulation to calculate the required time periods, which would result in relatively complex assessment installation. Thus, the obtained results might be used to improve laboratory systems with the aim of conducting mobile, ecologically valid experiments.

B. Implications and Outlook

To our knowledge, this study represents the first attempt to deploy the theoretical framework of the time-based resource sharing model for measuring cognitive load using fNIRS during a realistic training scenario. We believe this theoretical foundation might be a key to creating a measurement method that can be easily transferred to similar situations without requiring extensive calibration each time. The results, obtained in this study, seem to be promising and indicate feasibility of the proposed method. Nevertheless, additional research is needed to

investigate the scope of our approach. In subsequent research, we also plan to investigate whether the results obtained can be substantiated by using other measurement methods, such as eye tracking. Another important step will be the further development of the method (potentially by extension to other measurement modalities) and its verification on different time-critical resource management scenarios. In this context, the extent to which differences in induced time pressure might affect the validity of this method, is of particular interest, in order to address potential limitations. Another interesting research direction might be to investigate age-related differences in measures of cognitive load in time-critical situations.

V. CONCLUSION

In this study we presented a simple method to determine time periods that qualify for cognitive load detection in a single-trial time-critical resource-management emergency situation using the fNIRS method in combination with TBRS theory. Detection of proposed time periods is based on log data and can be easily run in the background. We found significant associations between cognitive load and neuronal activity within DLPFC during chosen time periods, whereas only a few or no significant effects were observed during later time intervals, substantiating that it is not always possible to determine differences in cognitive load by comparing random periods of time. We found no significant dependencies within IFG. These results illustrate how knowledge of task structure may be used advantageously for the identification of cognitive load. Although requiring further investigation in terms of reliability and generalizability, the presented approach seems promising evidence that fNIRS might be suitable for the assessment of cognitive load beyond classical experimental set-ups.

ACKNOWLEDGMENT

Author Natalia Sevchenko was employed by the company Daimler Trucks AG. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

- [1] Y. Liu *et al.*, "Human factors assessment in VR-based firefighting training in maritime: A pilot study," in *Proc. Int. Conf. Cyberworlds (CW)*, 2020, pp. 157–163.
- [2] J. G. Johnson, D. G. Rodrigues, M. Gubbala, and N. Weibel, "HoloCPR: Designing and evaluating a mixed reality interface for time-critical emergencies," in *Proc. 12th EAI Int. Conf. Pervasive Comput. Technol. Healthcare*, 2018, pp. 67–76.
- [3] J. P. Kincaid, J. Donovan, and B. Pettitt, "Simulation techniques for training emergency response," *Int. J. Emerg. Manage.*, vol. 1, no. 3, pp. 238–246, 2003.
- [4] S. Nebel and M. Ninaus, "New perspectives on game-based assessment with process data and physiological signals," in *Game-Based Assessment Revisited*, D. Ifenthaler and Y. Kim, Eds., Cham, Switzerland: Springer, 2019, pp. 141–161.
- [5] H. Ayaz, P. A. Shewokis, S. Bunce, K. Izzetoglu, B. Willems, and B. Onaral, "Optical brain monitoring for operator training and mental workload assessment," *Neuroimage*, vol. 59, no. 1, pp. 36–47, 2012.
- [6] R. M. Yerkes and J. D. Dodson, "The relation of strength of stimulus to rapidity of habit-formation," *J. Comp. Neurol. Psychol.*, vol. 18, no. 5, pp. 459–482, 1908.
- [7] M. Csikszentmihalyi, *Beyond Boredom and Anxiety*, 1st ed. San Francisco, CA, USA: Jossey-Bass Publishers, 1975.
- [8] K. Kiili, A. Lindstedt, and M. Ninaus, "Exploring characteristics of students' emotions, flow and motivation in a math game competition," in *Proc. GamiFIN Conf.*, Pori, Finland, May 2018, pp. 20–29.
