Directly Evaluating the Cognitive Impact of Search User Interfaces: a Two-Pronged Approach with fNIRS

Horia A. Maior[1,2], Matthew Pike[1], Max L. Wilson[1], Sarah Sharples[3]
[1]Mixed Reality Lab, [2]Horizon DTC, [3]Human Factors - School of Engineering
University of Nottingham, UK
{psxhama,psxmp8,max.wilson,sarah.sharples}@nottingham.ac.uk

ABSTRACT

Recent research has pointed towards further understanding the cognitive processes involved in interactive information retrieval, with most papers using secondary measures of cognition to do so. Our own research is focused on using direct measures of cognitive workload, using brain sensing techniques with fNIRS. Amongst various brain sensing technologies, fNIRS is most conducive to ecologically valid user studies, as it is less affected by body movement and can be worn while using a computer at a desk. This paper describes our two pronged approach focusing on a) moving fNIRS research beyond simple psychological tests towards actual interactive IR tasks and b) evaluating real search user interfaces.

Categories and Subject Descriptors

H5.2 [**Information interfaces and presentation**]: Evaluation/methodology, Theory and methods

Keywords

Functional near-infrared spectroscopy(fNIRS), Brain-computer interface(BCI), Human cognition, Information processing system, Multiple resource model, Limited resource model

1. INTRODUCTION

The cognitive aspects of Information Retrieval (IR) have repeatedly received focus over time, from Ingwersen's Cognitive Model [11], to recent analyses of cognitive workload during search tasks [2, 10]. The recurring interest is in what users think about at different task stages, and how much mental workload is involved. The benefits of knowing more about the searcher's cognitive state would come from providing better support for their needs, with Wilson et al suggesting that better designed Search User Interfaces (SUIs) could reduce unnecessary workload on the user [23].

Although some prior work (e.g. [2]) have used indirect techniques to analyse workload during search tasks, the decreasing cost of brain sensing hardware has meant that more recent research is using more objective techniques. Pike et al [17] and Gwizdka et al [10] used EEG technology, while Moshfeghi et al used fMRI to measure workload when making relevance judgements [15]. Each of these technologies have known limitations for studying actual interactive IR behaviour, with EEG being highly affected by even tiny body movement, and fMRI requiring users to lay in tunnel void of any metal objects. Recent Human-Computer Interaction research has listed the benefits of fNIRS brain sensing techniques, which are less affected by body movement, and can be more easily used in ecologically valid study conditions.

Functional Near Infrared Spectroscopy (fNIRS) is an emerging neuroimaging technique that is non-invasive, portable, inexpensive and suitable for periods of extended monitoring. fNIRS measures the hemodynamic response - the delivery of blood to active neuronal tissues. fNIRS is designed to be placed directly upon a participants scalp, typically targeting the prefrontal cortex. This paper describes our two-pronged approach to using fNIRS to study the cognitive workload created by SUIs, focused on a) task analysis and b) SUI analysis.

2. RELATED WORK

Understanding the cognitive aspects of interactive searching (as well as interaction in general) has been a long-standing goal for researchers in the field of Interactive IR. In the 1970s Bates suggested that searchers employ both search tactics and idea tactics [7]. In an attempt to explain an individual's path during IR, Bates' "Berrypicking" model [8] argued that search will vary as the user recognises information and has new ideas and questions.

In the main cognitive evolution of information seeking research, Ingwersen proposed a cognitive model of IR [11], where the searcher's understanding of the document collection, system, and task that would determine which path a search would take. The model again put the user's cognition as the central point of interest. More recently, Joho [12] argued that the cognitive effects typically observed in Psychology could provide a potential building block of theoretical development for evaluating interactive IR. Back et al [2], for example, examined the cognitive demands on users during the relevance judgement phase, suggesting that the amount of workload involved was the reason behind searchers rarely providing relevance judgements in previous work. Using a secondary measure, the Stroop task, Gwizdka [10] mapped varying levels of workload at multiple stages of search.

More recently, researchers have focused on objectively measuring interactive IR phases, in line with Back et al's work, Moshfeghi et al measured workload during relevance assessments by asking people to make judgements while lying in an fMRI machine. As making relevance judgements can be performed without directly interacting with a computer, this made use of an fMRI machine more realistic. Using more commercialised tools, Anderson [1] used an EEG sensor to compare visualization techniques in terms of the burden they place on a viewer's cognitive resources. Similarly, Pike et al [17] developed a prototype tool named CUES that was capable of collecting a variety of data including EEG whilst interacting with a website. Pike et al used this to monitor aspects such as frustration and concentration, but their work demonstrated the variability of EEG data across the several minutes involved in an interactive IR task.

