Investigating the Effectiveness of Mental Workload as a Predictor of Opportune Moments for Interruption

Shamsi T. Iqbal and Brian P. Bailey
Department of Computer Science
University of Illinois
Urbana, IL 61801 U.S.A
{siqbal, bpbailey}@uiuc.edu

ABSTRACT

This work investigates the use of workload-aligned task models for predicting opportune moments for interruption. From models for several tasks, we selected boundaries with the lowest (Best) and highest (Worst) mental workload. We compared effects of interrupting primary tasks at these and Random moments on resumption lag, annoyance, and social attribution. Results show that interrupting tasks at predicted Best moments consistently caused less resumption lag and annoyance, and fostered more social attribution. Thus, the use of workload-aligned task models offers a systematic method for predicting opportune moments for interruption.

Author Keywords

Attention, Interruption, Mental Workload, Task Models

ACM Classification Keywords

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

INTRODUCTION

Experiments show that when applications interrupt user tasks, they disrupt a user's task performance and affective state [1, 2, 11] and that these effects depend on at least the time at which the interruptions are delivered [1, 5].

Miyata and Norman [12] speculate that these differing effects are due to different mental workload at the point of interruption. They argue that boundaries in a task are best for interruption because they represent moments of lower workload. However, because tasks can be decomposed into recursive patterns of goal formulation, execution, and evaluation, many boundaries exist, making it unclear as to *which* boundaries are better. If a rank order of moments for interruption could be predicted for tasks, systems could use this information to make more effective decisions about when to interrupt users engaged in the tasks.

In prior work [8, 9], we designed several interactive tasks, constructed corresponding GOMS models, and validated the models. As users performed the tasks in an experiment,

we monitored their changing mental workload using pupil size, measured with a head-mounted eye-tracker. Workload was then precisely aligned to the task models.

Using these workload-aligned models, this work compares interrupting tasks at boundaries with the lowest (Best) and highest (Worst) workload. In our study, users performed tasks interrupted at Best, Worst, and Random moments, and performed one task without interruption. Effects were measured using resumption lag, annoyance, and respect.

Results show that interrupting tasks at the predicted Best moments consistently caused less resumption lag and annoyance, and fostered more respect for the interrupting system. Interestingly, the difference in timing between the Best and other moments was *small*, just a few seconds, but there was a *large* mitigation in disruption. This implies that systems need not delay information for an extended time, just a few seconds can cause large mitigation of disruption.

Our results demonstrate that the use of workload-aligned task models offers a systematic method for predicting how opportune moments in a primary task are for interruption.

RELATED WORK

In prior work [8, 9], we developed workload-aligned models for several interactive tasks. With the hypothesis that interruptions are less disruptive at points of lower mental workload, our current work uses these models to compare low and high workload moments for interruption.

Prior work shows that interruptions disrupt a user's task performance and affective state [1-3, 10, 11, 16] and that timing of the interruption relative to the primary task influences those effects [1, 2, 5, 11, 13]. However, no empirically validated method is available for reliably predicting which points in a task are more or less opportune for interruption. Our work seeks to provide such a method by comparing effects of interruption at boundaries with less and more workload, consistent with remarks in [11].

Miyata and Norman [12] speculate that opportune moments for interruption occur at task (and subtask) boundaries. However, tasks can be decomposed into recursive patterns of goal formulation, execution, and evaluation. We are investigating use of workload-aligned models for selecting *which* boundaries are most opportune for interruption.

Czerwinski et al. [4, 5] tested effects of interrupting users at different phases of a task, but did not compare the actual

boundaries between those phases. However, while differing effects were found, there was no underlying model to help explain why they were found. Using workload-aligned task models, our work seeks to provide a systematic method not only for predicting opportune moments for interruption, but also for explaining *why* differing effects would be found.

USER STUDY

Our study explores the use of workload-aligned task models for selecting opportune moments in a task for interruption. From our existing models, we selected the boundaries with the lowest and highest workload as the best and worst moments for interruption. Our expectation was that the Best (low workload) boundary would be more opportune than the Worst (high workload) boundary. We also included a Random condition and no-interruption (control) condition.

Experimental Design

A repeated measures within-S design was used with Timing (Best, Worst, Random, None) and Task (Route Planning, Document Editing, Object Manipulation) as factors.

Users and Tasks

Twelve users (6 female) participated in the study. Ages ranged from 21 to 42. Users were not compensated for participating. The experiment used the following tasks:

- *Route Planning.* Users retrieved distance and fare information for two routes from an interactive map, entered it into a structured table, and added them. Users then selected the shorter and the cheaper of the routes.

- *Document Editing.* Users edited a document based on three review comments placed, and then saved the document with a name of their choice.

