Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

Measuring Mental Workload Variations in Office Work Tasks using fNIRS

Serena Midha [*,a], Horia A. Maior [1,b], Max L. Wilson [2,c], Sarah Sharples [3,d]

[a] Horizon Centre for Doctoral Training, University of Nottingham, UK
[b] School of Computer Science, University of Lincoln, UK
[c] Mixed Reality Lab, School of Computer Science, University of Nottingham, UK
[d] Human Factors Research Group, Faculty of Engineering, University of Nottingham, UK

ARTICLE INFO

ABSTRACT

The motivation behind using physiological measures to estimate cognitive activity is typically to build technology that can help people to understand themselves and their work, or indeed for systems to do so and adapt. While functional Near Infrared Spectroscopy (fNIRS) has been shown to reliably reflect manipulations of mental workload in different work tasks, we still need to establish whether fNIRS can differentiate variety within common office-like tasks in order to broaden our understanding of the factors involved in tracking them in real working conditions. 20 healthy participants (8 females, 12 males), whose work included office-like tasks, took part in a user study that investigated a) the sensitivity of fNIRS for measuring mental workload variations in representations of everyday reading and writing tasks, and b) how representations of natural interruptions are reflected in the data. Results supported fNIRS measuring PFC activation in differentiating between workload levels for reading tasks but not writing tasks in terms of increased oxygenated haemoglobin (O2Hb) and decreased deoxygenated haemoglobin (HHb), for harder conditions compared to easier conditions. There was considerable support for fNIRS in detecting changes in workload levels due to interruptions. Variations in workload levels during the interruptions could be understood in relation to spare capacity models. These findings may guide future work into sustained monitoring of cognitive activity in real-world settings.

1. Introduction

Sensors to monitor our physiological wellbeing have arguably become ubiquitous; technologies to track how many steps have been taken or how much quality sleep was experienced, for example, are now widely available. Research has also shown that the same data captured by wearable sensors can reflect cognitive changes (Collet et al., 2014; Fridman et al., 2018; Shi et al., 2007; Shimomura et al., 2008). Although so far typically demonstrated in controlled laboratory conditions, it seems that we are not so far away from also being able to track our cognitive data as a form of personal informatics (Wilson et al., 2018), in a similar way to which we already track our physical activity. Indeed, a number of wearable technology start-ups aim to track cognitive activity and emotion[4].

Being able to track cognitive activity at work, and to use this data to make improvements to our working habits and lives falls within the areas of neuroergonomics and passive Brain-Computer Interfaces (pBCIs). There is in fact a central research aim within neuroergonomics to assess mental workload using neural measures (Parasuraman, 2011), and in the BCI community, monitoring mental workload at work to develop aids is increasingly popular (Aricò et al., 2018). For now, however, research lacks the insight into how our cognitive activity changes during our every day office-work and life tasks, as ground truth data are harder to observe than for physical actions. Additionally, Arico and colleagues (Aricò et al., 2018) outlined how very few pBCI applications have been tested in real life or realistic settings; due to the complexity of real-life conditions, laboratory settings are unable to take into account the factors that could impair the usability of pBCIs during

real-world applications. Thus, in order to develop our understanding of mental workload in office work, and whether different levels can be discerned in office-like conditions, this work used personalised reading and writing tasks alongside representations of interruptions to increase ecological validity, whilst still maintaining laboratory-style control, as a step towards the measurement of cognitive activity in the wild.

In terms of effort at work, there are several related models, including mental workload (Baddeley, 2000; Sharples and Megaw, 2015; Wickens, 2008), Cognitive Load Theory (Paas et al., 2003), and Mental Effort (Hart and Staveland, 1988). Across these different terminologies, however, mental workload may be described as the amount of mental effort required to complete a task within a limited time period (Maior et al., 2015). In Human-Computer Interaction (HCI), we have seen an increase in research that tries to objectively estimate mental workload, e.g. through deep learning of pupil dilation (Fridman et al., 2018), or through changes in facial temperature (Marinescu, 2018; Marinescu et al., 2018). One promising approach is the measurement of brain activity using functional Near-Infrared Spectroscopy (fNIRS), as a non-invasive and movement tolerable brain scanner (Afergan et al., 2014; Maior et al., 2015, 2018; Pike et al., 2014; Solovey et al., 2009; Yuksel et al., 2016). fNIRS uses near-infrared light to measure changes in blood oxygenation in the brain. Brain activity can be indirectly evaluated from this based on the concept of neurovasuclar coupling in which active brain regions require increased blood flow to meet enhanced energy demands. For understanding the effort involved in everyday work tasks, the prefrontal cortex (PFC) is one area of the brain often measured in mental workload studies (Baddeley, 2012; Foy et al., 2016; Gabrieli et al., 1998; Miller and Cohen, 2001) as it is an area associated with executive functions required for the cognitive processes that mental workload is comprised of.

The current study used fNIRS to examine the measurement of mental workload in more representative work-like tasks, which differs from the more tightly controlled psychology tasks that are most commonly seen in the literature. We used personalised reading and writing material to investigate the ability of fNIRS to detect different mental workload levels. Representations of natural interruptions were incorporated during the tasks to investigate whether fNIRS could detect the effect of interruptions.

Through this work we make the following contributions:

1. We show that fNIRS measurements can differentiate between reading task levels but saw no significant differences between writing levels (despite self-reported differences).
2. We consider fNIRS measurements in terms of spare capacity models to reflect on how interruptions are handled.
3. We provide a deepened understanding of the factors that will be involved in measuring mental workload in real-world environments.

2. Related Work

2.1. Mental Workload Theories

One important characteristic that is presented is limited capacity, meaning that humans can only process a limited amount of information at any one time. Early work by Wickens (Hollands and Wickens, 1999; Wickens, 2008) was fundamental in the development of the concept of mental workload, which described how the cognitive resources necessary for task completion are limited. In an updated model, Sharples and Megaw highlight the relationship between available cognitive resources and task performance (Sharples and Megaw, 2015). When a person has spare capacity of cognitive resources, they may have the ability to take on additional work concurrently; as task demands begin to exceed the amount of cognitive resources a person has available, mental workload levels become overloaded and performance may drop rapidly. This model further acknowledges that at underload, task performance may also drop from lack of attention to the task.

The Multiple Resource Model (Wickens, 2008) extends the notion of limited resources to consider different types of resources and different stages of processing. The theory has three dimensions and posits that encoded information (left dimension) can either be of a spatial (visual) or verbal (auditory) nature, the perception and cognition (processing; middle dimension) of that information can also be spatial or verbal, and the selected responses (right dimension) will either be of a spatial (manual) or verbal (vocal) nature. Multiple tasks may be completed at the same time with good performance only if, across the three dimensions, they do not compete for or overlap in the resources that are required. For example, driving whilst texting can be disastrous because according to the model they are both encoded visually, are spatially processed and require a manual response. Driving whilst having a conversation, however, is theoretically sound because different resources are required over the three dimensions.

