A Method, System, and Tools for Intelligent Interruption Management

Piotr D. Adamczyk, Shamsi T. Iqbal♦ and Brian P. Bailey♦

Division of Human Factors and Department of Computer Science♦
University of Illinois, Urbana, IL 61801
{pdadamcz, siqbal, bpbailey}@uiuc.edu

ABSTRACT

Interrupting users engaged in tasks typically has negative effects on their task completion time, error rate, and affective state. Empirical research has shown that these negative effects can be mitigated by deferring interruptions until more opportune moments in a user's task sequence. However, existing systems that reason about when to interrupt do not have access to task models that would allow for such finer-grained temporal reasoning. We outline our method of finding opportune moments that links a physiological measure of workload with task modeling techniques and theories of attention. We describe the design and implementation of our interruption management system, showing how it can be used to specify and monitor practical, representative user tasks. We discuss our ongoing empirical work in this area, and how the use of our framework may enable attention aware systems to consider a user's position in a task when reasoning about when to interrupt.

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 [1, 14].

Unfortunately, systems that attempt to manage human attention largely lack the facilities to make reasoned or even informed decisions about when to provide new information to users. To build systems that can accurately infer user availability and potentially forecast opportune moments for interruption, we have been exploring the use of task models augmented by an understanding of human attention and mental workload. Rather than focus on the specifics of any given instance of task execution, this approach allows us to focus on the factors that generalize across people, contexts, and tasks.

In a set of empirical studies, grounded in psychological theories of attention and interruption, we devised and adapted methods for building appropriate task models at various levels of detail. Using these, we selected and evaluated the systematically defined moments for interruption. Then we constructed a task monitoring system to help attention aware systems reason about a user's current position in a task, allow designers to build, learn, or infer models of user task execution, and engineered the system to be robust enough to accommodate a variety of tasks and use contexts.

INTERRUPTION MANAGEMENT

The rest of this section outlines our existing research in the field of interruption management and how it incorporates various aspects of task modeling. Each study's description includes a brief analysis of related work and presents the key results on which our methods and framework have been developed. We describe an initial empirical study using event perception theory to build simple task models grounded in psychology. We present a subsequent study using pupil size as a measure of mental workload to further validate the moments for interruption and refine a method for their selection. We go on to show how lessons from both of the empirical studies contribute to the design of a system and tools for task specification and monitoring.

Opportune Moments and Event Perception

McFarlane [18] suggests 4 strategies for coordinating interruption in HCI; immediate, negotiated, mediated, and scheduled. Each strategy has characteristics that make it more attractive for certain situations. For example, immediate interruption values new information over the current task, while scheduled interruption occurs at intervals to which users and systems may adjust their behavior.

Some studies [9, 10] place moments for interruption towards the beginning, middle, or end of a task. This rough temporal placement relates closely to work presented by

Miyata and Norman [20]. The authors explain that task execution occurs in three phases: planning, execution, and evaluation. But if this applies to a task, a logical extension would be that it would also apply to each of the component subtasks. As tasks in themselves, every subtask would then contain moments of planning, execution, and evaluation, making task execution a loop of these phases. There are clearly effects to interrupting during the various phases [27] but associating rough temporal placement (beginning, middle, end) might be an oversimplification of task execution.

Still others place interruptions between instances of repetitive action sequences [4, 19, 21]. The choice of these points is more intuitive but the reasoning behind these locations is often ill defined, sometimes producing internally inconsistent results [19].

Each of the approaches above depends on a model of attention when deciding how and when to interrupt, but none present a clear or easily generalized method for moment selection. To address this absence, we first used methods from studies on human event perception [24-26]. Human observers of events segment ongoing activity into temporal parts and sub-parts that are reliable, meaningful, and correlate with ecologically relevant features of the action [24]. This process of recognizing time-based boundaries was linked to distinct patterns of brain activity [24]. In an experimental follow-up [25], subjects were shown video recordings of tasks being performed, and then asked to communicate or recall the task structure. Subjects recalled events as hierarchies with two levels, coarse and fine. Coarse breakpoints largely represented the introduction of objects and broad actions on those objects, while Fine breakpoints were the more precise actions in the scene. The study showed how event structure influences recall of tasks and goals, and that moments that are best recalled are those that are more firmly related to schematic action – recognizable and well understood activities.

