The Effects of Highlighting and Pop-up Interruptions on Task Performance

Pablo-Alejandro Quinones, Jigar Vora, Aaron Steinfeld, Asim Smailagic, Jeffery Hansen, Dan P. Siewiorek, Pete Phadhana-Anake, Abhishek Shah
Carnegie Mellon University
Pittsburgh, PA

Prior research suggests that interruptions by software components are undesirable and detrimental in many work scenarios. However, there are clear instances where interruptions can conceivably provide a net benefit. For example, interruption is appropriate when reminding a user to accomplish an important task instead of working on lower value activities. This paper examines the pros and cons of interruption and how interruption should occur in the context of an integrated intelligent assistant system. Results from a study and future directions are discussed.

INTRODUCTION

As long as the crux between work and technology has existed, interruptive technologies have always existed, ranging from people to phones to instant messaging (IM) systems. As technologies become more available, the ubiquity of interruptions and interruptive technologies will become more evident. Interruptive technologies, by which we mean any software that makes use of alerts or interruptions, already have a bad reputation in part because of the assumption that interruptions are inherently unfavorable and extensive literature on their disruptive effects.

Interruptions can become disruptive to workflows, however there are instances where interruptions may be favorable compared to the alternative—remaining clueless to urgent situations. To illustrate some examples, urgent tasks may present themselves while doing other things and an important task may be ignored in favor of a less important one.

In recent years, research groups have created interruptive software such as recommender systems and mixed initiative intelligent systems, which are designed to make suggestions in real time to improve task performance. This paper will first present previous efforts related to work, interruptions, and intelligent assistants. Then we will present an empirical study concerned with the types of interruption to be used within a larger intelligent assistant system.

RELATED WORK

Intelligent Assistants

Complex programs, multiple tasks, and information overload are all factors that weigh heavily not only in performance metrics, but also in affect and cognitive load. With current advances in machine learning, many researchers and developers have sought to create technology to assist users in their activities, preempting their needs and sometimes adjusting to them (Nirenburg & Lesser, 1986). We call these types of technologies *Intelligent Assistants* due to their ability to adapt. Systems like these are designed to help users perform a specific task or a set of tasks more efficiently by undertaking some of the work, providing assistance, hints, and tips when needed or they provide important information. The work described here was used to direct the design of an interruption component within the RADAR intelligent assistant (Freed et al, 2008).

In general, developers are faced with the difficulty of balancing the tradeoffs of deciding if and when to interrupt, whether help is needed, the benefits of receiving help, the cost to tasks when receiving help, the cost of not helping, and the cost to providing the wrong kinds of help (Hui & Boutilier, 2006).

Nature of Interrupted Work

With increased technology, work tasks have become increasingly complex, consisting of multiple subtasks. This complexity often requires a degree of multi-tasking, which affects the order and methods in which tasks are completed.

It is a known fact in psychology and human factors that cognitive resources, specifically memory and attention are limited, forcing users to differentiate tasks and prioritize them for the sake of efficiency. According to Miyata and Norman (1986), the flow of tasks undergoes a transformation into a generally linear form with few tasks that are currently being worked on and tasks that, for any number of reasons, have been set aside to be returned to later (suspended tasks), however, over the course of daily activities, interruptions often occur.

Interruptions can be either internal (self-created) or external (from an outside source), introducing new information and updates, tasks, or advice. Interruptions may also be social in nature, such as in IM, email, or talking with coworkers. Interruptions create breaks in the workflow and cause one to suspend current work. Resuming suspended work may be difficult depending on factors such as the task or motivation from the worker. Continuation of disrupted tasks requires one to finish attending to the interruption and regain context in the original task.

Interruptions can be either relevant or irrelevant to the task. Czerwinski et al (2000) found that the total amount of time spent on a distraction and returning to the original task increased when the interruption was irrelevant.

According to Mark et al (2005), most research on interruptions highlights the negative impacts they have on work due to the belief that work fragmentation, breaks in continuous workflow, are unfavorable. Their work showed that interruptions to workflow cause people to resume work slowly. Other previous research suggests that that task performance and affective states (such as frustration, anxiety, and annoyance) are affected by interruptions introducing secondary tasks during activities (Adamczyk & Bailey, 2004, Adamczyk et al, 2005, Bailey et al, 2001, Bailey & Konstan, 2006), instant messages, as well as when receiving advertisement pop-ups (Edwards et al, 2002). Bailey and Konstan's research (2006) suggests that interruptions in themselves may not be the cause of errors, but perhaps the expectance of interruptions may be to blame. Speier et al (2003) found that interruptions affected main task speed-accuracy tradeoff. High frequency interruptions during complex decision making tasks reduces accuracy and increases speed, while low levels of interruption lower task speed and raise accuracy. Oulasvirta and Salovaara (2004) also found switching attention between tasks negatively impacts memory.