2.2. Measuring Mental Workload

There are a variety of subjective and objective methods used for assessing mental workload. Physiological methods to objectively measure mental workload have the advantage of being implemented continuously throughout, and independently of, the task, unlike subjective measures that require participants to self-report their experienced mental workload at intervals or after the task. Physiological measures stand on the assumption that as mental workload levels change, there will be a corresponding response in the autonomic nervous system which can be reflected and measured in a number of physiological parameters. Electrodermal activity (Collet et al., 2014; Shi et al., 2007; Shimomura et al., 2008) and changes in facial temperature (Marinescu, 2018; Marinescu et al., 2018) are examples of physiological measures that have evidence supporting their use as tools for distinguishing between mental workload levels.

Brain imaging techniques as a measure of mental workload are growing rapidly in popularity (Afergan et al., 2014; Ayaz et al., 2013; Maior et al., 2018; Pinti et al., 2018a; Yuksel et al., 2016). Traditionally, participant movement has interrupted the imaging signal, meaning that mental workload data was indistinguishable from movement artefacts. Due to technological progress, imaging methods are now available to enable the investigation of cognition in ecological settings (Dehais et al., 2018). Electroencephalography (EEG) is the most commonly used technique for measuring mental workload (Borghini et al., 2014), but in recent years fNIRS has gained a lot of momentum (Dehais et al., 2018). EEG measures brain activation directly and therefore has high temporal resolution, but has a relatively weak spatial resolution (Herold et al., 2018) and can be susceptible to artefacts (Herold et al., 2018; Pinti et al., 2018). fNIRS, however, has a relatively good spatial and temporal resolution (Herold et al., 2018) and is robust against motion artefacts (Herold et al., 2018; Pinti et al., 2018). As a non-invasive, portable and movement tolerant brain imaging method, fNIRS is perhaps the most effective tool for measuring mental workload *in-the-wild*.

There are many mental workload studies using fNIRS that have taken place in controlled laboratory environments, using tasks such as the n-back task (Ayaz et al., 2012; Fishburn et al., 2014) where mental workload levels can be manipulated by increasing the amount of numbers to memorise. These controlled studies also extend to tasks such as memorising verbally presented information and spatially presented patterns (Maior et al., 2015). Lab studies generally show strong support for fNIRS as a tool for measuring mental workload levels in controlled environments.

2.3. Moving from Controlled Studies to the Real World

For many years, mental workload has been studied in the workplace to ensure that work demands remain within the operators' capabilities in order to prevent poor performance outcomes (Sharples and Megaw, 2015) through task errors or slow completion (Maior et al., 2014).

Considering the promise fNIRS has shown in measuring mental workload in laboratory controlled studies, it could be a potential tool to measure mental workload in the workplace, which could benefit self-reflection and productivity. Cinaz et al. (Cinaz et al., 2013) also suggested that tracking mental workload could prevent a decline in wellbeing.

Dehais et al. (2020) highlight the importance of mental workload measurement in safety-critical jobs. Because of the huge implications to safety, tracking mental workload in safety-critical jobs is a key research area within neuroergonomics and pBCI. Research in the area of safety-critical jobs have used fNIRS as a measurement tool in mental workload studies. The tasks used in these studies can be realistic to workplace tasks. For example, fNIRS could detect different levels of mental workload in participants during a task representative of remotely operated vehicle operational tasks (Durantin et al., 2014). In another study example, certified professional controllers participated in realistic air traffic controller tasks and it was found that fNIRS could detect mental workload levels effectively (Ayaz et al., 2012).

There is research beginning to study mental workload using fNIRS in more uncontrolled environments. For example, fNIRS was effective in distinguishing low and high mental workload levels in pianists playing music pieces (Yuksel et al., 2016), computer programmers comprehending programming languages (Nakagawa et al., 2014), and a table tennis player playing at two levels of difficulty (Balardin et al., 2017). However, research so far does not represent mental workload measurements during standard office work tasks or take into account the factors that could impair mental workload measurements in real-life office-work applications due to the increased complexity of natural settings (Aricò et al., 2018).

In a working environment, being interrupted whilst undertaking a task is often inevitable. Research has suggested that people at work are interrupted four times per hour on average, and the most common form of interruption is verbally face to face (O'Conaill and Frohlich, 1995), though online distractions are ever-more common (Mark et al., 2017). Most research into interruptions has focused on their impact on task performance or completion (Mark et al., 2008), or into delivering interruptions at a timely stage of a task (Bailey and Iqbal, 2008; Iqbal and Bailey, 2008). The effect of interruptions on mental workload levels have not yet been investigated, and thus alongside investigating the sensitivity of fNIRS in measuring mental workload levels during personalised work-like tasks, we investigate whether fNIRS can detect changes in mental workload levels due to realistic verbal interruptions in order to deepen our understanding of the factors involved in tracking mental workload in the real-world. The current research, therefore, is another key stepping stone between controlled laboratory studies and naturalistic studies, which aims to progress research further towards the sustained objective monitoring of mental workload in office workers.

3. Experiment Design

Considering the previous research presented and the features of the present study, the following hypotheses apply:

H1a fNIRS will detect differences in brain activity between conditions that correspond to different mental workload levels of a reading task.

H1b fNIRS will detect differences in brain activity between conditions that correspond to different mental workload levels of a writing task.

H2 fNIRS will detect changes in brain activity corresponding to interruptions.

3.1. Participants

20 healthy participants took part in the study (8 females and 12 males, aged 31 ± 9.57). Opportunity sampling was used to recruit participants and each participant provided written and informed consent. Participants were eligible for participation if their work included typical office-like tasks, such as reading professional documents. The experiment was approved by the School's ethics board (approval ID: CS-2017-R13) and participants were provided with a £10 Amazon voucher as an inconvenience allowance.

3.2. Tasks and Conditions

There was a reading and writing task, each with three conditions of an 'easy,' 'medium,' and 'hard' difficulty designed to require corresponding levels of mental workload. The easy and medium conditions involved personalised materials, and the hard conditions were a continuation of the medium conditions with an addition of a secondary task designed to overload participants' mental workload according to the Multiple Resource Model.

3.2.1. Reading task.

The easy condition for the reading task involved participants reading basic material related to their area of research, work, or study. The medium condition involved reading a previously-unread academic journal article, also relevant to participants' individual areas. For the hard condition, participants continued reading the materials from the medium condition whilst completing a secondary task that competed for the same cognitive resources according to the Multiple Resource Model — this involved counting the amount of times the word 'the' was read. This secondary task was the same for all participants and aimed to represent tasks in which one is searching for words or information whilst reading.

The difficulty of the reading materials was formally assessed using the Flesch-Kincaid Grade Level and Flesch Reading Ease scores (Flesch, 1948). These measures are based upon word length and sentence length and can assess how difficult a piece of text is to read. The easy condition materials were at least 2 levels below the medium condition materials material in both the Flesch-Kincaid Grade Level and the Flesch Reading Ease measures to ensure a definite difference in task demands. All reading materials were presented in an identical format with images removed to reduce the effect of confounding variables.

3.2.2. Writing Task

All writing conditions were conducted in an email format which was addressed to the experimenter, and the difficulty of the conditions was based upon the assumption that an increased amount of required cognitive processes positively correlates with task demand. For the easy condition participants were asked to *"Describe the tasks that you have done so far in this experiment in some detail"*. This was designed to require retrospectively recent memory.

The premise for both the medium and hard conditions were to *"Pretend I have emailed you asking about your area of research. I'm interested in what you research, how you research it and why you research this area. Please reply to that email. You can assume I have a basic but limited knowledge of your field so you will need to explain certain terms to me. I also mentioned that I would be interested in meeting with you to discuss your research."* Wording was altered slightly when required to be relevant to participants.