- *Object Manipulation.* Users dragged email messages and dropped them into appropriate folders based on classification rules applied to the subject of the emails.

While we developed sets of similar tasks for each category, we were careful to design them such that they would not induce workload patterns largely different from the existing models. Screenshots of the tasks can be seen in [8, 9].

For the interrupting task, users read a news article and selected the most appropriate title from three choices. The interrupting task was adapted from our prior work [1, 2].

Moments for Interruption

For Route Planning, Best was between completing the second route and selecting the shorter/cheaper route. Worst was between recalling and entering information in the table. For Document Editing, Best was between completion of the last edit and accessing the save menu item. Worst was between positioning the mouse at the intended location and typing changes to the text. For Object Manipulation, Best was between placing an email into a folder and preparing to access the next email. Worst was between selecting an email and starting to drag it towards the destination folder.

Experimental Setup

Delivery of interrupting tasks used a Wizard of Oz model. The experimenter observed a user's task using a RealVNC client connected over a high-speed LAN to minimize latency. At pre-defined moments, the experimenter used custom software to send an interrupting task to a user. Best and Worst moments were pre-defined from the task models and Random moments were delivered within a time interval based around average task times collected in a pilot study.

Procedure

Before each category, specific instructions were given to the user and a practice task was performed. For each category, users performed four task trials, one for each timing condition. Users were instructed to attend to an interrupting task as soon as it appeared and, once complete, resume the primary task. The interrupted task was presented in a modal window and covered the main work area of the primary task. Users were instructed to complete all tasks as quickly and as accurately as possible. After each task trial, users completed the NASA TLX and scales for annoyance and respect. The order of the categories, tasks, and timing conditions were randomized. The study lasted an hour.

Measurements

In the study we measured the following:

- *Subjective workload.* This was measured using the NASA TLX [6]. Users responded by marking a vertical line along continuous scales from low to high. Scores were combined into a single workload value following [6].

- *Resumption Lag.* This was measured as the time to resume the primary task after completing the interrupting task. Lag was measured as the time from closing the interrupting task window to the first keyboard or mouse action in the primary task in direction of the task goal.

- *Annoyance.* This was measured on a scale from low to high, similar to those used in the TLX. Annoyance was used as a measure of the user's affective state.

- *Respect.* Users rated how respectful the interrupting system was to the primary task, i.e., social attribution.

These measures have been used to measure effects of interruption in many prior studies [1, 2, 14].

RESULTS

Two-way ANOVAs (Task x Timing) were used to analyze the dependent measures.

Subjective Workload

Task had a main effect on Subjective Workload ($F(2,22)=22.01$, $p<0.001$). Post hoc analysis showed that Route Planning ($\mu=2.74$) induced higher subjective workload than both Document Editing ($\mu=2.25$, $p<0.002$) and Object Manipulation ($\mu=1.77$, $p<0.001$), while Document Editing induced higher workload than Object Manipulation ($p<0.003$). Timing did not influence Subjective Workload and there were no interactions.
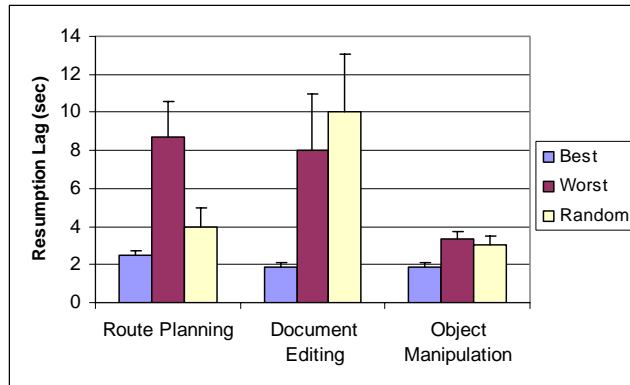
Figure 1. Time to resume primary task after attending to interrupting task (Resumption Lag)

Results show that the interrupting task did not induce workload beyond that of the primary task - regardless of timing. Results provide a consistent rank order of the tasks in terms of workload (Route Planning > Document Editing > Object Manipulation), which helps interpret later results.

Resumption Lag

Figure 1 shows a chart for Resumption Lag. Task had a main effect ($F_{(2,22)}=6.265$, $p<0.007$). Post hoc analysis showed that users resumed Object Manipulation tasks faster ($\mu=2.73s$) than Route Planning ($\mu=5.04s$, $p<0.012$) and Document Editing ($\mu=6.63s$, $p<0.004$) tasks after being interrupted. No other differences were found. This is roughly consistent with the workload measures, as the lowest workload task had the least resumption lag.