In an initial empirical work [1], we proposed that the best moments for interruption should be between two coarse breakpoints that are, on the whole, better understood and better recalled than other points in the task [6]. Having just completed a schematic event, the subject is utilizing fewer cognitive resources, leaving the rest immediately available for a peripheral task [6, 8]. Interruption triggers are based on behavior that there is good reason to believe is significant in the mind of the user, and the interruptions are not associated with a temporal phase, making it easier for them to be applied anywhere during execution.

To construct an event perception task model, subjects were shown an instance of a task and asked to construct their own two level hierarchies for document editing, directed web searching, and media summarization tasks. These procedures mirrored those in [25], and our own full procedures can be found in [1]. We posited that the level of agreement across user task models would reflect how schematic an action was for that task, reflecting its suitability as a point for interruption. We recorded how often certain breakpoints appeared in the task models, and selected the points with highest agreement as the best points for interruption. Worst points were those with least agreement.

While we found no improvement in objective measures, our results showed that our predicted best points for interruption consistently produced less annoyance, frustration, and time pressure, required less mental effort, and were deemed by the user, more respectful of the their primary task. Full results can be found in [1].

Use of Pupil Size and Validation of Moments

Our study using event perception methods suggested the importance of breakpoints in task models as key moments for interruption. To build on these results and explore automated methods for construction of task models and selection of opportune moments in these models we turned to pupil size as an immediate physiological metric reflecting the subject's mental workload.

Under conditions of controlled illumination, research shows that pupil size is an effective and reliable measure of mental workload [12, 22], where an increase in pupil size correlates with an increase in mental workload. Beatty reviewed a large corpus of experimental data, concluding that pupillary response is a reliable indicator of mental workload for a task, that the degree of pupillary response correlates with the workload of the task, and that this phenomenon holds true between tasks and individuals [5].

In [15], we showed that pupillary response correlates with the mental workload of *interactive* tasks and discovered that changes in mental workload seem to align well with the hierarchical model of the task being performed. To better understand this relationship we built GOMS models for two interactive tasks, route planning and document editing, and measured subject's pupil size through a head mounted eyetracker.

Briefly, an initial GOMS model was constructed through our own analysis of task execution, and the GOMS decomposition continued until there was no observable or meaningful separation between operators. The GOMS models were validated by a set of external reviewers, and based on their feedback, the models were refined as necessary. The average error rates for the route planning and document editing tasks were 2.81% and 2.3% respectively. A full description of procedures can be found in [14].
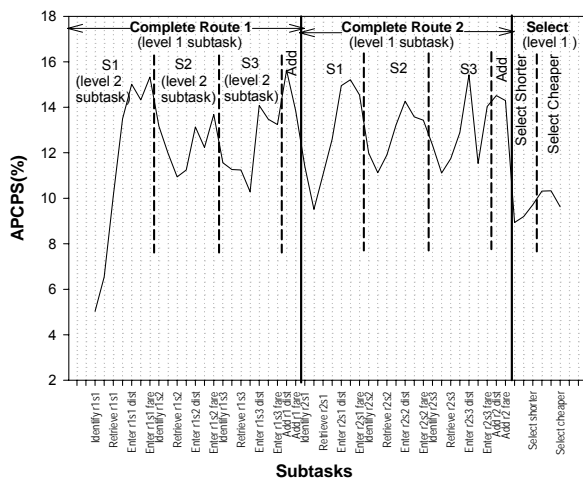
Figure 1: Average Percent Change in Pupil Size (APCPS) for the subtasks in the route planning task. Vertical lines indicate subtask boundaries. The x-axis enumerates level 3 GOMS subtasks. Notice how the graph dips lower at major (solid) boundaries than at minor (dashed) ones.