The medium condition required participants to start by outlining some real days and times that they were available to meet this week and then talk about their research. This required retrospective memory, short term prospective memory and working memory to remember the vast amount of information that was provided. The hard condition required participants to continue with the task as well as outlining some real days and times they would be available to meet the next week. The secondary task for the hard condition involved participants saying 'blah' repeatedly out loud whilst writing. This condition was designed to require retrospective memory, longer term prospective memory and

working memory, whilst completing a difficult secondary task that competed for the same resources according to the Multiple Resource Model. This secondary task was the same for all participants and aimed to represent the notion of speaking whilst working.

3.2.3. Interruptions

Verbal interruptions involved the experimenter briefly disrupting the condition with generic conversation, and were added to 3 out of the 6 conditions, counter-balanced between participants.

To further increase ecological validity and make the study environment less controlled, all participants were provided with a drink (coffee, tea or water - if a drink was declined, water was provided on the desk). Drink consumption was permitted as desired except during the baseline conditions.

3.2.4. Baseline Conditions

A fixation cross presented for 1 minute on a monitor was used as a baseline condition at the beginning of both the reading and writing task. Before each further task condition, the fixation cross was used again for 1 minute to allow brain activity to return to baseline.

3.3. Setup and Procedure

A standard office set-up was created. Participants sat behind an office desk and in front of a computer monitor and keyboard. A non-transparent board was placed between the participant and experimenter to provide a sense of open-plan office form of semi-privacy. In order to identify the times at which the verbal interruptions occurred, a GoPro Hero4 Silver placed in an audible protective case was used. This was placed inconspicuously behind participants and recorded their frame and monitor.

3.3.1. Study Procedure.

Participants were first provided with a drink and seated at the desk where the study instructions were given and informed consent was provided. The fNIRS device was fitted and the GoPro was started.

The reading task was completed first due to the writing task being partially based on the reading task. When the study started, participants first stared at the fixation cross for 1 minute before the easy reading condition began. The condition lasted for 5 minutes, and when time was up participants immediately filled out a NASA-TLX questionnaire. After the questionnaire was completed, the fixation cross appeared again for 1 minute before the next reading condition began. The order of the medium and hard reading conditions were counterbalanced across participants and all conditions lasted for 5 minutes. Once the second condition was completed, the NASA-TLX questionnaire was administered immediately again, followed by the 1 minute fixation cross and the final reading condition which was again followed by the NASA-TLX questionnaire. After the reading conditions, the writing conditions started and followed the same format as the reading conditions.

3.4. Measurements and Data Analysis

3.4.1. Mental workload questionnaire.

A NASA-TLX workload questionnaire (Hart, 2006) was used to collect subjective mental workload information. The self-assessed questionnaire comprises of six 21 point scales where a rating of 0 equates to 'Very Low' and 20 to 'Very High'. The scales include mental demand, physical demand, temporal demand, performance, effort, and frustration. Friedman tests were run to investigate whether there was a significant difference in ratings between the easy, medium, and hard reading and writing tasks; post hoc analysis was conducted with Wilcoxon signed-ranks with a Bonferroni correction applied.

3.4.2. fNIRS Measurements

A wireless fNIRS device (Octamon, Artinis Medical Systems) with 8

channels with a source-detector distance of 3.5cm measured oxygenated (O2Hb) and deoxygenated (HHb) haemoglobin across the PFC, as show in Fig. 1. The wavelengths used were 760 and 850nm with a differential pathlength factor of 6 and a sampling rate of 10Hz. Raw data was exported to Homer2 fNIRS processing package (Huppert et al., 2009). Data was converted into changes in optical density, and motion artifacts were corrected using a Wavelet filter (iqr=1.5) and a bandpass filter (0.5 LPF and 0.01 HPF). Physiological noise was reduced using a Principal Component Analysis (nSV=0.97) and concentration changes in O2Hb and HHb were calculated using the Modified Beer-Lambert Law.

Baseline correction was performed by subtracting baseline mean values from the task data; the first baseline condition from both the reading and writing tasks were used and subtracted from their respective tasks. To test whether there were significant differences in brain activity between the easy, medium, and hard difficulty reading and writing tasks, one-way repeated measures ANOVAs were conducted using the average values from the first 2.5 minutes of each condition before any interruption occurred.

For the interruption analysis, the interruption timings were marked down from the video footage and added as stimuli in Homer2. Data was baseline corrected in the same way as the task data. To compare brain activity during the task compared to during the interruptions, paired t-tests were used to compare 10 seconds of interruption data against the previous 10 seconds of task data for the participants that were interrupted in the same conditions as the interruptions. All interruption stimuli were shifted by 2 seconds after the onset of the interruptions to account somewhat for the temporal delay of fNIRS measurements.

When a region of the brain becomes activated, cerebral blood flow increases to meet the increase in oxygen demand; this is known as the hemodynamic response and is reflected by an increase in O2Hb and a decrease in HHb (Scholkmann et al., 2014). Measurements of O2Hb alone are vulnerable to physiological noise which risks false conclusions about neural activity being drawn (Herold et al., 2017; Pinti et al., 2018a). Whilst measurements of HHb are less affected by these confounds (page 364) and most highly correlated with other brain imaging methods (Huppert et al., 2006) the strongest indicator of functional brain activity is when there is an increase in O2Hb corresponding with a decrease in HHb (Herold et al., 2017), and thus this is what will be our main focus in the results.

Data from 2 male participants were excluded due to technical difficulties making the final analysis include data from 18 participants. All fNIRS measurements are reported in micromoles (μM). Post-hoc analysis for the fNIRS data was conducted using a Bonferroni correction.

4. Results

4.1. NASA-TLX ratings

Figures 2 and 3 display mean subjective scores for the Mental Demand, Effort and Performance subscales for reading and writing respectively.

A Friedman test revealed a significant effect of condition on mental demand ratings for the reading task ($\chi^2(2)$=18.250, p<0.001). Post hoc analysis with Wilcoxon tests was conducted with a Bonferroni correction applied which resulted in a p=0.017 significance value for all NASA-TLX post hoc analyses. This showed that the easy and medium reading conditions were rated significantly lower than the hard reading condition (Z=-3.335, p<0.001 and Z=-2.519, p=0.012 respectively) but did not significantly differ to each other. Mental demand ratings also showed significance in the writing task ($\chi^2(2)$=14.464, p<0.001), where the easy and medium conditions were both rated lower than the hard condition (Z=-3.463, p<0.001 and Z=-2.872, p=0.004 respectively) but did not significantly differ to each other.