Timing had a main effect ($F_{(2,22)}=6.866$, $p<0.005$). Post hoc tests showed that users were able to resume the primary task almost 3 times faster after being interrupted at Best moments ($\mu=2.06s$) than at Worst ($\mu=6.69s$, $p<0.03$) and Random ($\mu=5.66s$, $p<0.007$) moments. This may be explained by users needing to acquire fewer resources to resume the primary task after being interrupted [15]. No other differences were found and there were no interactions.

Annoyance

Figure 2 shows a chart of user ratings of annoyance. Task had a main effect ($F_{(2,22)}=7.057$, $p <0.004$). Post hoc tests showed that Route Planning ($\mu=2.43$) caused users to experience more Annoyance than Document Editing ($\mu=1.6$, $p<0.025$) and Object Manipulation ($\mu=1.5$, $p<0.004$). No other differences were found. The results are mostly consistent with the workload measures, as the highest workload task caused the most annoyance and the lowest workload task caused the least annoyance.

Timing had a main effect ($F_{(3,33)}=7.414$, $p<0.001$). Post hoc tests showed interruptions at Best ($\mu=1.69$) moments caused less annoyance than at Worst ($\mu=2.07$, $p<0.079$) and at Random ($\mu=2.35$, $p<0.052$) moments. Not surprisingly, users experienced less annoyance when not interrupted
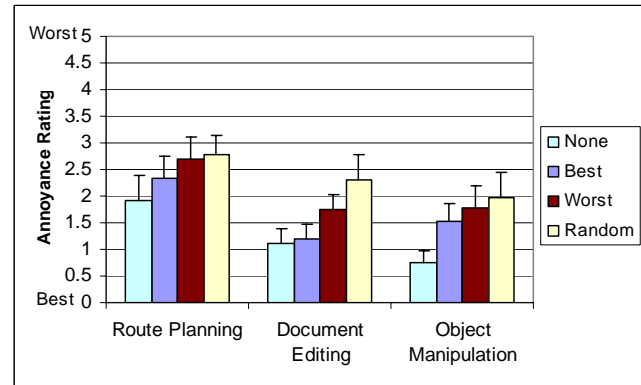

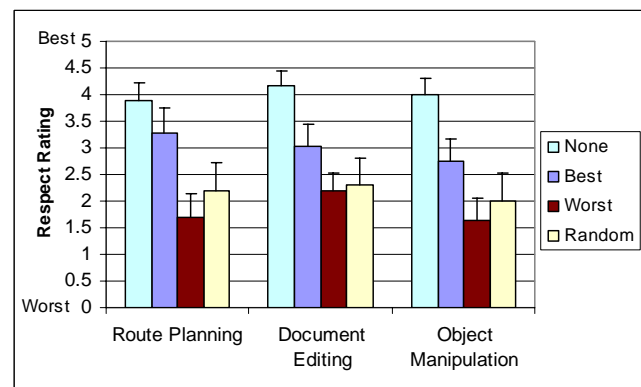**Figure 2. Ratings of Annoyance**

Figure 3. Ratings of Respect

($\mu=1.26$) than when interrupted ($p<0.043$ in all cases). No other differences were found and there were no interactions.

Respect

Figure 3 shows a chart of user ratings of respect they attributed to the interrupting system. Task had no main effect ($F_{(2,20)}=0.701$, $p<0.508$), while Timing did have a main effect ($F_{(3,33)}=12.105$, $p<0.001$). Post hoc tests showed that users rated the system to be *more* respectful to their primary task when interrupted at Best ($\mu=3.02$) moments than at Worst ($\mu=1.85$, $p<0.015$) and at Random ($\mu=2.17$, $p<0.086$) moments. Users rated the system to be most respectful to their primary task when they were not interrupted ($\mu=4.02$) than when they were interrupted ($p<0.001$). No interactions were found.

DISCUSSION

Two important lessons were learned from this study:

- *Timing of an interrupting task relative to a primary task influences the amount of disruption that it causes.* Across tasks, interrupting a task at the Best moments was *less* disruptive – it caused less resumption lag and annoyance, and fostered more respect - than at Random or Worst moments. Also, the difference in timing between the Best and other moments was small, a few seconds on average, but the mitigation in disruption was meaningfully large.

- *Mental workload is an effective predictor of opportune moments for interruption.* For each task, Best and Worst moments were selected by identifying the boundaries with the lowest and highest workload in the workload-aligned task models. Since interrupting the primary task at Best and Worst moments caused consistently less and more disruption, respectively, use of workload appears to be an effective predictor of how opportune boundaries are for interruption. Thus, the use of workload-aligned task models offers a promising step towards a systematic method for selecting opportune moments for interruption.

A practical limitation of this work is that it requires one to develop or have access to workload-aligned task models. However, these models only need to be developed *once*. Practitioners or researchers who have access to appropriate equipment (such as eye-trackers that measure pupil size) can develop these models in controlled settings and then make them available for broader use. As producing models for all possible tasks is probably not realistic, producing models targeted for high frequency or safety critical tasks is realistic and should yield valuable results.