- [9] K. J. Anderson, "Impulsivity, caffeine, and task difficulty: A within-subjects test of the Yerkes-Dodson law," *Pers. Individual Differences*, vol. 16, no. 6, pp. 813–829, 1994.
- [10] F. Montani, C. Vandenberghe, A. Khedhaouria, and F. Courcy, "Examining the inverted U-shaped relationship between workload and innovative work behavior: The role of work engagement and mindfulness," *Hum. Relations*, vol. 73, no. 1, pp. 59–93, 2020.
- [11] G. Orru and L. Longo, "The evolution of cognitive load theory and the measurement of its intrinsic, extraneous and germane loads: A review," in *Human Mental Workload: Models and Applications*, L. Longo and M. Leva, Eds., Cham, Switzerland: Springer, 2019, pp. 23–48.
- [12] P. Gerjets, C. Walter, W. Rosenstiel, M. Bogdan, and T. O. Zander, "Cognitive state monitoring and the design of adaptive instruction in digital environments: Lessons learned from cognitive workload assessment using a passive brain-computer interface approach," *Front. Neurosci.*, vol. 8, no. 385, 2014, doi: [10.3389/fnins.2014](https://doi.org/10.3389/fnins.2014).
- [13] T. Appel *et al.*, "Predicting cognitive load in an emergency simulation based on behavioral and physiological measures," in *Proc. Int. Conf. Multimodal Interact.*, W. Gao *et al.*, Eds., New York, NY, USA: Association for Computing Machinery, 2019, pp. 154–163.
- [14] R. Brünken, T. Seufert, and F. Paas, "Measuring cognitive load," in *Cognitive Load Theory*, New York: Cambridge University Press, pp. 181–202, 2010, doi: [10.1017/CBO9780511844744.011](https://doi.org/10.1017/CBO9780511844744.011).
- [15] F. T. Eggemeier, G. F. Wilson, A. F. Kramer, and D. L. Damos, "Workload assessment in multi-task environments," in *Multiple-task Performance*, D. L. Damos, Ed., Washington, DC, USA: Taylor & Francis, 1991, pp. 207–216.
- [16] M. W. Scerbo, "Theoretical perspectives on adaptive automation," in *Automation and Human Performance: Theory and Applications*, R. Parasuraman and M. Mouloua, Eds., Boca Raton, FL, USA: CRC Press, 1996, pp. 37–64.
- [17] G. B. Reid and T. E. Nygren, "The subjective workload assessment technique: A scaling procedure for measuring mental workload," in *Advances in Psychology*, vol. 52, Amsterdam, The Netherlands: Elsevier, 1988, pp. 185–218.
- [18] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Advances in Psychology*, vol. 52, Amsterdam, The Netherlands: Elsevier, 1988, pp. 139–183.
- [19] R. O'Donnell and F. Eggemeier, "Workload assessment methodology," in *Handbook of Perception and Human Performance*, vol. 2, K. R. Boff, L. Kaufman, and J. P. Thomas, Eds., Hoboken, NJ, USA: Wiley, 1986.
- [20] E. H. Magnusdottir, M. Borsky, M. Meier, K. Johannsdottir, and J. Gudnason, "Monitoring cognitive workload using vocal tract and voice source features," *Periodica Polytechnica Elect. Eng. Comput. Sci.*, vol. 61, no. 4, pp. 297–304, 2017.
- [21] N. Ruiz, G. Liu, B. Yin, D. Farrow, and F. Chen, "Teaching athletes cognitive skills: Detecting cognitive load in speech input," in *Proc. HCI*, 2010, pp. 484–488.
- [22] T. F. Yap, J. Epps, E. Ambikairajah, and E. H. Choi, "Voice source features for cognitive load classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2011, pp. 5700–5703.