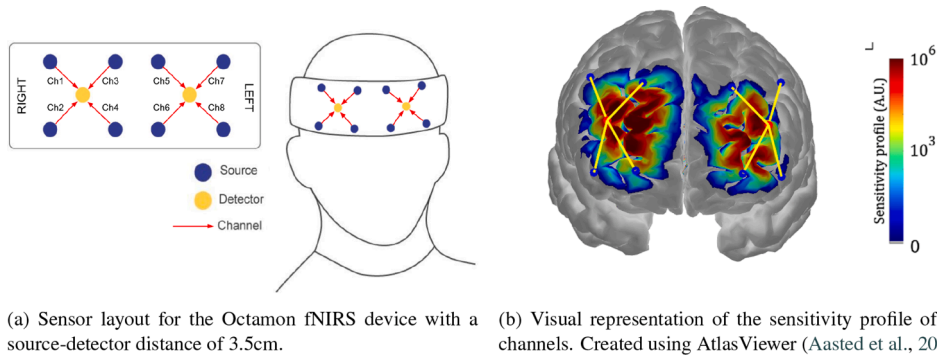There was a significant effect of condition on physical demand rat-

(a) Sensor layout for the Octamon fNIRS device with a source-detector distance of 3.5cm.

(b) Visual representation of the sensitivity profile of the channels. Created using AtlasViewer (Aasted et al., 2015).

Fig. 1. Sensor placement and sensitivity (Aasted, Yücel, Cooper, Dubb, Tsuzuki, Becerra, Petkov, Borsook, Dan, Boas, 2015).

ings for the reading task ($\chi^2(2)$=17.211, p<0.001). Wilcoxon tests showed that the easy and medium reading conditions were rated significantly lower than the hard condition (Z=-2.958, p=0.003 and Z=-2.534, p=0.011 respectively) but were not rated significantly different to each other. Physical demand was not significant for the writing task. There was also no significant effect of condition on temporal demand or performance ratings for the reading or writing tasks.

A significant effect of condition on effort ratings for the reading task was found ($\chi^2(2)$=25.054, p<0.001). Post hoc analysis revealed that the easy and medium conditions were rated as significantly lower than the hard conditions (Z=-3.830, p<0.001 and Z=-3.604, p<0.001 respectively), but again did not significantly differ to each other. The Friedman test was also significant for the writing task ($\chi^2(2)$=22.243, p<0.001). Wilcoxon tests revealed that the easy and medium conditions were rated lower than the hard difficulty condition (Z=-3.428, p<0.001 and Z=-3.732, p<0.001 respectively) but did not differ to each other.

A significant effect of condition on frustration ratings for the reading task was revealed ($\chi^2(2)$=18.613, p<0.001). The post hoc analysis showed that the easy and medium conditions were rated significantly lower than the hard condition (Z=-3.416, p<0.001 and Z=-3.316, p<0.001 respectively), but not significantly different to each other. Effort was also significant for the writing task ($\chi^2(2)$=15.254, p<0.001), where the easy and medium conditions were rated significantly lower than the hard condition (Z=-3.157, p=0.002 and Z=-3.465, p<0.001 respectively) but again were not significantly different to each other.

4.2. fNIRS data

Condition analysis On an individual channel basis, only channel 7

HHb (see Figure 1) showed a significant ANOVA result (F(2, 34)=4.258, p=0.022), where means showed that reading medium had the lowest brain activity, followed by reading easy, and reading hard had the highest brain activity, though this did not reach significance in the Bonferroni post hoc test.

When averaging across the different sides of the PFC, the ANOVA showed a significant reading result for channels 5-8 (left side of the PFC) for both O2HB (F(2, 34)=3.400, p=0.045) and HHb (F(2, 34)=3.425, p=0.044). For both O2Hb and HHb, the means showed that like channel 7, reading medium had the lowest brain activity followed by reading easy, and reading hard had the highest brain activity (Figure 4), though Bonferroni correction did not reach significance.

No writing results reached significance in the condition analysis.

Interruption analysis On an individual channel level, the paired t-tests showed that for channel 6, those who were interrupted during the reading easy task (n=8) had significantly higher O2Hb brain activity during the task compared to during the interruption (t(7)=3.119, p=0.017). On the contrary, channel 7 showed significantly more HHb (indicating less brain activation) during the reading easy task compared to during the interruption (t(7)=2.525, p=0.040) (Figure 5a). Channel 7 also showed significantly less HHb levels (more brain activity) for participants (n=8) during the writing hard task compared to during the writing hard interruptions (t(7)=2.749, p=0.029).

When considering the averages for the sides of the PFC, channels 1-4 (right side of the PFC) for both O2Hb and HHb were significant for writing hard (t(7)=2.496, p=0.041 and t(7)=2.514, p=0.040 respectively). Both the O2Hb and HHb showed, like channel 7, more brain activity during the writing hard task compared to during the interruption during the writing hard task (Figure 5b).
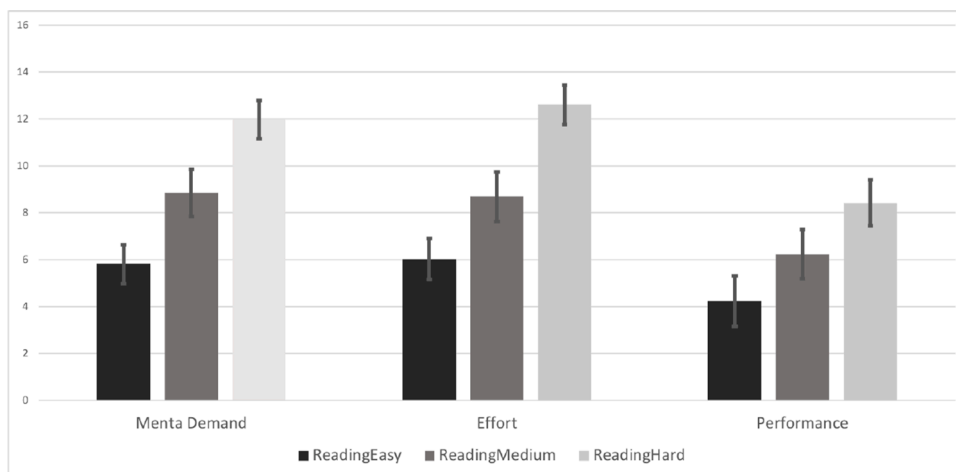

Fig. 2. Average NASA-TLX ratings across the reading task conditions for Mental Demand, Effort and Performance sub scales. All error bars represent standard error of the mean.
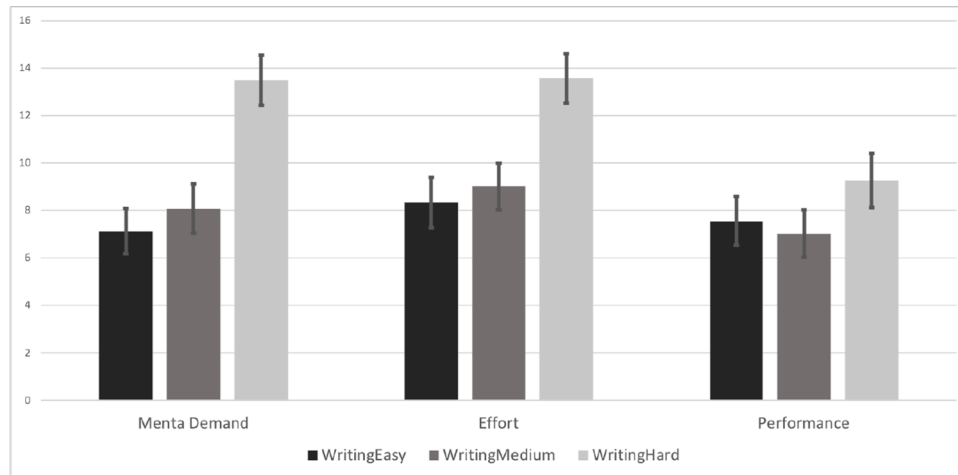
Fig. 3. Average NASA-TLX ratings across the writing task conditions for Mental Demand, Effort and Performance sub scales.
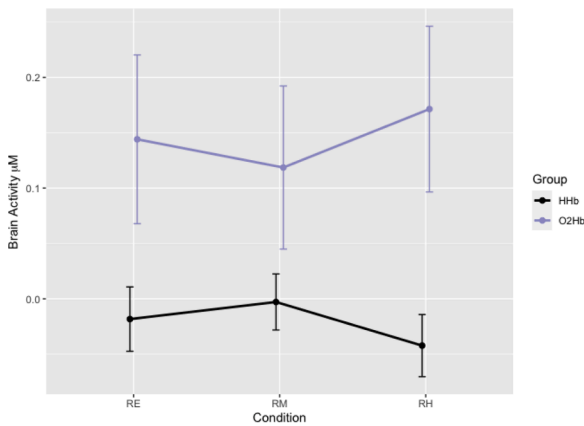

