Wilson, Max L. and Alsuraykh, Norah and Maior, Horia A. (2017) Measuring mental workload in IIR user studies with fNIRS. In: NeuroIIR 2017 - CHIIR Workshop on Challenges in Bringing Neuroscience to Research in Human-Information Interaction, 11 Mar 2017, Oslo, Norway.

Measuring Mental Workload in IIR User Studies with fNIRS

Max L. Wilson
School of Computer Science
University of Nottingham
Nottingam, UK
max.wilson@nottingham.ac.uk

Norah Alsuraykh
School of Computer Science
University of Nottingham
Nottingam, UK
psxna10@nottingham.ac.uk

Horia Maior
School of Computer Science
University of Nottingham
Nottingam, UK
horia.maior@nottingham.ac.uk

ABSTRACT

Gathering neuro-physiological data during user studies, and analysing the continuous data they produce, typically involves making a tradeoff between detail and practical utility. This paper describes our long-term work-in-progress towards developing study protocols for using functional Near-InfraRed Spectroscopy (fNIRS) with the aim of finding the ideal balance in this tradeoff. Our results show that fNIRS can be easily used in normal IIR user study conditions, is tolerant of minor movement artefacts (including speaking), and can still determine mental workload differences between different user interfaces designed for the same task.

1 INTRODUCTION

The aim of using Neuro-Physiological methods in Interactive Information Retrieval (IIR) is to take cognitive measurements about a user while they are searching. We aim to objectively gain insights into mental workload, emotional response, and other cognitive activity that go on behind, and motivate, the physical actions that we currently use to evaluate IIR. Rather than use qualitative, subjective or secondary-task methods to infer what the user understands or remembers about their cognitive activity, neuro-physiological methods give us more objective data for analysis.

Ultimately, our aim is to be able to add neuro-physiological sensors into our typical experiment protocols and run a 'normal' IIR user study, but doing so, however, is a hard challenge that involves many tradeoffs. In order to use many Neuro-Physiological sensors, we have to make tradeoffs in the ecological validity of the methods: in order to increase the value of the collected data, we have to reduce the normality of the user study protocol. Below are some of the challenges our team have faced when selecting appropriate neuro-physiological sensors for IIR studies.

Challenge 1: Physical Constraints - It may be necessary to constrain the environment in order to use a sensor: in detecting Information Needs in the brain, Moshfeghi et al had to place participants in an MRI scanner and interact indirectly via mirrors and an intercom [7]. If we try to avoid restricting physical activity, we typically have to use less invasive technology like eye trackers [3], or wrist-worn GSR sensors [9] that provide more indirect data. Natural user study behaviour, however, may involve looking away the screen, and thus from e.g. eye trackers.

Challenge 2: Designing Tasks - We may also need to restrict the tasks that people do in order to be able to take comparable measures: BCI studies typically involve abstract psychology tasks [6, 11]. However, a natural IIR study protocol might involve asking participants to perform exploratory search tasks that involve very complex behaviour. It becomes hard to determine when participants are e.g. making relevance judgements or simply reading.

Challenge 3: Confounding Variables - The main reason for either applying physical or experimental restrictions is to reduce possible confounding variables in the data. However, many actions might be counted as a step by a pedometer, many different mental activities have been associated with changes that can be seen in neuro-physiological sensors: affective responses, stress, and so on.

Challenge 4: Protocol Compatability - A typical user study protocol may involve taking many different types of measurements. Many IIR researchers utilise Think Aloud Protocols, but as well as being affected by physical movement, sensors like EEG data are heavily affected by speaking during tasks (see below). Many studies may also involve specific time constraints, however these may invoke stress-related emotional responses in some participants that create confounding variables noted above. Further, in order to correlate neuro-physiological signals with user behaviour, we often have to add elements to our protocols, like video footage of participants' behaviour, which can also reduce ecological validity.

