

Effects of Instant Messaging Interruptions on Computing Tasks

Edward B. Cutrell, Mary Czerwinski, and Eric Horvitz
 Microsoft Research
 One Microsoft Way, Redmond, WA 98052
 a-edcut@microsoft.com

Abstract

This paper describes a study that probes the cost of interrupting users with instant messages during different phases of a computing task. We found that interrupting users during the “evaluation phase” of the task resulted in significantly longer completion times than interruptions in other phases. We also found that interruptions that were irrelevant to the task resulted in longer times to process the message and longer task resumption times than relevant messages. These initial results have implications for the principled design of intelligent interrupters and instant messages.

Introduction

People using computers typically cope with a barrage of interruptions from email, the telephone, coworkers dropping by, and, increasingly, instant messages. The growing sophistication of user modeling suggests that it may be possible to mediate these interruptions based on assessments of users’ activity and the content of the interruption [3,4]. We are interested in how the content and context of interruptions influence the cost of notifications. We describe an initial study that explores the effect of instant messaging interruptions on realistic computing tasks.

There is a long history in the attention literature examining the allocation of attentional resources to competing events or tasks (e.g., [1]). While this literature certainly aids our understanding of how attention is allocated, it tells us less about how to design and guide notifications or messages during typical computing tasks. The results of the few applied experiments suggest that interruptions are particularly costly when the interruption is similar to the ongoing task or when it is particularly complex [2].

We set out to determine if different stages of a computer task were more or less amenable to interruptions that were or were not relevant to that task. Our labeling of stages of a web search task was based loosely on notions outlined in Miyata and Norman [5]. If certain task stages are robust to interruptions, intelligent systems may one day be used to govern the timing and nature of those interruptions to optimize user satisfaction and performance.

Methods

Participants: 9 advanced users of Microsoft Office, ages 30-56 years old participated in this study.

© Copyright on this material is held by the author(s).

Procedure: At the beginning of the session, participants were given 4 practice trials to get them used to the procedure. They then performed a series of 24 compound tasks. Each task comprised two parts: a targeted web search task, and a cursory analysis of the graphic design quality of the target site.

Participants began with the search portion of the task, which could be roughly broken into 3 phases. In the first phase, which we called the *planning phase*, participants were given the web search target in the form of a title and brief description of the web site. They were told to review this information and mentally construct three search terms to be used in a Boolean search (a & b & c). Participants were explicitly told to decide on their search terms before leaving this window. Participants then moved to what we termed the *execution phase* in which they entered their search terms in the AltaVista query window in Internet Explorer v. 5.0. Upon receiving the search results, participants entered what we labeled the *evaluation phase*, where they reviewed the results and selected the best match to their target. Participants were told to explore at most two or three pages of search results before selecting a match, and were not allowed to reformulate their query.

After selecting the match, participants inspected the site to evaluate its graphic design quality. They copied the URL and pasted it into a word file divided into three categories based on the probable designer: 1) a student or hobbyist 2) an upscale professional web design firm, or 3) a small company’s IT department. They were then given a brief pause before beginning the next trial.

While participants were performing the search task, they were sometimes interrupted with an instant message notification (including an audio alert) using MSN’s Instant Messenger Service v. 1.0 in the task bar. When this occurred, they were required to open the instant message and reply, “OK,” before going back to the search task. These interruptions occurred in one of the three phases of the search task (planning, execution, evaluation) or did not occur at all (no interruption). Half the time, the messages were “relevant” in that they told subjects which design category to place the site they found. Other times, the messages were “irrelevant”—these simply conveyed some factoid about the site they were searching for. All conditions were counter-balanced.

Results

Of the seven timing measures used in this study, three dealt exclusively with interruption trials: time to switch from the search task to message, total time spent on

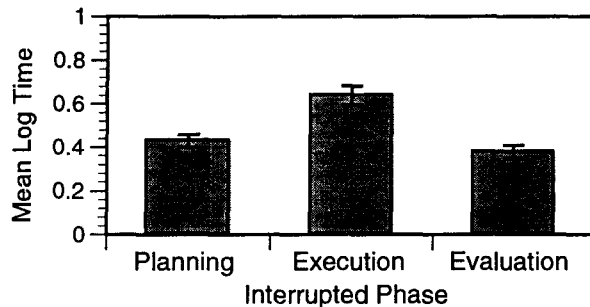


Figure 1. Mean log time to switch from task to notification for each interrupted phase (\pm SEM).

message before returning to the task, and time to resume the search task after leaving the message. Analyses of these measures omitted trials without interruptions. The other four measures included time spent in each task phase (planning, execution, and evaluation) and total time on task. These times were adjusted to subtract out time actually spent attending to the message. All timing measures were converted to log (time in seconds) before analysis to normalize the common skewing and variability associated with response time data.