Fig. 4. fNIRS results for the reading task show a mirroring pattern of increased O2Hb and decreased HHb during the hard difficulty condition - Channels 5-9.

5. Discussion

The study aimed to investigate the measurement of mental workload using fNIRS in tasks and environments more relevant to the workplace. The study can be seen as a stepping stone between tightly controlled lab studies and unconstrained real-world studies, and the findings can provide guidance for progressing research into workplace environments.

5.1. The Sensitivity of fNIRS

Results showed support for the ability of fNIRS to detect mental workload levels in reading tasks (H1a) but not writing tasks (H1b). For the reading task, fNIRS could detect significant differences between conditions in the left side of the PFC and the results aligned with the subjective ratings of mental workload in terms of the hard condition showing the highest mental workload. Subjective ratings of mental workload showed no significant differences between the easy and medium conditions, and fNIRS measurements showed only small differences between these conditions compared to the hard condition, also supporting the sensitivity of fNIRS (H1a).

The reading results are supported by the findings from other studies. Because of the different sizes of participant's heads and the fixed optode locations on the fNIRS device, we can broadly assume which channels correspond to the different areas of the PFC via the 10-20 electrode system from the EEG field (Herold et al., 2017; Jasper, 1958). Reading comprehension is heavily associated with the PFC (as well as temporal regions), namely Broca's area located in the left inferior frontal gyrus, and thus the consensus is for left hemispheric dominance in reading comprehension (Baretta et al., 2012; Paquette et al., 2015; Price, 2012). Our results were significant for channel 7, which is in the left hemisphere and where Broca's area is expected to be located (Vergotte et al., 2018), in line with the areas of activation one would expect to find. The results were also significant for channels 5-8, which is the left hemisphere, further corresponding to previous literature, including fNIRS studies into reading comprehension (Dieler et al., 2012).

As there was strong evidence for fNIRS detecting differences in
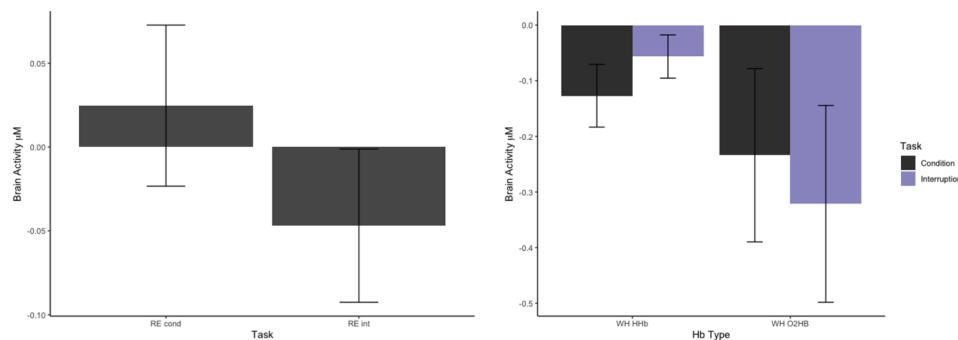
(a) fNIRS results show a drop in HHb during the interruption of reading easy compared to task alone - Channel 7.

(b) fNIRS results show decreased O2Hb and increased HHb during interruptions of the hard difficulty writing task - Channel 1-4.

Fig. 5. fNIRS measurements during task completion compared to during the interruption periods.

mental workload levels between conditions for channels 5-8 with significant results for both O2Hb and HHb, it seems sensible to infer that averaging across channels captured the activation area across participants, which may not have been captured fully by channel 7 if it didn't exactly correspond to the inferior frontal gyrus in all participants.

Regarding the non-significant post hoc tests of the ANOVA, Bonferroni is a conservative test as it attempts to control the overall alpha level. As ANOVA results for both O2Hb and HHb were significant, we consider this as strong supporting evidence for fNIRS detecting differences (Pinti et al., 2018b) between reading difficulties, and hence our focus is on this global effect.

As means for brain activity consistently showed the medium condition corresponding to slightly less activity than the easy condition, and non-significant differences in subjective ratings between these conditions, this is an interesting finding for the challenges associated with objectively tracking of mental workload in the workplace (and beyond). Even if task demands were harder for the medium condition, it appears that participant feelings could have impacted on the results. As the medium condition had more stimulating materials, it is expected that participants felt more engaged with the task. Indeed, previous work found that mental workload ratings were lower for demanding tasks when participants were more engaged (Horrey et al., 2009). Similarly, Lukanov et al. found that, for an insurance claim form, participants preferred the user interface condition that objectively and subjectively generated the highest levels of mental workload (Lukanov et al., 2016), and thus emotional factors could be a challenge for objectively measuring mental workload in the workplace.

In contrast to reading, fNIRS did not detect significant differences between conditions in the writing task, not aligning with the subjective results and not supporting H1b. Writing organisation in the brain appears to represent a complex human function that involves several language sub-components, and thus localisation in the brain is highly individualised between people (Lubrano et al., 2004). Nevertheless, writing localisation is thought to heavily involve the frontal lobe (and the anterior parietal lobe), more specifically the posterior part of the middle and superior frontal gyri (Exner's area, Katanoda et al., 2001; Lubrano et al., 2004). The fNIRS device in the current study measured isolated activation in the PFC, meaning activation from the 'writing centre' was not covered. This finding, where different levels of mental workload for the writing task were not distinguished in the PFC, does not only hold relevance for the current study, but also challenges perceptions from the wider HCI field. FNIRS studies of mental workload most often measure cognitive activity from the PFC (Maior et al., 2015, 2018; Pike et al., 2014; Solovey et al., 2009, 2012; Yuksel et al., 2016); there is a consensus that mental workload will consistently be exhibited and measurable in the PFC. However, the results from the writing task suggest that the processing involved when task demand increased and subsequent mental workload increased between the hard and the easy and medium conditions (as found in the subjective ratings) may not have been measurable from the PFC. As the writing task required a complex amount of neural processing and cognitive processes, it seems that it was not possible to capture the full picture of mental workload level from the PFC alone. Working memory cognitive load is measurable from the PFC (Fishburn et al., 2014; Tomasi et al., 2007), but the writing conditions might not have differed significantly from each other in the working memory aspect of the writing task like we intended with the study design, and the combination of cognitive processes that increased mental demand for the hard condition might have been localised outside of the PFC. This is supported by some of our recent work (Argyle et al., 2021) which did not find differences in oxygenation in the PFC for a visual search task despite significantly different subjective mental workload ratings; here, mental workload might have been detectable in the occipital and parietal lobes (Tomasi et al., 2007), and did not involve the PFC enough to detect objective changes in mental workload level. Thus, future studies of realistic or real-world tasks should consider carefully which brain areas mental workload might be most represented