Challenge 5: Data Analysis - Another benefit of choosing more abstracted psychology-based tasks, is in the data analysis and interpretation of results; as task complexity increases, it becomes harder to understand what causes certain reactions in participants behaviour, performance and physiology. It is therefore important to consider task complexity as a factor when designing and analyzing IIR studies, especially during more *natural*, exploratory tasks. We must further then decide which signals (EEG sensors have many sensors, covering various parts of the brain) to analyse and how to process them. Neuro-physiological data may peak and trough several times a second, but interactions that we wish to make judgements on may take several seconds: do we take an average of set periods of time? or measure number of peaks? or highest peak?

2 FNIRS AS A GOOD TRADEOFF

Based on the challenges above, choosing a neuro-physiological sensor, becomes a reasoned decision based upon the research questions driving the work and the ideal methods of answering those research questions. Moshfeghi et al designed a very careful methodology in order to detect Information Needs in the brain [7]. If we wished to evaluate a new user interface to help people resolve search tasks more quickly, or more accurately, we could not use an MRI to evaluate the cognitive demands in such tasks.

The sensors used to assess users' cognitive processes during natural IIR user studies should ideally provide useful information about them, without dramatically changing the protocols typically used, or further restricting the users during the studies. Therefore, an

NeuroIIR'17, Olso, Norway

DOI: -

MRI is too physically restrictive, but EEG data is too heavily affected by natural participant behaviour. Eye trackers are becoming much more flexible, and are complementary to the BCI studies. However, eye tracking data can be interrupted by the 'normal' behaviour we would ideally retain in our studies. Wrist-worn skin sensors are also too indirectly associated with the cognitive processes we wish to observe, because the sensors are responsive to arousal generated by both physical and mental activities.

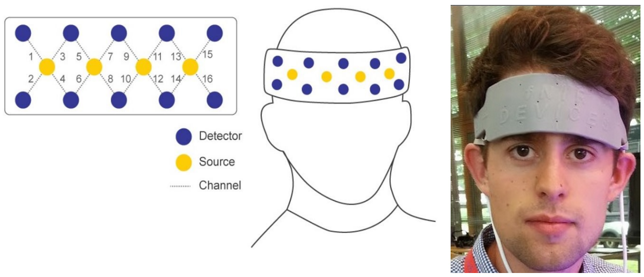

Figure 1: A Biopac 16 channel fNIRS

Our own choice of neuro-physiological sensor has been functional Near-InfraRed Spectroscopy (fNIRS) shown in Figure 1, which takes objective measurements of blood oxygen using near-infrared light, rathern than e.g. measures of electrical changes in the brain. By placing fNIRS on the forehead, it takes measurements of brain activity in the pre-frontal cortex, an area associated with working memory. As the participant experiences higher levels of mental workload, the pre-frontal cortex demands more oxygen from the heart and lungs. Consequently, fNIRS data, like MRI, is delayed by the BOLD response. Below we review how we work with fNIRS against the five challenges listed above.

Challenge 1: Physical Constraints - There are very few physical constraints in using fNIRS. Work by Solovey et al [10], and our own research [6], has shown that fNIRS is tolerant of most minor movement artefacts, which means that participants can behave quite naturally in e.g. a seated position in front of a computer and work within the range of the 2 metre cables. However, wireless fNIRS devices are now available.

Challenge 2: Designing Tasks - While we have used fNIRS to study more constrained psychology tasks ourselves [6, 8, 10], the technique was successfully deployed during more naturalistic studies, including tasks such as air traffic control [5], and usability testing the form filling process of an insurance claim [4].

Challenge 3: Confounding Variables - We are aware of the affect of related cognitive activities affecting the data. Some participants in our air traffic control task reported feeling anxious about the task conditions. Alsuraykh has begun her doctoral studies focused on examining stress as an artefact in Mental Workload measurements using fNIRS.

Challenge 4: Protocol Compatability - fNIRS involves minimal changes to our normal user study protocols. For example, there is some time allocated for setup and configuration at the beginning of each data collection session. We typically use n-back tasks [2, 11] to identify baseline measurements of high (3-back) and low (1-back) mental workload during this configuration period, as well as taking a measurement at rest. We also have to introduce 2 minute rest periods between tasks, and add time-markers using the fNIRS software. However, we have shown that fNIRS can be used freely along side other protocol elements, including Think Aloud Protocols [8].