Multivariate analyses of variance (MANOVA) revealed four significant results. First, for the time to switch to the message from the task, a significant main effect was found for interrupted phase, $F(2,149)=14.93$, $p<0.001$. Post hoc analyses showed that the time to switch to the message was significantly slower when the notification arrived during the execution phase than either other phase (see Fig. 1). Findings for the total time spent on messages and time to resume the primary search task were complementary: a significant main effect was found for relevance, $F(1,149)=32.34$, $p<0.001$, and $F(1,149)=4.64$, $p<0.033$, respectively. The total time spent on messages and time to resume the search task were both longer when the message was irrelevant than when it was relevant (see Fig. 2). Finally, for overall time spent in the evaluation phase, there was a significant main effect for the interrupted phase, $F(3,197)=3.90$, $p<0.01$. Post hoc tests showed that when the notification occurred during the evaluation phase, participants were slower on this phase than any other (mean log (time) = 1.6 vs. 1.4 seconds for the other two phases).

Discussion

The results of this study can be broadly split into two classes: those relating to an interruption itself and to the ongoing web search task. The finding that it took longer to switch to the message during our execution phase is reminiscent of the idea of *chunking behaviors*. This refers to the tendency to delay switching to another task until completion of a subtask (e.g., typing the search words). We informally observed users not attending to an interruption until they had completed typing their search keywords. The significant effect of relevance on message reading and task resumption times suggests that when interruptions are unrelated to ongoing tasks, they will take

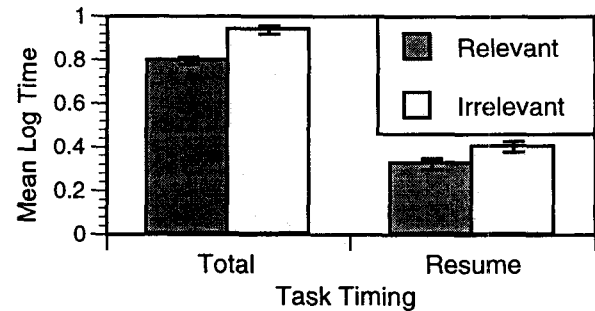


Figure 2. Mean log time attending to notification and resuming search task for each notification type (\pm SEM).

longer to process and it will be more difficult to get back on task following the interruption.

While we saw no cost for interruptions during the planning stage, there was a reliable effect for our evaluation stage. This result may reflect the time required for users to visually re-orient themselves to where they left off, and the concomitant re-scanning of the web search results after the interruption, or due to the user forgetting why/if a particular result was a candidate target. Further research will be needed to tease these possibilities apart.

These initial results suggest first that interruptions that are relevant to ongoing tasks are less disruptive than those that are irrelevant, so determining message content prior to notifying the user may be useful. Second, the results suggest that notifications delivered while a user is typing or potentially rendering other kinds of "chunked" information may be problematic, and should be avoided. Finally, the results for interruptions in the evaluation phase of our task were intriguing. Further research will explore just which parts of a task suffer most from the interruption, and how innovative policies and designs for notifications might mitigate these effects.

References

- Bryan, W. and Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits, *Psych Rev*, 6, 345-75.
- Gillie, T. and Broadbent, D. (1989). What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychol Res*, 50, 243-50.
- Horvitz, E., Jacobs, A. and Hovel, D. (1999). Attention-sensitive alerting. *15th Conf. on Uncertainty and AI (UAI '99)*, Morgan Kaufmann Publishers, Stockholm, Sweden, pp. 305-13.
- McFarlane, D. (1999). Coordinating the interruption of people in human-computer interaction. , *Human-Computer Interaction - INTERACT'99*, IOS Press, Inc., The Netherlands, pp. 295-303.
- Miyata, Y. and Norman, D.A. (1986). Psychological issues in support of multiple activities. In D.A. Norman and S.W. Draper (Eds.), *User Centered Design: New Perspectives on Human Computer Interaction*. Lawrence Erlbaum, pp. 265-84.