With so much work focusing on the negative aspects of interruption, it would seem that interruption should be avoided. Indeed, very little research highlights the positive aspects of interruption. Interruptions are almost necessary, since they can provide new and desirable information as well as urgent updates and tasks. They may also be used to direct someone to neglected tasks.

This is in line with work by Miyata and Norman (1986) describing two extremes of active work states: one where the focus is on a primary task and interruptions and secondary tasks are ignored (task-driven), and another where one is sensitive to interruptions to the primary task (interrupt-driven). A system with good interaction design will recognize the latter state interrupt at those times.

Interruption Rendering

Methods of interruption are many. Pop-ups are undeniably ubiquitous in computing systems. Instant message windows usually pop-up when initiating conversation or may blink when an already open conversation window is in the background. Dialog boxes pop-up when one closes a document without saving first. It is also common to receive notification through icons, such as the System Tray in Windows, where icons may blink or change form. Billsus et al (2005) created and tested a toolbar for Microsoft Internet Explorer and Gluck (2007) introduced a large number of notification methods based on icons. Notifications may also occasionally make sounds. With different alert methods to choose from, selecting the correct one can be a difficult design decision.

When deciding upon notification types, there is a balance between drawing attention and trying not to be bothersome, subtlety vs. intrusiveness. This consideration, attentional draw, refers to the amount of attention attracted by the notification signal and can be thought of as the amount of distraction. This matter is very closely related to the cost of interruption.

Gluck's studies (2007) suggest that users dislike notifications with high attentional draw when the notification has no content and is irrelevant to the task. Also, users view notifications favorably when there is content and are relevant to the task. The implication is that interruptions that attract a lot of attention should convey important information.

McCrickard et al's model for notification systems (2003) produces similar design advice. This model is composed of three factors: *Interruption*, or the degree to which the signal disrupts workflow; *Response*, or the requirement to act once the signal is detected; and *Comprehension*, or being able to remember or make sense of it at later points in time. According to the model, these ought to be balance. But moreover, these three factors can correlate to different signal types. For example, an alarm is highly disruptive and requires a response, but has low comprehension. They advise that to create proper notification types, one must consider the user and the system, and then decide which kind of notification is most appropriate. Under this design model, highly disruptive notifications may actually be desirable.

The purpose of this study was to determine, which method of interruption is best given an interruption notification of an urgent task. The two methods of user alerts were pop-up messages and highlighting a line from a list of items. These represent the two domains of alerts, intrusive alerts which interrupt an action by placing themselves prominently, and unobtrusive alerts, which present themselves quietly, usually in a taskbar as an icon with some form of attention getting signal, i.e., blinking, moving, pulsing, etc (Gluck, 2007).

Highlighting a task from a list might be less obtrusive and less annoying than a popup. This type of interruption does not seem to have been examined in detail and may be helpful in recommender systems if task lists have high visibility. Pop-ups may elicit faster reaction times because of their high level of intrusiveness. This study compares the two methods and their combination to see which method is best for notifying a user about urgent tasks. Also, combining the two methods may be redundant, however, making them less desirable than pop-ups.

> Hypothesis 1a: Highlighting tasks from a list will be seen as more effective than pop-ups.

> Hypothesis 1b: Highlighting tasks from a list in addition to a pop-up will be seen as less effective than pop-ups.

Users may be annoyed with pop-ups and may respond more slowly to them. At the same time, pop-ups grab attention easily and may elicit faster responses than highlighting simply because they are noticed more quickly. The danger of highlighting, or any other sort of non-invasive interruption techniques, is that they may remain unnoticed by the users if they are too mild. The combination of methods might work best, so that if the highlighting is not noticed, an eventual pop-up might be.

> Hypothesis 2a: Highlighting tasks from a list will elicit slower response times than pop-ups.

> Hypothesis 2b: Highlighting tasks from a list in addition to a pop-up will elicit response times as fast as pop-ups.

> Hypothesis 2c: Highlighting tasks from a list in addition to a pop-up will elicit response times as faster than highlighting.

EXPERIMENT

Task

To simulate real world situations, we used three kinds of tasks – computation, 'Look Up' (information retrieval/cross-reference), and Sudoku (a popular number game) Each type of task had three levels of difficulty – easy, medium, and hard. Each task was composed of subtasks, each requiring a previous action to be done before being complete. This enables the interrupting agent to find task and subtask boundaries.