- [23] C. S. Ikehara and M. E. Crosby, "Assessing cognitive load with physiological sensors," in *Proc. 38th Annu. Hawaii Int. Conf. Syst. Sci.*, New York, NY, USA, 2005, p. 295a, doi: [10.1109/HICSS.2005.103](https://doi.org/10.1109/HICSS.2005.103).
- [24] Y. M. Lim, A. Ayes, and M. Stacey, "Using mouse and keyboard dynamics to detect cognitive stress during mental arithmetic," in *Intelligent Systems in Science and Information SAI 2014. Studies in Computational Intelligence*, vol. 591, K. Arai, S. Kapoor, and R. Bhatia, Eds., Cham, Switzerland: Springer, 2015, pp. 335–350.
- [25] G. Johannsen, "Workload and workload measurement," in *Mental Workload*, vol. 8, N. Moray, Ed., Boston, MA, USA: Springer, 1979, pp. 3–11.
- [26] A. Fowler, K. Nesbitt, and A. Canossa, "Identifying cognitive load in a computer game: An exploratory study of young children," in *Proc. IEEE Conf. Games*, Aug. 2019, pp. 1–6.

- [27] M. Buchwald, S. Kupiński, A. Bykowski, J. Marcinkowska, D. Ratajczyk, and M. Jukiewicz, "Electrodermal activity as a measure of cognitive load: A methodological approach," in *Proc. IEEE Signal Process.: Algo., Architect., Arrangements, Appl. (SPA)*, Poznan, Poland, 2019, pp. 175–179, doi: [10.23919/SPA.2019.8936745](https://doi.org/10.23919/SPA.2019.8936745).
- [28] D. McDuff, S. Gontarek, and R. Picard, "Remote measurement of cognitive stress via heart rate variability," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 2957–2960.
- [29] Y. Liang, W. Liang, J. Qu, and J. Yang, "Experimental study on EEG with different cognitive load," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2018, pp. 4351–4356.
- [30] R. F. Ahmad, A. S. Malik, N. Kamel, and F. Reza, "Machine learning approach for classifying the cognitive states of the human brain with functional magnetic resonance imaging (fMRI)," in *Proc. 6th Int. Conf. Intell. Adv. Syst. (ICIAS)*, 2016, pp. 1–4.
- [31] M. Ninaus *et al.*, "Neurofeedback and serious games," in *Psychology, Pedagogy, and Assessment in Serious Games*. E. T. M. Connolly, T. Boyle, G. Hainey, P. Baxter, and Moreno-Ger, Eds., Hershey, PA, USA: IGI Global, 2013, pp. 82–110.
- [32] R. Brunken, J. L. Plass, and D. Leutner, "Direct measurement of cognitive load in multimedia learning," *Educ. Psychol.*, vol. 38, no. 1, pp. 53–61, 2003.
- [33] F. F. Jobsis, "Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters," *Sci.*, vol. 198, no. 4323, pp. 1264–1267, 1977.
- [34] G. Strangman, D. A. Boas, and J. P. Sutton, "Non-invasive neuroimaging using near-infrared light," *Biol. Psychiatry*, vol. 52, no. 7, pp. 679–693, 2002.
- [35] A. Fallgatter, A. Ehlis, A. Wagener, T. Michel, and M. Herrmann, "Near-infrared spectroscopy in psychiatry," *Der Nervenarzt*, vol. 75, no. 9, 2004, Art. no. 911.
- [36] F. Herold, P. Wiegel, F. Scholkmann, and N. G. Müller, "Applications of functional near-infrared spectroscopy (fNIRS) neuroimaging in exercise-cognition science: A systematic, methodology-focused review," *J. Clin. Med.*, vol. 7, no. 12, 2018, Art. no. 466.
- [37] G. E. Strangman, Z. Li, and Q. Zhang, "Depth sensitivity and source-detector separations for near infrared spectroscopy based on the Colin27 brain template," *PLoS One*, vol. 8, no. 8, 2013, Art. no. e66319.