in, and perhaps measure multiple lobes to gather a richer insight into brain activation as these types of task come with more cognitive complexities compared to laboratory studies (Aricò et al., 2018). Future studies could also benefit from further investigating varying levels of mental workload and specific cognitive task demands for tasks and activities relevant to daily life. This perhaps would also explain why papers using machine learning to classify mental workload levels (Chan et al., 2012; Herff et al., 2014; Sassaroli et al., 2008) typically achieve fairly low levels of accuracy. In support, it has been shown that measuring a larger neural area resulted in a higher accuracy of mental workload classification (Saadati et al., 2019). That being said, the PFC may more often than not provide a reasonable insight into mental workload levels due to its involvement in a wide variety of tasks. It is notable that the reading results support the finding that changes in mental workload for reading are detectable in the PFC, where more complex reading tasks are associated with increased neural activity (Just et al., 1996; 2003).

A limitation of the study could be that the data was analysed in 2.5-minute blocks which is likely to contain more artefacts compared to shorter trials (Zhu et al., 2020) that are often seen in laboratory studies. A main research area in pBCIs and neuroergonomics is to use fNIRS to measure mental workload of workers in safety-critical jobs and to develop aids to improve the safety of these jobs based on their mental state. Our aim is related to this, in the sense that we wish to investigate mental workload levels in office workers and this data might progress to aid improvements to working habits and lives. These types of research aims (in addition to many other research areas using fNIRS) require long-time continuous monitoring of brain activity, which is an acknowledged advantage of fNIRS (Pinti et al., 2018). With rigorous processing of the data, the noise should not impact the validity of the data, and longer blocks may reveal a more representative picture of brain activity compared to a short snapshot. We did opt to shorten the analysis from 5 minutes of data to 2.5 minutes of data with the aim of somewhat ensuring control over the data considering the study was still essentially lab based.

5.2. Interruptions

In terms of the effect of verbal interruptions on fNIRS measurements (H2), there was less brain activity in the right side of the PFC during the interruption of the writing hard condition, compared to before the interruption, and this was significant for both O2Hb and HHb. If we consider this in relation to spare capacity (Sharples and Megaw, 2015), as the hard writing condition was subjectively rated as requiring the most mental workload, there might not have been enough cognitive resources available to take on the interruption concurrently with the primary task. This means the interruption would have become a primary task which was less demanding than the hard writing condition. Whereas for the easy reading condition, HHb levels (which are interpreted a bit more cautiously alone) showed an increase in brain activity during the interruption compared to during the task. In relation to the model, this could be explained by the notion that there was spare capacity during the easy reading condition which meant that responding to the interruption could have been achieved through multi-tasking, which increased mental demand and hence mental workload. Observationally, we note that verbal responses to interruptions during the reading easy conditions tended to be briefer and more distractible, whereas in the writing hard participants could not keep saying 'blah' *and* respond to the interruption, so they would pause the task to give a proper response.

Even though results were only significant for two of our conditions, perhaps due to the small number of participants for each interruption, they do contribute to a deeper understanding of the factors involved in measuring mental workload objectively in the workplace, which are not encountered during controlled lab studies. The results emphasise that changes in mental workload levels are not 'black and white'; instead they often depend on situational factors which might be different from

one person to the next (Sharples, 2018). This also means that it is difficult to evaluate the effectiveness of objective measures because determining changes in mental workload in these situations depends on subjective interpretation. Objective measures might not reflect the interpretation made about a person's mental workload but that does not necessarily mean that the measurement is wrong. Such variability in factors contributing to mental workload levels and the different responses between people could mean that future studies might benefit from considering results on a participant by participant basis. Additionally, whilst we have incorporated natural interruptions into the study design to gain an understanding of challenges to do with subtle within task variations of mental workload whilst completing a single task, workplaces are increasingly made up of multi-tasking activities (Mark et al., 2005), which most likely means that objectively measuring mental workload in the workplace will come with increased complexity. It should be noted that the sample size was limited for the interruption analysis; with more data to analyse the statistical power and validity would have been increased. We do believe our results have opened an interesting area for discussion and further research area on the effect of interruptions on mental workload levels and how these measurements can be dealt with in real-world settings.

We further increased ecological validity by including uncontrolled drink consumption. Because this data was messy, due to participants drinking at different frequencies and during different conditions, it was not possible to analyse the effect of drink consumption on brain activity. This type of interruption data, however, could prove to be valuable if incorporated plausibly in future studies, as it could provide insight into whether different types of interruptions seem to follow the same trend in which data can be understood in relation to spare capacity models, or whether different factors need to be considered.

Finally, continuing to bridge the gap between controlled lab studies and real-world studies is hugely important. Ladouce suggested that true cognition and its complexities can often only be understood correctly when examined in ecologically valid environments (Ladouce et al., 2017), and mental workload appears to belong to this category. Progressing this research to in-the-wild studies of mental workload will enable further understanding of the research and challenges associated with the sustained measurement of mental workload in the workplace with fNIRS as a potential candidate. Our own future work aims to make progress towards being able to track cognitive activity as a form of personal informatics.

6. Conclusions

Through using personalised tasks and verbal interruptions in a workplace-like setting, we show that fNIRS placed over the PFC alone was able to detect the differences in mental workload experienced by participants during personalised reading tasks, but was not sensitive to the reported differences in writing tasks. This finding could be due to the PFC not exhibiting mental workload levels for all tasks, and thus careful consideration of optode placement over a larger region of interest is emphasised for naturalistic studies especially. Verbal interruptions appeared to cause within task mental workload variation, causing increased load if done in parallel with tasks and decreased load if becoming the primary task temporarily. These findings demonstrated the complexity of mental workload as concept that is non-quantifiable and often affected by situational factors reliant on interpretation. With the goal in mind of the sustained objective monitoring of mental workload in the workplace for self-improvements, further work is needed to establish the sensitivity of fNIRS for further work-like tasks and to build on understanding of the factors involved in these measurements.

CRediT authorship contribution statement

Serena Midha: Conceptualization, Methodology, Investigation, Formal analysis, Data curation, Visualization, Writing - original draft, Writing - review & editing. **Horia A. Maior:** Conceptualization, Methodology, Visualization, Writing - original draft. **Max L. Wilson:** Conceptualization, Methodology, Resources, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Sarah Sharples:** Conceptualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

Aasted, C.M., Yücel, M.A., Cooper, R.J., Dubb, J., Tsuzuki, D., Becerra, L., Petkov, M.P., Borsook, D., Dan, I., Boas, D.A., 2015. Anatomical guidance for functional near-infrared spectroscopy: Atlasviewer tutorial. Neurophotonics 2 (2), 020801.

