Using Brain Measurement to Evaluate Reality Based Interactions

*Leanne M. Hirshfield** *Erin Treacy Solovey** *Audrey Girouard** *James Kebinger** *Michael Horn** *Orit Shaer+* *Jamie Zigelbaum^* *Robert J. K. Jacob**

*Tufts University
Medford, MA 02155, USA
{leanne.hirshfield, erin.solovey, audrey.girouard, james.kebinger, robert.jacob}@tufts.edu

+Wellesley College
106 Central Street
Wellesley, MA 02481, USA
oshaer@wellesley.edu

^MIT Media Lab
20 Ames Street
Cambridge, MA 02139
zig@media.mit.edu

ABSTRACT

We proposed the notion of Reality-Based Interaction (RBI) as a unifying concept to tie together a large subset of emerging interaction styles [7]. Interfaces should adhere to reality-based principles unless designers explicitly choose to make a tradeoff where they sacrifice reality in order to gain other positive qualities. When a reality-based interface has been designed, and the reality tradeoffs have been carefully considered, it is very difficult to determine whether or not the reality-based principles or reality tradeoffs resulted in a '*good*' interface. To address this problem, we use functional near infrared spectroscopy (fNIRs), a new, non-invasive brain imaging technique, to complement standard usability experiments. We demonstrate an experimental protocol and data analysis algorithms which can help evaluators of RBI systems to acquire objective, real time metrics of users' workload, and other states while working with a given interface.

Author Keywords

Usability, reality-based interaction, RBI, brain, cognition

ACM Classification Keywords

H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

INTRODUCTION

A well designed computer interface should be nearly transparent, allowing the user to focus on the task at hand [12]. This is a common mantra for experts in Human Computer Interaction (HCI) who conduct a wealth of research on designing and evaluating user interfaces. It is well known that effectively designing and evaluating interfaces enhances user performance, increases user satisfaction, and increases safety[14]. Therefore, determining the most effective techniques for evaluating human computer interfaces remains an important and popular area of research. Although we can measure accuracy and time to complete a task, measuring factors such as user workload, frustration, enjoyment, and distraction are often done by qualitatively observing subjects or by administering subjective surveys to subjects. These surveys are often administered after a task has been completed, lacking valuable insight into the user's changing experiences throughout the task.

The need for accurate and objective usability metrics has been of interest to researchers for decades, and it has become increasingly important due to the recent explosion of emerging human-computer interaction techniques. These new techniques redefine our understanding of both computers and interaction. We proposed the notion of *Reality-Based Interaction* (RBI) as a unifying concept to tie together a large subset of these emerging interaction styles[7] such as virtual, mixed and augmented reality, tangible interaction, ubiquitous and pervasive computing, context-aware computing, handheld, or mobile interaction, perceptual and affective computing as well as lightweight, tacit or passive interaction.

Direct Manipulation moved interfaces away from the command line and moved them closer to *real world* interaction by allowing users to directly manipulate objects rather than instructing the computer to do so by typing commands. New interaction styles push interfaces further in this direction[7]. They increase the realism of interface objects and allow users to interact even more directly with them—using actions that correspond to daily

CHI 2009, April 4–9, 2009, Boston, Massachusetts, USA.
ACM 978-1-60558-247-4/09/04…$5.00

practices within the non-digital world. In our recent CHI paper[7], we described a framework to understand and compare these emerging HCI technologies. We discussed implications for design, and a set of tradeoffs that interface designers should consider when adding or removing reality-based principles from their interface. We believe that elements of reality should only be removed from an interface when the designers of that interface make an explicit choice to make a tradeoff between reality and one of the following RBI tradeoffs[7].

RBI Tradeoffs

- *Expressive Power:* i.e., users can do a variety of tasks within the application domain
- *Efficiency:* users can do a task with the system rapidly
- *Plasticity:* users can do many tasks from different application domains
- *Ergonomics:* users can do a task without physical injury or fatigue
- *Accessibility:* users with a variety of abilities can perform a task
- *Practicality:* the system is practical to develop and produce[7]

It is up to designers of emerging interfaces to make knowledgeable decisions about the reality tradeoffs made within the design of their interfaces, and the effect of those tradeoffs. A designer whose primary goal is *accessibility for all* may be willing to make a less reality based, intuitive interface in order to gain a high degree of accessibility. We argue that designers should be mindful of the reality tradeoffs that they make, why they are making those tradeoffs, and what the effects of that tradeoff will be on the intuitiveness, enjoyment, efficiency, etc. of their interface.

This new generation of interaction styles poses new problems for the already challenging area of usability testing. In particular, evaluation techniques for direct manipulation interfaces may be insufficient for the newly emerging generation of interaction styles. Many new interfaces claim to be “intuitive,” a claim that is often difficult to quantify, and users’ goals for these emerging interfaces often differ from the more common productivity goals that are associated with many command line and GUI applications. This makes objective measurements such as workload, engagement, frustration, boredom, and fatigue of paramount importance for this new generation.