- [38] M. Cope, D. Delpy, E. Reynolds, S. Wray, J. Wyatt, and P. Van der Zee, "Methods of quantitating cerebral near infrared spectroscopy data," in *Oxygen Transport to Tissue X*. New York, NY, USA: Springer, 1988, pp. 183–189.
- [39] S. G. Kim, E. Rostrup, H. B. Larsson, S. Ogawa, and O. B. Paulson, "Determination of relative CMRO2 from CBF and BOLD changes: Significant increase of oxygen consumption rate during visual stimulation," *Magn. Reson. Med.: Official J. Int. Soc. Magn. Reson. Med.*, vol. 41, no. 6, pp. 1152–1161, 1999.
- [40] R. D. Hoge, J. Atkinson, B. Gill, G. R. Crelier, S. Marrett, and G. B. Pike, "Linear coupling between cerebral blood flow and oxygen consumption in activated human cortex," in *Proc. Nat. Acad. Sci.*, vol. 96, no. 16, 1999, pp. 9403–9408.
- [41] A. R. Nippert, K. R. Biesecker, and E. A. Newman, "Mechanisms mediating functional hyperemia in the brain," *Neuroscientist*, vol. 24, no. 1, pp. 73–83, 2018.
- [42] E. K. Miller and J. D. Cohen, "An integrative theory of prefrontal cortex function," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 167–202, 2001.
- [43] P. Thier, "Die funktionelle architektur des präfrontalen kortex," in *Neuropsychologie*. Berlin, Germany: Springer, 2006, pp. 471–478.
- [44] J. M. Murkin and M. Arango, "Near-infrared spectroscopy as an index of brain and tissue oxygenation," *Brit. J. Anaesth.*, vol. 103, pp. i3–i13, 2009.
- [45] R. Parasuraman and M. Rizzo, "Introduction to neuroergonomics," in *Neuroergonomics: The Brain at Work*, Eds., New York, USA: Oxford University Press, pp. 3–12, 2007.
- [46] G. Strangman, R. Goldstein, S. L. Rauch, and J. Stein, "Near-infrared spectroscopy and imaging for investigating stroke rehabilitation: Test-retest reliability and review of the literature," *Arch. Phys. Med. Rehabil.*, vol. 87, no. 12, pp. 12–19, 2006.
- [47] G. Derosière, K. Mandrick, G. Dray, T. E. Ward, and S. Perrey, "NIRS-measured prefrontal cortex activity in neuroergonomics: Strengths and weaknesses," *Front. Hum. Neurosci.*, vol. 7, 2013, Art. no. 583.
- [48] S. E. Kober, G. Wood, K. Kiili, K. Moeller, and M. Ninaus, "Game-based learning environments affect frontal brain activity," *Plos one*, vol. 15, no. 11, 2020, Art. no. e0242573.
- [49] M. Witte, M. Ninaus, S. E. Kober, C. Neuper, and G. Wood, "Neuronal correlates of cognitive control during gaming revealed by near-infrared spectroscopy," *Plos one*, vol. 10, no. 8, 2015, Art. no. e0134816.
- [50] M. A. McIntosh, U. Shahani, R. G. Boulton, and D. L. McCulloch, "Absolute quantification of oxygenated hemoglobin within the visual cortex with functional near infrared spectroscopy (fNIRS)," *Invest. Ophthalmol. Vis. Sci.*, vol. 51, no. 9, pp. 4856–4860, 2010.
- [51] J. Pringle, C. Roberts, M. Kohl, and P. Lekeux, "Near infrared spectroscopy in large animals: Optical pathlength and influence of hair covering and epidermal pigmentation," *Vet. J.*, vol. 158, no. 1, pp. 48–52, 1999.
- [52] F. A. Fishburn, M. E. Norr, A. V. Medvedev, and C. J. Vaidya, "Sensitivity of fNIRS to cognitive state and load," *Front. Hum. Neurosci.*, vol. 8, 2014, Art. no. 76.