Afergan, D., Peck, E.M., Solovey, E.T., Jenkins, A., Hincks, S.W., Brown, E.T., Chang, R., Jacob, R.J., 2014. Dynamic difficulty using brain metrics of workload. Proceedings of the 2014 CHI Conference on Human Factors in Computing Systems. ACM, pp. 3797–3806.

Argyle, E.M., Marinescu, A., Wilson, M.L., Lawson, G., Sharples, S., 2021. Physiological indicators of task demand, fatigue, and cognition in future digital manufacturing environments. International Journal of Human-Computer Studies 145, 102522.

Aricò, P., Borghini, G., Di Flumeri, G., Sciaraffa, N., Babiloni, F., 2018. Passive BCI beyond the lab: current trends and future directions. Physiological measurement 39 (8), 08TR02.

Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P.A., McKendrick, R., Parasuraman, R., 2013. Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. Frontiers in human neuroscience 7, 871.

Ayaz, H., Shewokis, P.A., Bunce, S., Izzetoglu, K., Willems, B., Onaral, B., 2012. Optical brain monitoring for operator training and mental workload assessment. Neuroimage 59 (1), 36–47.

Baddeley, A., 2000. The episodic buffer: a new component of working memory? Trends in cognitive sciences 4 (11), 417–423.

Baddeley, A., 2012. Working memory: Theories, models, and controversies. Annual review of psychology 63, 1–29.

Bailey, B.P., Iqbal, S.T., 2008. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. ACM Transactions on Computer-Human Interaction (TOCHI) 14 (4), 21.

Balardin, J.B., Zimeo Morais, G.A., Furucho, R.A., Trambaiolli, L., Vanzella, P., Biazoli Jr, C., Sato, J.R., 2017. Imaging brain function with functional near-infrared spectroscopy in unconstrained environments. Frontiers in human neuroscience 11, 258.

Baretta, L., Tomitch, L.M.B., Lim, V.K., Waldie, K.E., 2012. Investigating reading comprehension through EEG. Ilha do Desterro: A Journal of English Language, Literatures in English and Cultural Studies (63), 69–99.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., Babiloni, F., 2014. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neuroscience & Biobehavioral Reviews 44, 58–75.

Chan, J., Power, S., Chau, T., 2012. Investigating the need for modelling temporal dependencies in a brain-computer interface with real-time feedback based on near infrared spectra. Journal of Near Infrared Spectroscopy 20 (1), 107–116.

Cinaz, B., Arnrich, B., Marca, R., Tröster, G., 2013. Monitoring of mental workload levels during an everyday life office-work scenario. Personal and ubiquitous computing 17 (2), 229–239.

Collet, C., Salvia, E., Petit-Boulanger, C., 2014. Measuring workload with electrodermal activity during common braking actions. Ergonomics 57 (6), 886–896.

Dehais, F., Dupres, A., Di Flumeri, G., Verdiere, K., Borghini, G., Babiloni, F., Roy, R., 2018. Monitoring pilot's cognitive fatigue with engagement features in simulated and actual flight conditions using an hybrid fNIRS-EEG passive BCI. 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, pp. 544–549.

Dehais, F., Lafont, A., Roy, R., Fairclough, S., 2020. A neuroergonomics approach to mental workload, engagement and human performance. Frontiers in Neuroscience 14, 268.

Dieler, A., Tupak, S., Fallgatter, A., 2012. Functional near-infrared spectroscopy for the assessment of speech related tasks. Brain and Language 121 (2), 90–109.

Durantin, G., Gagnon, J.-F., Tremblay, S., Dehais, F., 2014. Using near infrared spectroscopy and heart rate variability to detect mental overload. Behavioural brain research 259, 16–23.

Fishburn, F.A., Norr, M.E., Medvedev, A.V., Vaidya, C.J., 2014. Sensitivity of fNIRS to cognitive state and load. Frontiers in human neuroscience 8, 76.

Flesch, R., 1948. A new readability yardstick. Journal of applied psychology 32 (3), 221.

Foy, H.J., Runham, P., Chapman, P., 2016. Prefrontal cortex activation and young driver behaviour: a fNIRS study. PLoS one 11 (5), e0156512.

Fridman, L., Reimer, B., Mehler, B., Freeman, W.T., 2018. Cognitive load estimation in the wild. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, p. 652.

Gabrieli, J.D., Poldrack, R.A., Desmond, J.E., 1998. The role of left prefrontal cortex in language and memory. Proceedings of the national Academy of Sciences 95 (3), 906–913.

Hart, S.G., 2006. NASA-task load index (NASA-TLX); 20 years later. Proceedings of the human factors and ergonomics society annual meeting, 50. Sage publications Sage CA: Los Angeles, CA, pp. 904–908.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research. Advances in psychology 52, 139–183.

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., Schultz, T., 2014. Mental workload during n-back task-quantified in the prefrontal cortex using fNIRS. Frontiers in human neuroscience 7, 935.

Herold, F., Wiegel, P., Scholkmann, F., Müller, N.G., 2018. Applications of functional near-infrared spectroscopy (fNIRS) neuroimaging in exercise–cognition science: a systematic, methodology-focused review. Journal of clinical medicine 7 (12), 466.

Herold, F., Wiegel, P., Scholkmann, F., Thiers, A., Hamacher, D., Schega, L., 2017. Functional near-infrared spectroscopy in movement science: a systematic review on cortical activity in postural and walking tasks. Neurophotonics 4 (4), 041403.

Hollands, J.G., Wickens, C.D., 1999. Engineering psychology and human performance. Journal of surgical oncology.

Horrey, W.J., Lesch, M.F., Garabet, A., 2009. Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task conditions. Journal of safety research 40 (1), 7–12.

Huppert, T.J., Diamond, S.G., Franceschini, M.A., Boas, D.A., 2009. Homer: a review of time-series analysis methods for near-infrared spectroscopy of the brain. Applied optics 48 (10), D280–D298.

Huppert, T.J., Hoge, R.D., Diamond, S.G., Franceschini, M.A., Boas, D.A., 2006. A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. Neuroimage 29 (2), 368–382.

Iqbal, S.T., Bailey, B.P., 2008. Effects of intelligent notification management on users and their tasks. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 93–102. https://doi.org/10.1145/1357054.1357070.

Jasper, H., 1958. Report of the committee on methods of clinical examination in electroencephalography. Electroencephalogr Clin Neurophysiol 10, 370–375.

Just, M.A., Carpenter, P.A., Keller, T.A., Eddy, W.F., Thulborn, K.R., 1996. Brain activation modulated by sentence comprehension. Science 274 (5284), 114–116.

Just, M.A., Carpenter, P.A., Miyake, A., 2003. Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. Theoretical Issues in Ergonomics Science 4 (1-2), 56–88.

Katanoda, K., Yoshikawa, K., Sugishita, M., 2001. A functional MRI study on the neural substrates for writing. Human brain mapping 13 (1), 34–42.

Ladouce, S., Donaldson, D.I., Dudchenko, P.A., Ietswaart, M., 2017. Understanding minds in real-world environments: toward a mobile cognition approach. Frontiers in human neuroscience 10, 694.

Lubrano, V., Roux, F.-E., Démonet, J.-F., 2004. Writing-specific sites in frontal areas: a cortical stimulation study. Journal of neurosurgery 101 (5), 787–798.