Figure 2: A region of a task hierarchical task model (top) aligned with a subject's pupillary response curve (bottom) as displayed in TAPRAV [2]. Time is displayed along the x-axis. Colored lines denote boundary moments across both data structures.

At this point the GOMS models accurately reflected the full structure of the task, and when aligned to the pupillary response graph (see Figure 1), gave further evidence for which moments were opportune for interruption.

Specifying and Monitoring User Tasks
Informed by the lessons learned from our studies using event perception and GOMS task models, we have developed a task specification and monitoring framework that facilitates the creation of interruption management, and more broadly, attention aware systems.

Our framework consists of four components: a task description language that draws upon GOMS, regular expressions, and schema descriptions to support expressive specification of tasks using a concise notation; a graphical tool that enables rapid assembly of task specifications; an event database and handler that manages user events from instrumented applications; and a task monitor that follows a user's progress through specified tasks, notifying user-level services when task-related events occur. Our framework has been evaluated for its ease of use by designers and effectiveness as a task specification and monitoring tool [3]. While a full description of our framework is described in [3], the following presents some key system features.

An important contribution of our framework is that it provides an open architecture, enabling tasks involving any application with appropriate instrumentation to be monitored and any user-level service to be notified when task-related events occur. Our framework thus enables systems to have access to accurate information about a user's current position in a task sequence, important for intelligent tutoring systems [7], software agents [16, 17], and attention aware systems that manage interruption [13].

In contrast to other similar systems [7] that allow context-specific instructions to be integrated into task models at multiple levels of detail, our monitoring system can accurately follow tasks if the user switches from executing an ongoing task to another task or performs multiple tasks at the same time.

Many systems, e.g., [7, 11, 23] monitor the user event stream and compare events to a task model in order to provide context-sensitive instruction or feedback. While our system provides a similar function, it also attempts to learn a flexible model of task execution and record that model in a user profile.

Future Work
As future work, we are building tools like TAPRAV [2] which align task models with real time (e.g. physiological) data streams for more in depth task analysis (see Figure 2). The TAPRAV interface may be combined with the graphical tool in our framework to provide a way to interactively build detailed task models through visual inspection of task recordings.

Further empirical research needs to be conducted comparing moments for interruption outlined by alternate task models. Also key, are acceptable amounts of system lag for real time interruption systems. These would serve as a metric by which to compare and evaluate systems.

We are continuing work validating the opportune moments suggested by our event perception and GOMS/Pupil size models. In particular we are considering applications beyond the desktop for aircraft cockpits, in-vehicle systems, and ubiquitous computing.

CONCLUSION
Refined task models are of crucial importance to interruption management systems if they are to effectively provide users significant benefit. Through our ongoing work, we hope to provide researchers and systems designers with effective tools for creating, analyzing, and implementing task models for existing attention aware applications, and support the creation of new implementations. Our methods identifying opportune moments in tasks are grounded in psychological theory

allowing for a more informed analysis. Our framework for specifying and monitoring tasks is tuned for office applications, but components can be used independently, making it immediately useful in other domains.

REFERENCES

1. Adamczyk, P.D. and Bailey, B.P., If Not Now When? The Effects of Interruptions at Various Moments Within Task Execution. in *Proceedings of the ACM Conference on Human Factors in Computing Systems*, (2004), 271-278.

2. Adamczyk, P.D., Busbey, C.W. and Bailey, B.P. TAPRAV: A Tool for Exploring Physiological Data Aligned to Task Models. *Report No. UIUCDCS-R-2005-2562*.

3. Bailey, B.P., Adamczyk, P.D., Chang, T.Y. and Chilson, N.A. A Framework for Specifying and Monitoring User Tasks. *Journal of Computers in Human Behavior, special issue on attention aware systems*.