- [53] 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," *Front. Hum. Neurosci.*, vol. 7, 2014, Art. no. 935.
- [54] H. Ayaz, M. Izzetoglu, S. Bunce, T. Heiman-Patterson, and B. Onaral, "Detecting cognitive activity related hemodynamic signal for brain computer interface using functional near infrared spectroscopy," in *Proc. Neural Eng. CNE'07, 3rd Int. IEEE/EMBS Conf.*, May 2007, pp. 342–345.
- [55] C. Li, H. Gong, Z. Gan, and Q. Luo, "Monitoring of prefrontal cortex activation during verbal n-back task with 24-channel functional NIRS imager," in *Proc. Int. Soc. Opt. Photon.*, 2005, vol. 5630, pp. 882–891.
- [56] E. E. Smith and J. Jonides, "Working memory: A view from neuroimaging," *Cogn. Psychol.*, vol. 33, no. 1, pp. 5–42, 1997.
- [57] M. J. Herrmann *et al.*, "D4 receptor gene variation modulates activation of prefrontal cortex during working memory," *Eur. J. Neurosci.*, vol. 26, no. 10, pp. 2713–2718, 2007.
- [58] X. Xu, Z.-Y. Deng, Q. Huang, W.-X. Zhang, C.-Z. Qi, and J.-A. Huang, "Prefrontal cortex-mediated executive function as assessed by stroop task performance associates with weight loss among overweight and obese adolescents and young adults," *Behav. Brain Res.*, vol. 321, pp. 240–248, 2017.
- [59] A.-C. Ehlis, M. Herrmann, A. Wagener, and A. Fallgatter, "Multi-channel near-infrared spectroscopy detects specific inferior-frontal activation during incongruent stroop trials," *Biol. Psychol.*, vol. 69, no. 3, pp. 315–331, 2005.
- [60] J. L. Bruno *et al.*, "Mind over motor mapping: Driver response to changing vehicle dynamics," *Hum. Brain Mapping*, vol. 39, no. 10, pp. 3915–3927, 2018.
- [61] A. Unni, K. Ihme, M. Jipp, and J. W. Rieger, "Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: A realistic driving simulator study," *Front. Hum. Neurosci.*, vol. 11, 2017, Art. no. 167.
- [62] P. Barrouillet, S. Bernardin, and V. Camos, "Time constraints and resource sharing in adults' working memory spans," *J. Exp. Psychol.: Gen.*, vol. 133, no. 1, 2004, Art. no. 83.
- [63] N. Sevcenko, M. Ninaus, F. Wortha, K. Moeller, and P. Gerjets, "Measuring cognitive load using in-game metrics of a serious simulation game," *Front. Psychol.*, vol. 12, no. 906, 2021, doi: [10.3389/fpsyg.2021.572437](https://doi.org/10.3389/fpsyg.2021.572437).
- [64] Promotion Software GmbH, "World of emergency," *Promot. Softw. GmbH*. Accessed: Feb. 2022. [Online]. Available: <https://www.world-of-emergency.com/?lang=en>.
- [65] NIRX Medical Technologies. Accessed: Feb. 2022. [Online]. Available: <https://nirx.net/nirsport>
- [66] S. G. Hart, "NASA-task load index (NASA-TLX); 20 years later," *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 50, no. 9, pp. 904.908, 2006.
- [67] J. K. Haerle, M. J. Daly, H. H. Chan, A. Vescan, W. Kucharczyk, and J. C. Irish, "Virtual surgical planning in endoscopic skull base surgery," *Laryngoscope*, vol. 123, no. 12, pp. 2935–2939, 2013.
- [68] J. G. Temple, W. N. Dember, J. S. Warm, K. S. Jones, and C. M. LaGrange, "The effects of caffeine on performance and stress in an abbreviated vigilance task," *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 41, no. 2, pp. 1293–1297, 1997.