Using fNIRS, as introduced above, Peck [16] performed a similar study of different visualisation techniques, while a system called Brainput [18] was able to identify and correlate brain activity patterns among users during multitasking studies, and intervene when it sensed workload exceeding a certain level. Our work intends to build upon these HCI studies, to study interactive IR tasks and SUIs in more ecologically valid user study situations.

3. RESEARCH PATHS

Pike et al [17] highlighted the challenges of using brain sensing technologies to evaluate IIR tasks: that tasks have different stages, that behaviour quickly diverges after the first interaction (and thus is hard to compare), and that brain measurements vary dramatically over time. In order to address these challenges, we have initiated two clear research paths, both utilising fNIRS technology: 1) evaluating the cognitive aspects of Interactive IR tasks and 2) methods to evaluate the design of SUIs. The aim of the first path, is to move beyond using fNIRS to measure workload in simplistic psychology memory tasks (like Peck et al [16]), towards being able to break down real search tasks into primary components. This implies three considerations:

- Collected data would be meaningless if is not related to existing knowledge. Therefore, to interpret sensed fNIRS data we use proposed theories and models.
- It is known that fNIRS can sense cognition information [19, 16] related to so called working memory (if placed on the forehead). Assuming this is correct, we are using models of working memory.
- The proposed models will help us interpret the sensed data with fNIRS and have a better understanding of the cognitive impact of various complex tasks (such as a IR).

Such a technique would allow researchers to analyse data by stage, and find effective points of comparison during several minutes of continuous measurements. The second path is focused on identifying which aspects of working memory are affected by different features of SUIs, such that researchers can objectively evaluate the effect of different SUI design decisions. A combination of both paths works towards being able to proactively evaluate how SUIs support searchers.

4. PATH 1: WORKLOAD MODELS

To understand the cognitive aspects of IIR, it is essential to learn about user's capabilities and limitations in terms of their cognition: how people perceive, think, remember, and process information. This path of research focuses on existing models from Cognitive Psychology and Human Factors, models that conceptualize and highlight aspects that typically describe or influence elements of human cognition. One important part of cognition during interactive searching involves human memory systems. There are two different types of memory [21]: working memory (sometimes called short-term memory) and long-term memory. Wickens describes working memory as the temporary holding of information that is "active", while long-term memory involving the unlimited, passive storage of information that is not currently in working memory.

Working memory. Working memory, proposed by Baddeley and Hitch (1974) [6], refers to a specific system in the brain which "provides temporary storage and manipulation of information..." [3]. Working memory [6, 4, 5] processes information in two forms: verbal and spatial, and has four main components (Figure 1):

- A **central executive** managing attention, acting as supervisory system and controlling the information from and to its "slave systems".
- A **visuo-spatial sketch pad** holding information in an analogue **spatial** form (e.g. Colours, shapes, maps, etc.), specialised on learning by means of visuospatial imagery.
- A **phonological loop** holding **verbal** information in an acoustical form (e.g. Numbers, words, etc.); specialised on learning and remembering information using repetition.
- A **episodic buffer** dedicated to linking verbal and spatial information in chronological order. It is also assumed to have links to long-term memory.

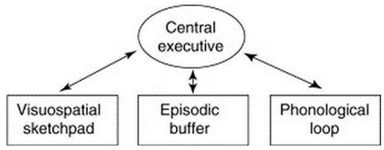

Figure 1: Baddeley's Working Memory Model

Information processing system. As humans, we are exposed to large amounts of information via our sensory systems. One of our strengths is in selecting information from our environment, perceiving it, processing it, and creating a response. Therefore we can use this understanding of brain activity to identify which elements of an interactive IR environment need to be considered when measuring brain activity, and how we can reduce rather than increase a user's mental workload via interface and system design.

Wicken's Information Processing Model [21] aims to illustrate how elements of the human information processing system such as attention, perception, memory, decision making and response selection interconnect. We are interested in observing how and when these elements interconnect during IR. He describes three different 'stages' (see STAGES dimension in Figure 2) at which information is transformed: a perception stage, a processing or cognition stage, and a response stage, the first two being processes involved in cognition. The first stage involves perceiving information that is gathered by our senses and provide meaning and interpretation of what is being sensed. The second stage represents the step where we manipulate and "think about" the perceived information. This part of the information processing system takes place in working memory and consists of a wide variety of the mental activities. In relation to IR, it is interesting to observe how elements of cognition, such as rehearsal of information, planning the search strategy and deciding on the search keywords interconnect.