4. Bannon, L., Cypher, A., Greenspan, S. and Monty, M.L., Evaluation and analysis of users' activity organization. in *CHI*, (1983), 54-57.

5. Beatty, J. Task-evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin*, *91* (2). 276-292.

6. Cellier, J.M. and Eyrolle, H. Interference between switched tasks. *Ergonomics*, *35* (1). 25-36.

7. Cheikes, B.A., Geier, M., Hyland, R., Linton, F., Rodi, L. and Schaefer, H.-P. Embedded Training for Complex Information Systems. *International Journal of Artificial Intelligence in Education*. 314-334.

8. Cohen, S. Aftereffects of Stress on Human Performance and Social Behavior: A Review of Research and Theory. *Psychological Bulletin*, *88* (1). 82-108.

9. Cutrell, E., Czerwinski, M. and Horvitz, E., Notification, Disruption and Memory: Effects of Messaging Interruptions on Memory and Performance. in *Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction*, (Tokyo, Japan, 2001), 263-269.

10. Czerwinski, M., Cutrell, E. and Horvitz, E., Instant Messaging: Effects of Relevance and Timing. in *People and Computers XIV: Proceedings of HCI*, (2000), British Computer Society, 71-76.

11. Franklin, D., Budzik, J. and Hammond, K., Plan-based Interfaces: Keeping Track of User Tasks and Acting to Cooperate. in *International Conference on Intelligent User Interfaces*, (2002), 79-86.

12. Hoecks, B. and Levelt, W. Pupillary Dilation as a Measure of Attention: A Quantitative System Analysis. *Behavior Research Methods, Instruments, & Computers*, *25*. 16-26.

13. Horvitz, E., Jacobs, A. and Hovel, D., Attention-Sensitive Alerting. in *Conference Proceedings on Uncertainty in Artificial Intelligence*, (1999), 305-313.

14. Iqbal, S.T., Adamczyk, P.D., Zheng, S. and Bailey, B.P., Towards an Index of Opportunity: Understanding Changes in Mental Workload during Task Execution. in *Proceedings of the ACM Conference on Human Factors in Computing Systems*, (2005), 313-322.

15. Iqbal, S.T., Zheng, X.S. and Bailey, B.P., Task Evoked Pupillary Response to Mental Workload in Human-Computer Interaction. in *Proceedings of the ACM*

16. Lieberman, H., Autonomous Interface Agents. in *Proceedings of the ACM Conference on Human Factors in Computing Systems*, (1997), 67-74.

17. Maes, P. Agents that Reduce Work and Information Overload. *Communications of the ACM*, *37* (7). 30-40.

18. McFarlane, D.C. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction*, *17* (1). 63-139.

19. Miller, S.L., Window of opportunity: Using the interruption lag to manage disruption in complex tasks. in *To appear in: Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society.*, (Santa Monica, CA, 2002), Human Factors and Ergonomics Society.

20. 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.

21. Monk, C.A., Boehm-Davis, D.A. and Trafton, J.G., The Attentional Costs of Interrupting Task Performance at Various Stages. in *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, (2002).

22. Nakayama, M. and Takahashi, K., The Act of Task Difficulty and Eye-movement Frequency for the Ocul-motor Indices. in *Proceedings of Eye Tracking Research and Application*, (2002), 37-42.

23. Rich, C. and Sidner, C.L. COLLAGEN: A Collaboration Manager for Software Interface Agents. *User Modeling and User-Adapted Interaction*, *8* (3/4). 315-350.

24. Zacks, J., Braver, T.S., Sheridan, M.A., Donaldson, D.I., Snyder, A.Z., Ollinger, J.M., Buckner, R.L. and Raichle, M.E. Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4* (6). 651-655.

25. Zacks, J. and Tversky, B. Event structure in perception and cognition. *Psychological Bulletin*, *127* (1). 3-21.

26. Zacks, J., Tversky, B. and Iyer, G. Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130* (1). 29-58.

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