- [69] EASYCAP, "EASYCAP EEG recording caps and related products." Accessed: Feb. 2022. [Online]. Available: <https://www.easycap.de/>
- [70] H. Jasper, "Report of the committee on methods of clinical examination in electroencephalography," *Electroencephalogr. Clin. Neurophysiol.*, vol. 10, pp. 370–375, 1958.
- [71] A. Riecker, K. Mathiak, W. Grodd, I. Hertrich, and H. Ackermann, "Functional MRI reveals two distinct cerebral networks subserving speech motor control," *J. Acoustical Soc. Amer.*, vol. 117, no. 4, pp. 2574–2574, 2005.

- [72] P. McGuire, D. Silbersweig, R. Murray, A. David, R. Frackowiak, and C. Frith, "Functional anatomy of inner speech and auditory verbal imagery," *Psychol. Med.*, vol. 26, no. 1, pp. 29–38, 1996.
- [73] F. A. Fishburn, R. S. Ludlum, C. J. Vaidya, and A. V. Medvedev, "Temporal derivative distribution repair (TDDR): A motion correction method for fNIRS," *Neuroimage*, vol. 184, pp. 171–179, 2019.
- [74] X. Cui, S. Bray, and A. L. Reiss, "Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics," *Neuroimage*, vol. 49, no. 4, pp. 3039–3046, 2010.
- [75] X. Zhang, J. A. Noah, and J. Hirsch, "Separation of the global and local components in functional near-infrared spectroscopy signals using principal component spatial filtering," *Neurophotonics*, vol. 3, no. 1, 2016, Art. no. 015004.
- [76] R Core Team, "R: A language and environment for statistical computing," 2020. [Online]. Available: <https://www.R-project.org/>
- [77] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," 2014, *arXiv:1406.5823*.
- [78] D. Makowski, D. Lüdecke, and M. Ben-Schachar, "Automated reporting as a practical tool to improve reproducibility and methodological best practices adoption," *J. Open Source Softw.*, vol. 5, 2020, Art. no. 2815.
- [79] P. Hancock, "The effect of performance failure and task demand on the perception of mental workload," *Appl. Ergonom.*, vol. 20, no. 3, pp. 197–205, 1989.
- [80] T. Nihashi *et al.*, "Monitoring of fatigue in radiologists during prolonged image interpretation using fNIRS," *Japanese J. Radiol.*, vol. 37, no. 6, pp. 437–448, 2019.
- [81] B. Magerko, B. S. Stensrud, and L. S. Holt, "Bringing the schoolhouse inside the box—a tool for engaging, individualized training," *SOAR Technol. Inc.*, Ann Harbor, MI, USA, 2006. [Online]. Available: <https://apps.dtic.mil/sti/pdfs/ADA481593.pdf>
- [82] P. Spronck, M. Ponsen, I. Sprinkhuizen-Kuyper, and E. Postma, "Adaptive game AI with dynamic scripting," *Mach. Learn.*, vol. 63, no. 3, pp. 217–248, 2006.
- [83] A. E. Zook and M. O. Riedl, "A temporal data-driven player model for dynamic difficulty adjustment," in *Proc. 8th Artif. Intell. Interactive Digit. Entertainment Conf.*, 2012, pp. 93–98.
- [84] J. M. Kivikangas *et al.*, "A review of the use of psychophysiological methods in game research," *J. Gaming Virtual Worlds*, vol. 3, no. 3, pp. 181–199, 2011.
- [85] F. Scholkmann *et al.*, "A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology," *Neuroimage*, vol. 85, pp. 6–27, 2014.
- [86] C. M. Aasted *et al.*, "Anatomical guidance for functional near-infrared spectroscopy: AtlasViewer tutorial," *Neurophotonics*, vol. 2, no. 2, 2015, Art. no. 020801.