Multiple Resource Model. One model of mental workload that has been widely accepted in Human Factors is Wickens Multiple Resource Model [20] (Figure 2). The elements of this model overlap with the needs and considerations of evaluating complex tasks (such as IR). He describes the aspects of human cognition and the multiple resource theory in four dimensions:

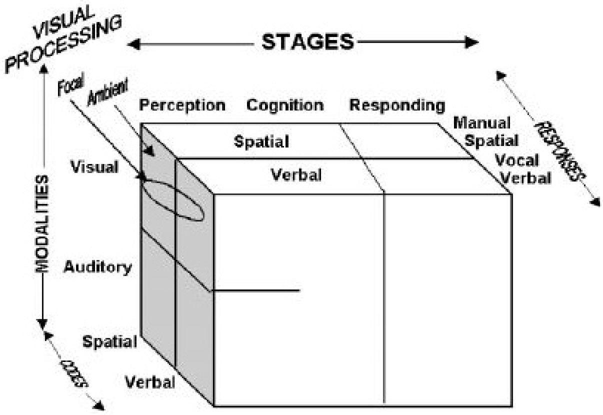

Figure 2: The 4-D multiple resource model [20]

- The STAGES dimension refers to the three main stages of information processing system (Wickens, 2004 [21]).
- The MODALITIES dimension indicating that auditory and visual perception have different sources.
- The CODES dimension refers to the types of memory encodings which can be spatial or verbal.
- The VISUAL PROCESSING dimension refers to a nested dimension within visual resources distinguishing between focal vision (reading text) and ambient vision (orientation and movement).

Our aim is to understand how these elements link together and compose more complex components/tasks. Additionally we want to consider how complex tasks (such as a search task) can be divided into primary components according to the models described. This will help identify possible problems in SUI design as well as indicating a possible solution to the problem (suggested implications by Wickens [21]):

- Minimize working memory load of the SUI system and consider working memory limits in instructions;
- Provide more visual echoes (cues) of different types during IR (verbal vs spatial);
- Exploit chunking (Miller, 1956 [14]) in various ways: physical size, meaningful size, superiority of letters over numbers, etc;
- Minimize confusability;
- Avoid unnecessary zeros in codes to be remembered;
- Encourage regular use of information to increase frequency and redundancy;
- Encourage verbalization or reproduction of information that needs to be reproduced in the future;
- Carefully design information to be remembered;

Resource vs Demands. One other model that is of interest is the limited resource model [22] describing the relationship between the demands of a task, the resources allocated to the task and the impact on performance.

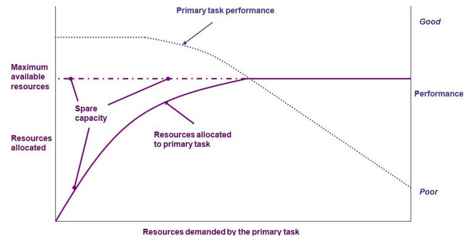

Figure 3: Resources available vs task demands → impact on performance [22]

The graph from Figure 3 is used to represent the limited resource model. The X-axes represent the resources demanded by the primary task and as we move to the right of the axes, the resources demanded by the primary task increase. The axes on the left indicate the resources being used, but also the maximum available resources point (if we think of working memory that is limited in capacity). The right axes indicate the performance of the primary task (the dotted line on the graph). The key element of this model is the concept of a limited set of resources which, if exceeded, has a negative impact on performance. However, it does not distinguish between resource modality, therefore we propose to use both the limited and multiple resources models to inform our work.

5. PATH 2: SUI EVALUATION

Relating quantitative data from brain sensing devices into feedback about SUI designs is one of our ultimate goals in conducting this research. SUIs are inherently information rich and thus affect both visual (results page layout) and verbal (text based results) memory. Detecting a change in either verbal or spatial working memory would help determine if a workload difference was caused by SUI design (spatial) or the amount of information the design provides (verbal). Our first in-progress study has stimulated each memory type in different tasks - Verbal memory was tested by performing an n-back [13] number memory task, whereas spatial memory was tested using an n-back visual block matrix task. Other studies have also looked at each type of memory and confirmed fNIRS ability to detect changes in heamodynamic responses accordingly [9].

In addition to developing an understanding of the extent to which we can monitor different memory, our initial study also sought to measure the effect of artefacts on the fNIRS data. Controlling the environment and human derived sources of noise is a potentially difficult factor to control without effecting the ecological validity of a study.

Solovey et al [19] showed that fNIRS is relatively resilient to motion derived artefacts when compared to EEG [17] for example, but still required some consideration by researchers conducting studies. In our own experience, we found that asking participants to remain still as much as possible was fairly successful. We are additionally looking at possible methods for correcting motion derived artefacts using an external gyroscope connected to the participant.