EVALUATING RBI WITH FUNCTIONAL NEAR INFRARED SPECTROSCOPY

To tackle these issues, current research focuses on developing objective techniques to measure user states such as workload, emotion, and fatigue [8, 9] in real time. Although this ongoing research has advanced user experience measurements in the HCI field, finding accurate, non-invasive tools to measure computer users’ states in real working conditions remains a challenge. Our research addresses these evaluation challenges. We use a new, non-invasive brain imaging technique called functional near infrared spectroscopy (fNIRs) to make real time, objective measurements of users’ mental state while working with interfaces. fNIRs was introduced in the early-mid 1990’s [1, 2] to complement, and in some cases overcome practical and functional limitations of EEG and other brain monitoring techniques. This tool has been shown to quantitatively measure attention, working memory, target categorization, and problem solving [6].

The tool, still a research modality, uses light sources in the near infrared wavelength range (650-850 nm) and optical detectors to probe brain activity. Light sources and detection points are defined by means of optical fibers which are held on the scalp with an optical probe (Figure 1). Deoxygenated and oxygenated hemoglobin are the main absorbers of near infrared light in tissues during hemodynamic and metabolic changes associated with neural activity in the brain [1]. We can detect these changes by measuring the diffusively reflected light that has probed the brain cortex [1, 5, 13]. Researchers have shown that by placing the probes on a subject’s forehead, fNIRs provides an accurate measure of activity within the frontal lobe of the brain [5]. These results are promising when combined with the fact that fNIRs is safe, portable, less invasive than other imaging techniques, and has been implemented wirelessly, allowing for use in real world environments [5, 10].

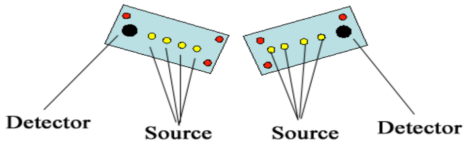

Figure 1: two probes and their sources and detectors.

Much of our research focuses on the measurement of mental workload with the fNIRs device[3, 11], but research suggests that we may be able to measure a variety of other useful user states with this device. With respect to workload, interface designers and evaluators want to minimize the workload attributed to the interface while maximizing the workload attributed to the task. This principle holds for Direct Manipulation and for Reality-Based Interactions. We follow the concept of Shneiderman's theory of syntactic/semantic components of a user interface [12]. The semantic component involves the workload and effort needed to complete the task. The syntactic component includes interpreting the interface’s feedback and formulating and inputting commands to the interface. An important goal in interface design is to reduce the amount of mental effort devoted to the syntactic aspects so that more mental workload can be devoted to the underlying task or semantic aspects.

We believe that brain measurement can be used as an additional metric in usability studies to acquire real time, objective measurements that shed light on the syntactic workload associated with UIs[4]. We focus on the interacting cognitive subsystems, or cognitive resources, that work together to process information (i.e., working memory, executive processing or visual search) while a user works with a UI and task. The brain is a complex structure, making it nearly impossible to completely separate resources devoted to processing the semantic (task) and syntactic (UI) components of workload. However, we argue that much information can be acquired about syntactic workload by breaking mental workload down into its multiple cognitive resources[4].

While experts in HCI have little control over semantic workload, we can attempt to modify UIs to decrease their syntactic workload, and we posit that lowering syntactic workload in emerging interfaces will have a relationship with the amount of *reality* in a given RBI interface. There may be situations when we increase reality in order to make an interface more intuitive. However, referring back to our RBI tradeoffs, there may also be cases when we make a tradeoff and decrease the amount of reality in an interface in order to add, for example, efficiency to the interface. For instance, if we have an interface that is based completely on reality, it may take a user time *t*, to complete task *x* using mental workload *w*. If we can make that interface more efficient, (reality/efficiency tradeoff), users may be able to expend a comparable amount of workload *w*, over less time *t*, to achieve the same task *x*.

We present a novel experimental protocol and data analysis algorithms that can help usability experts to gain information about the workload experienced by computer users in the various cognitive resources in their brain while working with a computer system[4]. We show how this can be related to the objective measurement of various user states such as frustration, boredom, enjoyment, workload, and fatigue during usability studies of RBIs. Our methodology can be used by designers of new interfaces, while they iterate on their design ideas and evaluate the workload, enjoyment, or frustration of making various tradeoffs between reality and power, ergonomics, efficiency, plasticity, accessibility, or practicality. The methodology can also be used post-development to compare the intuitiveness, or enjoyment associated with different interaction styles.

USABILITY EXPERIMENT PROTOCOL

We designed an experimental protocol to shed light on the workload experienced by a user's various cognitive resources while working with a computer system[4]. Given a UI to evaluate and an underlying task, we conduct a task analysis on the UI and task. For each subtask, we determine the cognitive subsystems that one would use while conducting the subtasks (i.e., spatial WM, visual search, etc.). We gather benchmark exercises from cognitive psychology designed to elicit high and low levels of workload on the target cognitive resource(s) associated with our UI.

Next, we run an experiment where users complete the well established cognitive psychology exercises, giving us a measure of their brain activity while they experience high and low workload levels in their various cognitive subsystems. Users also work with the UI that we are attempting to evaluate. Lastly, we use fNIRs data analysis tools to find similarities and differences in the user's brain activity while working with the UI to be evaluated and while completing the cognitive psychology exercises.