The computation task requires subjects to solve math problems. Easy problem sets involved simple addition or subtraction between one-digit or two-digit numbers with no subtasks. Medium problem sets involved simple addition or subtraction between two-digit numbers, which should require some mental math and optional use of paper and pencil. Subtasks were included in this problem set. Hard problems involved multiplication of two-digit numbers and normally required the use of paper and pencil. Subtasks were included in this problem set as well.

Look Up tasks involved referencing lists of compiled information which may be sorted or unsorted, and extracting a particular piece of information. Furthermore, several lists may need to be examined before an answer could be discovered. We classified levels of difficulty based on the level of 'cross-referencing' needed to answer the question. Easy questions involved a search into one data set, such as looking for a phone number or a person's email address. There were no subtasks. Medium questions required the user to search two different data sets to find the answer. Subtasks were included and are defined as the switch between lists. Hard questions required the user to search two different data sets to find the answer and may have required noting pieces of information in addition to cross-referencing the two data sets. Subtasks were defined as switches between lists.

The last task was Sudoku, a popular logic number game. The objective is to fill a 9x9 grid so that each column, each row, and each of the nine 3x3 boxes contained the digits 1 through 9. The puzzle provides a partially completed grid so that there is only one possible solution. This task generally requires a more constant stream of concentration as answers are highly correlated. Levels of difficulty are classified by the number of fields missing from the puzzle. All Sudoku fields are considered subtasks as the solutions are dependent on the other values in the puzzle. Easy puzzles required subjects to solve a Sudoku puzzle with 7 missing fields. Medium puzzles required subjects to solve a Sudoku puzzle with 10 missing fields. Hard puzzles required subjects to solve a Sudoku puzzle with 12 missing fields.

Design and Procedure

Over the course of an hour, subjects participated in 15 minute rounds, each having a different form of interruption styles. Subjects could choose in which order they wanted to do their given tasks. Interruptions were administered at intervals determined by experimenters. Interruptions requested that a particular task be done, although they do not prevent the subjects from continuing the experiment in their own order. The order of the conditions was randomized for each subject. Five minutes were devoted to consent forms at the beginning and a five-minute break was provided between each of the two rounds.

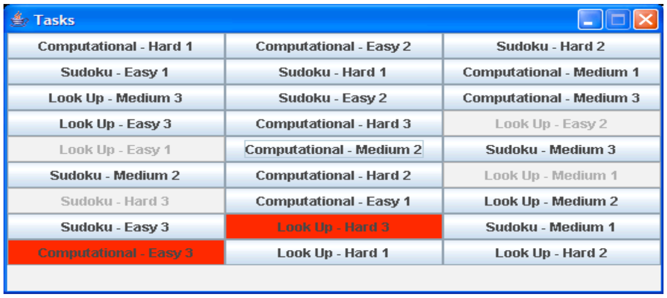

Figure 1: Highlight Condition

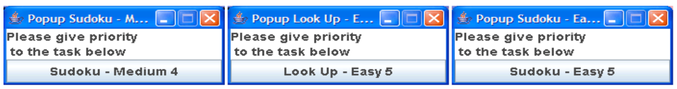

Figure 2: Pop-up Condition

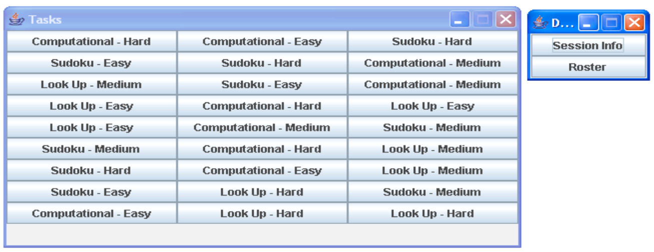

Figure 3: Example screen from the Interruption Study

Each round was composed of 27 completely different sets of questions. In each of the 3 rounds, it was estimated that a subject would have about 8 interruptions for a total of 24 interruptions for the duration of the experiment. Subjects were seated at a computer with an interface that listed 27 tasks (See Figure 3). To select a task to do, the subjects clicked on the desired task from the task list, which opens the task in a new window. Subjects received interruptions on the side (pop-up) or within the task list itself (highlight). Clicking on pop-ups opened the task in a new window. Clicking on a highlighted task opened it in a new window exactly like clicking a regular task from the list. In the case where both a highlight and pop-up were present, subjects could choose either method to open the task.