From the models, a rank order of opportune moments for interrupting the tasks would be made available to reasoning systems, e.g. [7]. These systems would then perform fine-grained manipulation because, as our results show, a *small* difference in timing causes a *large* mitigation of disruption.

Mitigating disruption by manipulating when primary tasks are interrupted is important for safety critical and office environments. In safety critical domains, enabling operators of complex systems to get back on task quickly after an interruption may prevent catastrophic accidents [11].

In office environments, reducing user annoyance enables a more pleasing and satisfying work environment. Increasing social attribution to applications such as peripheral displays is important so that users find the provided services useful. Otherwise, users just shut them off, losing the benefits they seek to provide. While the difference made by manipulating a *single* interruption may not make a difference in practical settings, we believe that over weeks or months, even years, manipulating interruptions such that they occur at opportune moments in task sequences will indeed make an enormous practical and positive difference for users.

CONCLUSION AND FUTURE WORK
This paper explored the use of workload-aligned task models for predicting opportune moments for interruption. From the models, we selected boundaries with the lowest and highest workload as the best and worst moments for interruption and then compared the effects of interrupting users at these and other moments. Results showed that interrupting primary tasks at the predicted best moments reliably caused less resumption lag and annoyance, and fostered more social attribution. By providing interruption reasoning systems with moments predicted by workload-

aligned task models, these systems can make more effective decisions about when to interrupt users engaged in tasks. Future work seeks to replicate these findings using a more diverse set of interactive tasks and to further compare interruption points predicted by alternative methods.

REFERENCES
1. Adamczyk, P.D. and B.P. Bailey. If Not Now When? The Effects of Interruptions at Various Moments within Task Execution. *CHI*, 2004, 271-278.

2. Bailey, B.P., J.A. Konstan and J.V. Carlis. The Effects of Interruptions on Task Performance, Annoyance, and Anxiety in the User Interface. *INTERACT*, 2001, 593-601.

3. Cutrell, E., M. Czerwinski and E. Horvitz. Notification, Disruption and Memory: Effects of Messaging Interruptions on Memory and Performance. *INTERACT*, 2001, 263-269.

4. Czerwinski, M., E. Cutrell and E. Horvitz. Instant Messaging and Interruption: Influence of Task Type on Performance. *OZCHI 2000 Conference Proceedings*, 2000, 356-361.

5. Czerwinski, M., E. Cutrell and E. Horvitz. Instant Messaging: Effects of Relevance and Timing. *People and Computers XIV: Proceedings of HCI*, 2000, 71-76.

6. Hart, S.G. and L.E. Stateland. Development of Nasa-Tlx (Task Load Index): Results of Emperical and Theoretical Research. In Hancock, P.A. and Meshkati, N. (eds.) *Human Mental Workload*, North-Holland, Amsterdam, 1988, 139-183.

7. Horvitz, E. and J. Apacible. Learning and Reasoning About Interruption. *Proc. ICMI*, 2003.

8. Iqbal, S.T., P.D. Adamczyk, X.S. Zheng and B.P. Bailey. Towards an Index of Opportunity: Understanding Changes in Mental Workload During Task Execution. *CHI*, 2005, to appear

9. Iqbal, S.T., X.S. Zheng and B.P. Bailey. Task Evoked Pupillary Response to Mental Workload in Human-Computer Interaction. *CHI*, 2004, 1477-1480.

10. Latorella, K.A. Effects of Modality on Interrupted Flight Deck Performance: Implications for Data Link. *Proc. HFES*, 1998.

11. McFarlane, D.C. Coordinating the Interruption of People in Human-Computer Interaction. *INTERACT*, 1999, 295-303.

12. Miyata, Y. and D.A. Norman. 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, 1986.

13. Monk, C.A., D.A. Boehm-Davis and J.G. Trafton. The Attentional Costs of Interrupting Task Performance at Various Stages. *Proc. HFES*, 2002.

14. Trafton, J.G., E.M. Altmann, D.P. Brock and F.E. Mintz. Preparing to Resume an Interrupted Task:Effects of Prospective Goal Encoding and Retrospective Rehearsal. *Intl. Journal of Human-Computer Studies*, 58, 583-603, 2003.

15. Wickens, C.D. *Engineering Psychology and Human Performance*. Scott, Foresman & Co, Boston, 1984.

16. Zijlstra, F.R.H., R.A. Roe, A.B. Leonora and I. Krediet. Temporal Factors in Mental Work: Effects of Interrupted Activities. *Journal of Occupational and Organizational Psychology*, 72, 163-185, 1999.