Challenge 5: Data Analysis - Measures of brain activity are recorded using an fNIRS300 device and the associated Cognitive Optical Brain Imaging (COBI) Studio platform provided by Biopac Systems Inc. The sensor headband is a sixteen-channel transducer for continuous Near Infrared Spectroscopy (NIRS), producing two types of measurements for each channel: Oxygenated (HbO) and deoxygenated (Hb) hemoglobin. Relative changes in HbO and Hb strongly correlate with changes in Mental Workload (when targeting the PFC). A typical fNIRS experiment involves various levels of data pre-processing and processing. Filtering techniques (e.g. low pass filters) can be used in order to remove high-frequency noise, physiological artefacts such as heartbeats and motion derived artefacts. The Correlation Based Signal Improvement (CBSI) method can be applied [1], for example, which combines HbO and Hb measurements in order to improve detection of Mental Workload.

3 CONCLUSIONS

Over the last three years, we have worked hard on evaluating the utility of fNIRS as a neuro-physiological sensor that can be used effectively to more objetively measure Mental Workload changes within HCI-style IIR user studies. In particular, we have shown that it is tolerant of minor movement artefacts [6], can be used along side other user study protocols like Think Aloud [8], and can be used within a usability study to examine the difference between three user interface conditions for the same task [4]. Our work to date, therefore, demonstrates that we are ready to go beyond constrained neuro-physiological studies of e.g. relevance judgements, and begin to evaluate different search user interfaces designs in terms of the cognitive impact they have on searchers.

REFERENCES

[1] Xu Cui, Signe Bray, and Allan L Reiss. 2010. Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage* 49, 4 (2010), 3039–3046.

[2] Jacek Gwizdka. 2010. Distribution of cognitive load in web search. *JASIST* 61, 11 (2010), 2167–2187.

[3] Jacek Gwizdka and Yinglong Zhang. 2015. Differences in eye-tracking measures between visits and revisits to relevant and irrelevant web pages. In *Proc. SIGIR2015*. ACM, 811–814.

[4] Kristiyan Lukanov, Horia A Maior, and Max L Wilson. 2016. Using fNIRS in usability testing: understanding the effect of web form layout on mental workload. In *Proc. CHI2016*. ACM, 4011–4016.

[5] Horia Maior, Max L. Wilson, and Sarah Sharples. 2017. Workload Alerts - Using Physiological Measures of Mental Workload to Provide Feedback during Tasks. *TOCHI (in submission)* (2017).

[6] Horia A Maior, Matthew Pike, Sarah Sharples, and Max L Wilson. 2015. Examining the reliability of using fNIRS in realistic HCI settings for spatial and verbal tasks. In *Proc. CHI2015*, Vol. 15. 3807–3816.

[7] Yashar Moshfeghi, Peter Triantafillou, and Frank E Pollick. 2016. Understanding information need: An fmri study. In *Proc. SIGIR2016*. ACM, 335–344.

[8] Matthew F Pike, Horia A Maior, Martin Porcheron, Sarah C Sharples, and Max L Wilson. Measuring the effect of think aloud protocols on workload using fNIRS.

[9] Yu Shi, Natalie Ruiz, Ronnie Taib, Eric Choi, and Fang Chen. 2007. Galvanic skin response (GSR) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems*. ACM, 2651–2656.

[10] Erin Treacy Solovey, Audrey Girouard, Krysta Chauncey, Leanne M Hirshfield, Angelo Sassaroli, Feng Zheng, Sergio Fantini, and Robert JK Jacob. 2009. Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines. In *Proc. UIST*. ACM, 157–166.

[11] Shouyi Wang, Jacek Gwizdka, and W Art Chaovalitwongse. 2016. Using Wireless EEG Signals to Assess Memory Workload in the n-Back Task. *IEEE Transactions on Human-Machine Systems* 46, 3 (2016), 424–435.