As new interruptions were delivered, they were presented in conjunction to the older unresolved ones. The interruption types were as follows: Highlight: Urgent tasks were highlighted on the task list. Pop-up: A popup message notified the user that an urgent task should be done. Multiple pop-ups were possible. Both: Both forms of notification.

All notifications occurred at task boundaries (as concluded from a pilot study). After each round, subjects were asked to rate the effectiveness of the notification method.

Participants

For this experiment, 20 subjects (14 male, 6 female) participated in the experiment in two cohorts of 10 subjects each. They were recruited from the university's online experiment listings website. Subjects from an earlier pilot study were excluded from this study. Subjects were compensated $15 for the hour.

Measures

Dependent measures included the response times to urgent tasks and subject perception of effectiveness on a 10-point scale (10 being highest).

Results

A one-way ANOVA on condition by Effectiveness ratings was significant, $F(2,51) = 5.82$, $p = .005$. A post-hoc student's t-test showed no significant differences between the Highlight condition (M= 8.78) and the Highlight+Pop-up condition (M= 7.56), but revealed a significant difference between those two conditions and the Pop-up condition (M= 5.78). Hypothesis 1a was confirmed. Hypothesis 1b was rejected.

A one-way ANOVA on Response Time by conditions revealed significant differences, $F(2,333)= 13.53$, $p < 0.0001$ (Figure 4). A post-hoc student's t-test suggested no difference between the Highlight condition (M= 24.32) and the Highlight+Pop-up condition (M= 9.38), but found a significant difference between those two conditions and the Popup condition (M=68.52). Participants in the Highlighting condition responded to the interruption significantly faster than the Pop-up condition, so Hypothesis 2a was not confirmed. The Highlight+Pop-up condition elicited the fastest reaction times, thus Hypothesis 2b was confirmed. Participants in the Highlight+Pop-up condition responded to the interrupt faster than the Highlight condition, though not significantly, thus Hypothesis 2c was somewhat confirmed.

We also performed a one-way ANOVA on condition by the percent of Completed Prioritized Tasks, which did not prove to be significant, $F(2,51) = 2.42$, $p = .09$. A post-hoc student's t-test again showed no significant differences between the Highlight condition (M= 0.82) and the Highlight+Pop-up condition (M= 0.85), but

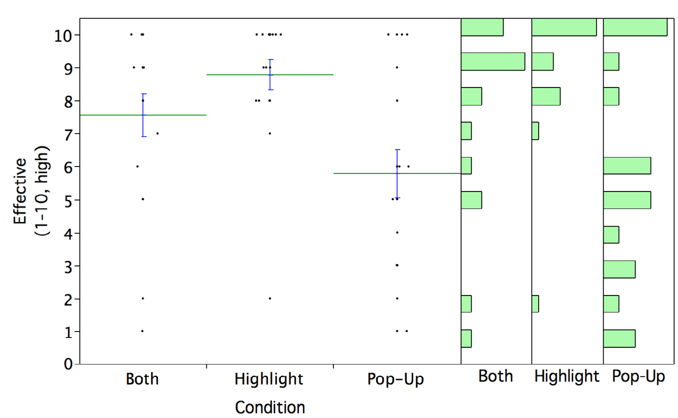

Figure 4: ANOVA: Effectiveness x Condition

found a significant difference between those two conditions and the Popup condition (M=0.71), which suggests a slight trend towards participants with pop-ups completing fewer of the interrupt-driven tasks.

DISCUSSION

This study suggests that users respond more slowly to pop-ups, doubt their effectiveness, and might complete fewer prioritized tasks. It was interesting to find that having highlighting along with pop-ups did not perform worse than pop-ups. While we do not know why this occurred, there may be various reasons. Perhaps redundancy was helpful as opposed to detracting from the experience and performance, which we expected. It might also be the case that participants could ignore pop-ups in favor of the highlighting because they could rely on the latter rather than depending solely on the pop-ups. In terms of user ratings, higher rating of dual-signal than to the pop-ups may be due to a feeling of choice in regards to which interruptions to attend to. More research is needed in order to discern the exact reasons behind these findings.

In this study, the interruptions did not provide new content or bring old important content back into focus—instead, they only alerted the user to urgent tasks. Because no new necessary information was given, the nature of our interruptions might differ from previous work that focuses on the use of content in interruptions. In order to provide interruptions with content, some proposed ideas for future tasks include document editing, information seeking, and summarizing of information. It might also be worthwhile to compare content-laden and content-less interruptions for the same kinds of tasks to determine whether one is better than the other.

