Changes in Mental Workload during Task Execution

Shamsi T. Iqbal, Piotr D. Adamczyk♦, Xianjun Sam Zheng†, and Brian P. Bailey
Department of Computer Science, School of Library and Information Science♦, Beckman Institute†
University of Illinois, Urbana-Champaign 61801
{siqbal, pdadamcz, xzheng, bpbailey}@uiuc.edu

ABSTRACT

To contribute to systems that reason about human attention, our work empirically demonstrates how a user's mental workload changes during task execution. We conducted a study where users performed an interactive hierarchical task and measured mental workload through the use of pupil size. Results show that (i) different types of subtasks impose different mental workload, (ii) workload decreases at subtask boundaries, and (iii) effective understanding of why changes in workload occur requires that the measure of workload be tightly coupled to a validated task model.

CATEGORIES AND SUBJECT DESCRIPTORS

H.5.2 [Information Interfaces and Presentation]: User Interfaces — evaluation/methodology, user-centered design

GENERAL TERMS

Human Factors, Design, Experimentation, Measurement

KEYWORDS

Interruption, task models, attention, forecasting, pupil size

INTRODUCTION

When applications interrupt a user at an inopportune moment during task execution, the user performs tasks slower, commits more errors, makes worse decisions, and experiences more frustration, annoyance, and anxiety than if it had interrupted at a more opportune moment [8]. To mitigate the disruptive effects of interruption, researchers are investigating systems that reason about when to interrupt users [6, 7]. These systems compute the cost of interruption using external cues such as desktop activity, visual, and acoustical analyses of the physical environment, and scheduled activities of the user. To compute a more accurate cost of interruption, these systems need a direct measure of a user's mental workload. This work investigates the use of pupil size to provide such a measure.

RELATED WORK

Researchers have long argued that opportune moments for interruption occur at periods of low mental workload during task execution [2, 4]. Miyata and Norman posit that moments of low mental workload occur at subtask boundaries during task execution [10]. However, interactive tasks are composed of hierarchical patterns of goal formulation, execution, and evaluation, creating many levels of boundaries in the task model. Our work seeks to empirically demonstrate how a user's mental workload changes during execution of an interactive task, focusing on subtask boundaries in the task model.

Research shows that pupil size is a reliable measure of mental workload [3, 5], where the increase in pupil size correlates with the increase in workload. A review [3] of a large corpus of experimental data concludes that pupillary response reliably indicates mental workload for a task, that the degree of pupillary response correlates with the workload of the task, and that this holds true between tasks and individuals. In [9], we showed that pupillary response correlates with the mental workload of interactive tasks and discovered that changes in mental workload align well with the hierarchical model of the task being performed.

USER STUDY

The purpose of our study is to better understand how a user's mental workload changes during task execution, whether it decreases at subtask boundaries, and whether and to what extent it aligns with a validated task model. 12 users (1 female, ranging from 23 to 50 years of age, M=27.1, SD=7.45) were asked to perform an interactive route planning task as quickly and accurately as possible. While a user performed the task, we measured mental workload by measuring relative changes in the user's pupil size using a head-mounted SR Inc. Eyelink II eye tracking system. Pupil and eye movement data and videos of screen interaction were time-synched and logged for later analysis.

Task, Subtasks, and Validation

In the route planning task, users were required to add distances and fares associated with route segments, and asked to select the cheapest and shortest routes. We designed the task to be comprised of meaningful subtasks of varying difficulty, to have a prescribed execution sequence, to have well defined boundaries among subtasks, and to provide a representative sample of user interaction. Though users do not typically follow a set execution sequence, this had to be controlled to reliably link changes in mental workload to task execution. We developed and validated a GOMS model for the task. Across users, the average error rate was 2.81%, ranging from 0% to 5.66%. Thus, the GOMS model accurately reflects a user's execution of the task and enabled us to precisely align each user's pupillary response to the task model.

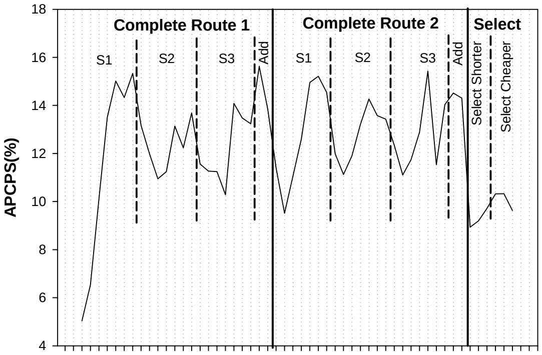

Figure 1: The graph shows the APCPS for each subtask in the model. Solid lines indicate high level boundaries and dashed lines indicate low level boundaries. Notice the graph dips lower at high level boundaries – a pattern that could be exploited in a system reasoning about or forecasting user attention.