Lukanov, K., Maior, H.A., Wilson, M.L., 2016. Using fNIRS in usability testing: understanding the effect of web form layout on mental workload. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, pp. 4011–4016.

Maior, H.A., Pike, M., Sharples, S., Wilson, M.L., 2015. Examining the reliability of using fNIRS in realistic HCI settings for spatial and verbal tasks. Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems. ACM, pp. 3807–3816.

Maior, H.A., Pike, M., Wilson, M.L., Sharples, S., 2014. Continuous detection of workload overload: an fNIRS approach. Contemporary Ergonomics and Human Factors 2014: Proceedings of the international conference on Ergonomics & Human Factors 2014, Southampton, UK, 7-10 April 2014. CRC Press, p. 450.

Maior, H.A., Wilson, M.L., Sharples, S., 2018. Workload alerts-using physiological measures of mental workload to provide feedback during tasks. ACM Transactions on Computer-Human Interaction (TOCHI) 25 (2), 9.

Marinescu, A.C., 2018. Facial thermography for assessment of workload in safety critical environments. University of Nottingham. Ph.D. thesis.

Marinescu, A.C., Sharples, S., Ritchie, A.C., Sánchez López, T., McDowell, M., Morvan, H. P., 2018. Physiological parameter response to variation of mental workload. Human factors 60 (1), 31–56.

Mark, G., Gonzalez, V.M., Harris, J., 2005. No task left behind?: examining the nature of fragmented work. Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, pp. 321–330.

Mark, G., Gudith, D., Klocke, U., 2008. The cost of interrupted work: more speed and stress. Proceedings of the SIGCHI conference on Human Factors in Computing Systems. ACM, pp. 107–110.

Mark, G., Iqbal, S., Czerwinski, M., 2017. How blocking distractions affects workplace focus and productivity. Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers. ACM, pp. 928–934.

Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annual review of neuroscience 24 (1), 167–202.

Nakagawa, T., Kamei, Y., Uwano, H., Monden, A., Matsumoto, K., German, D.M., 2014. Quantifying programmers' mental workload during program comprehension based on cerebral blood flow measurement: a controlled experiment. Companion Proceedings of the 36th International Conference on Software Engineering. ACM, pp. 448–451.

O'Conaill, B., Frohlich, D., 1995. Timespace in the workplace: Dealing with interruptions. Conference companion on Human factors in computing systems. ACM, pp. 262–263.

Paas, F., Renkl, A., Sweller, J., 2003. Cognitive load theory and instructional design: Recent developments. Educational psychologist 38 (1), 1–4.

Paquette, N., Lassonde, M., Vannasing, P., Tremblay, J., González-Frankenberger, B., Florea, O., Béland, R., Lepore, F., Gallagher, A., 2015. Developmental patterns of expressive language hemispheric lateralization in children, adolescents and adults using functional near-infrared spectroscopy. Neuropsychologia 68, 117–125.

Parasuraman, R., 2011. Neuroergonomics: Brain, cognition, and performance at work. Current directions in psychological science 20 (3), 181–186.

Pike, M.F., Maior, H.A., Porcheron, M., Sharples, S.C., Wilson, M.L., 2014. Measuring the effect of think aloud protocols on workload using fNIRS. Proceedings of the 32nd annual ACM conference on Human factors in computing systems. ACM, pp. 3807–3816.

Pinti, P., Aichelburg, C., Gilbert, S., Hamilton, A., Hirsch, J., Burgess, P., Tachtsidis, I., 2018a. A review on the use of wearable functional near-infrared spectroscopy in naturalistic environments. Japanese Psychological Research 60 (4), 347–373.

Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., Tachtsidis, I., 2018b. Current status and issues regarding pre-processing of fnirs neuroimaging data: An investigation of diverse signal filtering methods within a general linear model framework. Frontiers in human neuroscience 12, 505.

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., Burgess, P.W., 2018. The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. Annals of the New York Academy of Sciences.

Price, C.J., 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage 62 (2), 816–847.

Saadati, M., Nelson, J., Ayaz, H., 2019. Convolutional neural network for hybrid fNIRS-EEG mental workload classification. International Conference on Applied Human Factors and Ergonomics. Springer, pp. 221–232.

Sassaroli, A., Zheng, F., Hirshfield, L.M., Girouard, A., Solovey, E.T., Jacob, R.J., Fantini, S., 2008. Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy. Journal of Innovative Optical Health Sciences 1 (02), 227–237.

Scholkmann, F., Kleiser, S., Metz, A.J., Zimmermann, R., Pavia, J.M., Wolf, U., Wolf, M., 2014. A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. Neuroimage 85, 6–27.

Sharples, S., 2018. Workload II: A future paradigm for analysis and measurement. Congress of the International Ergonomics Association. Springer, pp. 489–498.

Sharples, S., Megaw, T., 2015. Definition and mesurement of human workload. In: Wilson, J.R., Sharples, S. (Eds.), Evaluation of human work. CRC Press.

Shi, Y., Ruiz, N., Taib, R., Choi, E., Chen, F., 2007. Galvanic skin response (GSR) as an index of cognitive load. CHI'07 extended abstracts on Human factors in computing systems. ACM, pp. 2651–2656.

Shimomura, Y., Yoda, T., Sugiura, K., Horiguchi, A., Iwanaga, K., Katsuura, T., 2008. Use of frequency domain analysis of skin conductance for evaluation of mental workload. Journal of physiological anthropology 27 (4), 173–177.

Solovey, E.T., Girouard, A., Chauncey, K., Hirshfield, L.M., Sassaroli, A., Zheng, F., Fantini, S., Jacob, R.J., 2009. Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines. Proceedings of the 22nd annual ACM symposium on User interface software and technology. ACM, pp. 157–166.

Solovey, E., Schermerhorn, P., Scheutz, M., Sassaroli, A., Fantini, S., Jacob, R., 2012. Brainput: enhancing interactive systems with streaming fNIRS brain input. Proceedings of the 2012 CHI Conference on Human Factors in Computing Systems. ACM, pp. 2193–2202.

Tomasi, D., Chang, L., Caparelli, E., Ernst, T., 2007. Different activation patterns for working memory load and visual attention load. Brain research 1132, 158–165.

Vergotte, G., Perrey, S., Muthuraman, M., Janaqi, S., Torre, K., 2018. Concurrent changes of brain functional connectivity and motor variability when adapting to task constraints. Frontiers in physiology 9, 909.

Wickens, C.D., 2008. Multiple resources and mental workload. The Journal of the Human Factors and Ergonomics Society 50 (3), 449–455.

Yuksel, B.F., Oleson, K.B., Harrison, L., Peck, E.M., Afergan, D., Chang, R., Jacob, R.J., 2016. Learn piano with BACh: An adaptive learning interface that adjusts task difficulty based on brain state. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, pp. 5372–5384.

Zhu, Y., Rodriguez-Paras, C., Rhee, J., Mehta, R.K., 2020. Methodological approaches and recommendations for functional near-infrared spectroscopy applications in HF/E research. Human factors 62 (4), 613–642.

Wilson, M. L., Sharon, N., Maior, H. A., Midha, S., Craven, M. P., Sharples, S., 2018. Mental workload as personal data: designing a cognitive activity tracker. In Proceedings of the 3rd Symposium on Computing and Mental Health: Understanding, Engaging, and Delighting Users.