CONCLUSION

In this paper, we presented a study that brings attention to the manner in which interruptions are delivered. More research is needed to expand the body of research on this topic. Intelligent assistants are being developed at increasing rates and designers need to be able to identify the costs and benefits of interruption types. In this study, we also examined a novel way of introducing interruptions by highlighting tasks from a to-do list.

ACKNOWLEDGEMENTS

We thank the Attention Management team, as well as the entire Radar project team. This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. NBCHD030010 and FA8750-07-D-0185.

REFERENCES

1. Adamczyk, P.D. & Bailey, B.P. If Not Now, When?: The Effects of Interruption at Different Moments Within Task Execution. *CHI 2004 Proceedings.* ACM Press (2004), 271-278.
2. Adamczyk, P.D., Iqbal, S.T., & Bailey, B.P. A Method, System, and Tools for Intelligent Interruption Management. *TAMODIA 2005*, (2005), ACM Press, 123-126.
3. Bailey, BP, Konstan, JA, Carlis, JV. The Effects of Interruption on Task Performance, Annoyance, and Anxiety in User Interface.*Interact*, 2001, 593-601.
4. Bailey, B.P. & Konstan, J.A. On the Need for Attention-aware Systems: Measuring Effects of Interruption on Task Performance, Error Rate, and Affective State. Computers in Human Behavior, 22, (2006), 685-708.
5. Bilsus, D., Hilbert, D.M., & Maynes-Aminzade, D. Improving Proactive Information Systems. *Proceedings of the Conference on Intellent User Interfaces*. ACM Press (2005).
6. Czerwinski, M, Cutrell, & E, Horvitz, E. Instant Messaging: Effects of Relevance and Timing. ftp://ftp.research.microsoft.com/pub/ejh/hci2000.pdf.
7. Edwards, S.M., Li, H, & Lee, J. Forced Exposure and Psychological Reactance: Antecedents and Consequences of the Perceived Intrusiveness of Pop-up Ads. Journal of Advertising, (2002), 83-95.
8. Freed, M., Carbonell, J., Gordon, G., Myers, B., Siewiorek, D., Smith, S., Steinfeld, A., & Tomasic, A. (2008). RADAR: A personal assistant that learns to reduce email overload. *AAAI Integrated Intelligence*.
9. Gluck, J., Bunt, A., & McGrenere, J. Matching Attentional Draw with Utility in Interruption. *CHI 2007 Proceedings*. ACM Press (2007), 41-50.
10. Horvitz, E & Apacible, J. Learning and Reasoning about Interruption. *Proceedings of ICMI,* 2003, 20-27.
11. Hudson, J.M., Christensen, J., Kellog, W.A., Erickson, T. "I'd be Overwhelmed, but It's Just One More Thing to Do:" Availability and Interruption in Research Management. *CHI 2002 Proceedings*, ACM Press (2002), 97-104.
12. Hudson, S, Fogarty, J, Atkeson, C, Avrahami, D, Forlizzi, J, Kieler, S, Lee, J, Yang, J. Predicting Intteruptibility with Sensors: A Wizard of Oz Feasibility Study. *CHI Proceedings 2003*, 257-264.
13. Hui, B & Boutilier, C. Who's Asking for help? A Bayesian Approach to Intelligent Assistance.
14. Mark, G, Gonzalez, VM, Harris, J. No Task Left Behind? Examining the Nature of Fragmented Work. *CHI 2005 Proceedings*, ACM Press 2005, 321-330.
15. McCrickard, D.S., Czerwinski, M, & Bartram, D. Introduction: Design and Evaluation of Notification User Interfaces. *International Journal of Human-Coputer Studies* (2003), 509-514.
16. McCrickard, D.S., Chewar, C.M., Somervell, J.P., Ndiwalana, A. A Model for Notification Systems Evaluation—Assessing Usr Goals for Multitasking Activity. ACM Transactions on Computer-Human Interation, (2003), 312-338.
17. Miyata, Y & Norman, D.A. Psychological Issues in Support of Multiple Activities. In: Norman, D.A., Draper, S.W. (Eds.), *User Centered System Design: New Perspectives on Human–Computer Interaction* (1986). Lawrence Erlbaum Associates, Hillsdale, NJ.
18. Nirenburg, S & Lesser, V. Providing Intelligent Assistance in Distributed Office Environments. *ACM SIGOIS Bulletin*, Vol 7, (1986), 104-112.
19. Oulasvirta, A & Alovaara, A. A Cognitive Meta-Analysis of Design Approaches to Interruptions in Intelligent environments. *CHI '04 extended abstracts on Human factors in computing systems*, (2004), 1155-1158.