Measurements

To measure changes in mental workload, we calculated the percentage change in pupil size (PCPS). This value was calculated by subtracting the baseline pupil size from the measured pupil size and dividing the result by the baseline. Consistent with prior work [3], PCPS is used to minimize pupillary response differences across users. The term APCPS refers to the average PCPS over a time window.

RESULTS

Figure 1 shows the pupillary response curve for users. The curve's shape shows changes of a user's mental workload during execution of the task. While we can only briefly discuss results here, a complete analysis is available in [8].

Mental workload at subtasks

We found a 12.7% increase over the baseline level for subtasks associated with memory and computation. Additionally, computation subtasks induced more mental workload than memory subtasks (1.7 point difference in APCPS, with $p<0.037$). Together with the short duration of some high workload subtasks, these results imply that a system can mitigate the disruptive effects of an interruption by deferring until a user is executing a lower workload subtask. This does not, however, require the system to defer the interruption for an extended period of time; waiting just a few seconds can considerably mitigate disruption [1, 2].

Mental workload at subtask boundaries

We compared the minimum PCPS at a subtask boundary to both the last PCPS measure in the preceding subtask as well as to the APCPS over the execution of the preceding subtask. From both perspectives, we found that a user's mental workload decreased at the subtask boundary. We compared paired differences between the minimum PCPS at a subtask boundary and the last PCPS measured in the preceding subtask across different levels in the task model. The difference between the pairs was greater at higher levels in the task model and smaller at lower levels in the task model (See Figure 1). A reasonable interpretation is that a user releases more cognitive resources when completing the final subtask of a larger goal chunk (higher in the model) than when completing the final subtask of a smaller goal chunk.

IMPLICATIONS

Our results show that a system could use a task model alone to roughly infer *where* a user's mental workload may change during task execution. However, our results empirically demonstrate that a system requires a tightly coupled measure of mental workload to understand *how much* a user's mental workload changes at those points. Combining the model of task execution and this fine-grained and direct measure of mental workload gives strong evidence pointing to a set of more opportune moments for interruption. This can be incorporated in systems that forecast mental workload or reason about when to interrupt users. Knowing not only how much a user's mental workload has changed, but also patterns of future workload like those in Figure 1, should enable a system to make better decisions about when to interrupt the user. Our ongoing work is investigating such a system.

REFERENCES

1. Adamczyk, P.D. and Bailey, B.P., If Not Now When? The Effects of Interruptions at Various Moments Within Task Execution. in *CHI*, (2004), 271-278.
2. Bailey, B.P., Konstan, J.A. and Carlis, J.V., The Effects of Interruptions on Task Performance, Annoyance, and Anxiety in the User Interface. in *Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction*, (Tokyo, Japan, 2001), 593-601.
3. Beatty, J. Task-evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin*, *91* (2). 276-292.
4. Gillie, T. and Broadbent, D. What Makes Interruptions Disruptive? A Study of Length, Similarity, and Complexity. *Psychological Research*, *50*. 243-250.
5. Hoecks, B. and Levelt, W. Pupillary Dilation as a Measure of Attention: A Quantitative System Analysis. *Behavior Research Methods, Instruments, & Computers*, *25*. 16-26.
6. Horvitz, E. and Apacible, J., Learning and Reasoning about Interruption. in *ICMI*, (2003).
7. Hudson, S.E., Fogarty, J., Atkeson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C. and Yang, J., Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. in *CHI*, (2003), 257-264.
8. Iqbal, S.T., Adamczyk, P.D., Zheng, X.S. and Bailey, B.P. Towards an Index of Opportunity: Understanding Changes in Mental Workload During Task Execution. *UIUC Technical Report* (UIUCDCS-R-2004-2431).
9. Iqbal, S.T., Zheng, X.S. and Bailey, B.P., Task Evoked Pupillary Response to Mental Workload in Human-Computer Interaction. in *CHI*, (2004), 1477-1480.
10. Miyata, Y. and Norman, D.A. The Control of Multiple Activities. in Norman, D.A. and Draper, S.W. eds. *User Centered System Design: New Perspectives on Human-Computer Interaction*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.