Designing tasks for experiments that measure cognitive effect via a brain sensor require careful consideration in order to ensure that results can be attributed to a cause. Thankfully this problem space has been well explored in the field of Psychology and we are able to adapt the approaches described in the literature to suit our task type requirements. A primary example of this adaptation is demonstrated by Peck et al [16], where 2 data visualisations techniques were compared using a methodology based loosely on the n-back task - a widely used psychology task that is designed to increase load on working memory.

Additionally, we are interested in exploring standard search studies (without following a psychological study layout) and seeing whether interesting states can be detected. Solovey et al [18] performed a similar function by utilising a machine learning algorithm that had classified "states of interest" prior to performing a task.

Using a similar approach, we could evaluate a SUI to determine whether a particular change in layout has a positive or negative impact on visual memory. Alternatively, to test the relevance of a results page (which would be dependant on the textual results), we could analyse the effects on verbal memory between 2 varied results pages, we could then reflect these changes to the Wickens Multiple Resource Model [20]. We are also working towards enabling the interpretation of data within the context of complex multimodal tasks to further extending our knowledge of the processes involved during IR and how they interact and effect one another.

6. SUMMARY

This paper has aimed to summarise our two-pronged approach towards actually evaluating the design of search user interfaces, in realistic ecologically valid study conditions, using fNIRS technology. The approach first involves braking down interactive IR tasks into how they effect the different elements of working memory, and second understanding how SUIs are processed by different parts of working memory. Our two paths of research will build towards a stage where we can combine them and objectively evaluate cognitive workload involved in interactive IR. We believe that this research will provide a novel new direction that SUI's and indeed HCI in a broader sense can benefit from. The association of physical recordings in ecological valid settings, to an existing theoretical model, provides a new measure from which future SUI development and evaluation could benefit.

7. REFERENCES

[1] E. W. Anderson, K. Potter, L. Matzen, J. Shepherd, G. Preston, and C. Silva. A user study of visualization effectiveness using EEG and cognitive load. *Computer Graphics Forum*, 30(3):791–800, 2011.

[2] J. Back and C. Oppenheim. A model of cognitive load for IR: implications for user relevance feedback interaction. *Information Research*, 6(2):6–2, 2001.

[3] A. Baddeley. Working memory. *Science*, 255(5044):556–559, 1992.

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

[5] A. D. Baddeley. Is working memory still working? *European psychologist*, 7(2):85–97, 2002.

[6] A. D. Baddeley and G. Hitch. Working memory. *The psychology of learning and motivation*, 8:47–89, 1974.

[7] M. J. Bates. Idea tactics. *JASIST*, 30(5):280–289, 1979.

[8] M. J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5):407–424, 1989.

[9] X. Cui, S. Bray, D. M. Bryant, G. H. Glover, and A. L. Reiss. A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *Neuroimage*, 54(4):2808–2821, 2011.

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

[11] P. Ingwersen. Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of documentation*, 52(1):3–50, 1996.

[12] H. Joho. Cognitive effects in information seeking and retrieval. In *Proc. CIRSE2009*, 2009.

[13] W. K. Kirchner. Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology*, 55(4):352, 1958.

[14] G. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The psychological review*, 63:81–97, 1956.

[15] Y. Moshfeghi, L. R. Pinto, F. E. Pollick, and J. M. Jose. Understanding Relevance: An fMRI Study. In *Proc. ECIR2013*, pages 14–25. Springer, 2013.

[16] E. M. Peck, B. F. Yuksel, A. Ottley, R. J. Jacob, and R. Chang. Using fNIRS Brain Sensing to Evaluate Information Visualization Interfaces. In *Proc. CHI2013*. ACM, 2013.

[17] M. Pike, M. L. Wilson, A. Divoli, and A. Medelyan. CUES: Cognitive Usability Evaluation System. In *EuroHCIR2012*, pages 51–54, 2012.

[18] E. Solovey, P. Schermerhorn, M. Scheutz, A. Sassaroli, S. Fantini, and R. Jacob. Brainput: enhancing interactive systems with streaming fNIRs brain input. In *Proc. CHI2012*, pages 2193–2202. ACM, 2012.

[19] E. T. Solovey, A. Girouard, K. Chauncey, L. M. Hirshfield, A. Sassaroli, F. Zheng, S. Fantini, and R. J. Jacob. Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines. In *Proc. UIST2009*, pages 157–166. ACM, 2009.

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

[21] C. D. Wickens, S. E. Gordon, and Y. Liu. *An introduction to human factors engineering*. Pearson Prentice Hall Upper Saddle River, 2004.

[22] J. R. Wilson and E. N. Corlett. *Evaluation of human work*. CRC Press, 2005.

[23] M. L. Wilson. Evaluating the cognitive impact of search user interface design decisions. *EuroHCIR 2011*, pages 27–30, 2011.