While the protocol, in its most general form, will not yield exact measures of syntactic workload for any given UI, usability experts can incorporate the protocol into their current usability studies and use the knowledge gained as an added usability metric. For example, designers of a virtual reality system may use this protocol in a usability study and find that their users were visually overloaded while searching for items in the virtual space. They could determine this by finding that the users' brain activity while working in the VR space was similar to the users' brain activity while conducting a cognitive psychology exercise designed to cause high visual search workload. In this case, the designers could re-design the interface to place less demand on users' visual search—perhaps by highlighting possible searched objects within the space. *This would be an example of an RBI reality tradeoff where reality is lessened in order to gain efficiency.*

In the future, one could imagine a *training period*, where users work with a set of benchmark cognitive psychology exercises designed to target particular cognitive resources (i.e., verbal WM, spatial WM, visual scanning, auditory processing). After determining the patterns of brain activity induced by the various benchmark exercises, users could work with a computer system and usability experts could search for similarities between the users' brain activity while working with the computer system, and the brain activity already established during the training period.

FEASIBILITY EXPERIMENT

We conducted an experiment that used this usability experiment protocol to measure the spatial WM load of two user interfaces that involve traversing through hyperspace while conducting an information retrieval (IR) task[4]. We analyzed our experiment data using Analysis of Variance (ANOVA) and clustering routines. We also used a k-nearest neighbor classifier with Dynamic Time Warping to classify users' spatial workload on a single trial basis. Our experiment results indicated that we could use our novel protocol and our fNIRs data analysis algorithms to measure the spatial working memory load that users experienced while working with our two user interfaces[4].

CONCLUSION

We use fNIRs to acquire real time, objective measurements of users' states while working with various interfaces. We discussed the reality-based interaction framework, and we placed particular emphasis on the tradeoffs that designers may make when choosing to lower the reality of a system in order to gain efficiency, expressive power, ergonomics, practicality, accessibility, and plasticity. We demonstrated a novel usability experimental protocol that demonstrates how brain measurement could be used to learn about the user's mental state while working with a reality-based interface. RBI designers can use these resulting brain measurements to guide them while making their reality-based tradeoffs.

ACKNOWLEDGMENTS

We thank Krysta Chauncey, Martin Paczynski, Angelo Sassaroli, and Sergio Fantini. In addition, we thank the National Science Foundation for support of this research (NSF Grant No. IIS-#0713506). Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

1. Chance, B., et al. A novel method for fast imaging of brain function, non-invasively, with light. *Optics Express*, *10* (2). 411-423, 1988.
2. Gratton, G., et al. Rapid Changes of Optical Parameters in the Human Brain During a Tapping Task. *Journal of Cognitive Neuroscience*, *7*. 446-456, 1995.
3. Hirshfield, L.M., et al., Human-Computer Interaction and Brain Measurement Using Functional Near-Infrared Spectroscopy. in *Symposium on User Interface Software and Technology: Poster Paper*, (2007), ACM Press.
4. Hirshfield, L.M., et al., Uncovering Syntactic Workload in the Brain Using Functional Near Infrared Spectroscopy. in *Conference on Human Factors in Computing Systems: Proceeding of the twenty-seventh annual SIGCHI conference on Human factors in computing systems*, ((submitted) 2009).
5. Izzetoglu, K., et al. Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks. *International Journal of Human-Computer Interaction*, *17* (2). 211-231, 2004.
6. Izzetoglu, M., et al. Functional Near-Infrared Neuroimaging. *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, *13* (2). 153-159, 2005.
7. Jacob, R.J.K., et al., Reality-based interaction: a framework for post-WIMP interfaces. in *Conference on Human Factors in Computing Systems: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, (2008).
8. John, M.S., Kobus, D., Morrison, J. and Schmorrow, D. Overview of the DARPA Augmented Cognition Technical Integration Experiment. *International Journal of Human-Computer Interaction*, *17* (2). 131-149, 2004.
9. Marshall, S., Pleydell-Pearce, C. and Dickson, B. Integrating psychophysiological measures of cognitive workload and eye movements to detect strategy shifts. *IEEE. Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, 2003.
10. Parasuraman, R. and Caggiano, D. Neural and Genetic Assays of Human Mental Workload. in *Quantifying Human Information Processing*, Lexington Books, 2005.
11. Sassaroli, A., et al. Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy. *accepted in the Journal of Innovative Optical Health Sciences*, 2009.
12. Shneiderman, B. and Plaisant, C. Designing the User Interface: Strategies for Effective Human-Computer Interaction, Fourth Edition, Addison-Wesley, Reading, Mass., 2005.
13. Son, I.-Y., et al. Human performance assessment using fNIR. *Proceedings of SPIE The International Society for Optical Engineering*, *5797*. 158–169, 2005.
14. Wickens, C., Lee, J., Liu, Y. and Becker, S. *An Introduction to Human Factors Engineering*. Pearson, 2004.