Instant Messaging: Effects of Relevance and Timing

Mary Czerwinski, Edward Cutrell and Eric Horvitz

Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
marycz@microsoft.com

ABSTRACT

Instant messaging (IM) has grown rapidly to involve millions of users spanning a variety of platforms. This paper outlines two preliminary studies that examined the effects of IM notifications on ongoing desktop computer productivity tasks. Results from the studies show that the disruptiveness of IM to productivity tasks is reduced if the incoming message is highly relevant to the current task, or if messages are queued until certain key computing operations have been completed. User interface design principles for the control of messaging are proposed based on the results.

KEYWORDS: Instant messages, attention, task switching, empirical studies

INTRODUCTION

Instant messaging systems such as America Online's Instant Messenger, Microsoft Network's Messenger, or Yahoo!'s Messenger service have over 70 million users and research shows that the number of users is growing rapidly (MediaMetrix, December, 1999). The benefits of instant messaging (IM) are numerous, including the ability to know when personal contacts are available, nearly instantaneous communication, and the ability to carry on several informal conversations at once. However, the effects of the growing number of incoming instant messages, combined with email, news alerts and other updates, on ongoing computing tasks have been relatively unexplored. Accompanied by audio alerts, or visual indicators, just how disruptive are these messages to normal computing tasks? How does the relevance of the content influence disruption? Does the kind of computing task the user is performing make a difference in a user's ability to disengage and attend to the notification? Does the disruption associated with a notification depend on the point in the primary task at which it occurs? These are the issues examined in the current research.

McFarlane (1999) examined four methods for deciding when to interrupt someone during multitasked computing. He explored *immediate* (requiring immediate user response), *negotiated* (user chooses when to attend), *mediated* (an intelligent agent might determine when best to interrupt) and *scheduled* (interruptions come at prearranged time intervals) interruption methods. None of these methods was the single best way to interrupt users in tasks across all performance measures. McFarlane concluded that giving people control to negotiate for the onset of interruptions resulted in good performance; however, he cautions that users may postpone attending to interrupting messages in these cases. Also, if forced to acknowledge an interruption immediately, users in his study got the interrupting task done promptly but were less efficient overall.

Gillie and Broadbent (1989) presented a series of experiments aimed at elucidating features of interruptions that make them more or less disruptive to an ongoing computer task. They manipulated interruption length, similarity to the ongoing task, and the complexity of the interruption. They showed that simply being able to rehearse one's position in the main task does not protect one from the disruptive effects of an interruption. In addition, they discovered that interruptions with similar content could be quite disruptive even if they are extremely short (replicating findings in earlier work by Kreifeldt and McCarthy, 1981). Finally, if the interrupter imposed high memory load or processing demands on the user, it was harmful to the primary task. Hess and Detweiler (1994) showed that interruptions that were similar to an ongoing computer task were quite disruptive over the first two of three sessions, but were significantly less disruptive by the third session. In addition, they found that, if participants were allowed to train on the primary task without interruptions for two sessions, presenting a third session of interruptions was significantly harmful to performance, despite the task being highly trained. It would appear from these last results that experience handling the interrupting tasks is what reduces their harmful effects over time.

Miyata and Norman (1986) discussed reminders and interruptions as they outlined a set of suggestions about system support for multiple activities. Pulling from the large body of psychological studies in attention, memory and action, they came up with a number of conjectures. In particular, the authors predicted that interruptions after important actions or between task execution and evaluation would be less harmful than interruptions occurring at other times. Work presented in the following two studies comes in the spirit of the writings of Miyata and Norman. We were interested in testing the influence of the phase of the task (planning, execution or evaluation) on the disruptiveness of IM notifications. In addition, we wondered if varying the type of primary computing task had an effect. For instance, would word processing be more amenable to interruptions than use of a drawing program? For both of these two questions, this work is exploratory in nature. Finally, we looked at the relevance of the content of the IM to the ongoing primary task. Based on the research and conjectures reviewed above, we predicted that relevant interruptions would be less disruptive, overall, to primary task performance.

EXPERIMENT 1—EFFECTS OF TASK AND TEMPORAL PHASE

METHODOLOGY

14 experienced Microsoft Office 2000 users, aged 26-55, participated in this study.

DESIGN AND PROCEDURE

Before participants began working, they were given a list of six major stocks and their strike prices. Participants were told that there was a special program on their computers that would notify them when there were significant changes in the prices of these stocks. They were also informed that they would have an opportunity to sell their stock at some point in the day; thus, it was very important for them to pay attention to each notification to monitor the prices of each stock. Participants were instructed to sell the stock only if it increased in value by a point or more.

Participants then performed a set of 18 tasks divided into 6 blocks of 3 tasks each. There were 3 kinds of tasks: 6 drawing tasks using the draw and paint tools in MS Word; 6 spreadsheet manipulation tasks using MS Excel; and 6 text editing tasks using MS Word. Blocks were presented in random order, and the tasks within each block were also randomized (i.e., the order of the three tasks associated with a given block was randomized, as was the order of the block itself). During each task, subjects were notified twice. Each notification occurred in one of three temporal positions: 1) while accessing the file menu at the beginning of the task (opening the file), 2) while accessing tools (e.g., toolbar), or 3) while manipulating the content of the file (e.g., typing text or selecting cells in Microsoft's Excel spreadsheet application).

Notifications were presented in a counterbalanced order on a separate computer screen situated next to the primary working screen. Each notification consisted of an auditory alert, a text box announcing the arrival of stock information, and a button. Clicking on the button replaced the box with another box containing updated information about a stock. A second button in this box dismissed the notification. Timing information was recorded on the time to respond to the notification (by clicking on the first button) and on the time to dismiss the notification (by clicking on the second button).

DESIGN

This study was a 3 (temporal interrupt position: file menu; tools menu; content manipulation) x 3 (task type: drawing; typing or selecting cells in Excel; text editing) x 4 (number of replications) repeated measures within subjects design, for a total of 36 notifications per single subject session. Both studies reported in this paper were run on a Dell Precision Workstation 610 MT with a dual P450 processor and 256 MB RAM. The computing system also included a Sony 15" flat panel monitor, a Microsoft serial mouse and keyboard.

RESULTS

A 3 (interrupt position) x 3 (task type) x 4 (replications) repeated measures within subjects ANOVA was performed for both time to respond to notifications and dwell time (time spent viewing notification). No significant effect was found for any factor for dwell time. For time to respond to the notification, however, significant main effects were found for both interrupt position ($\underline{F}(2,26)=3.93$, $p=0.032$) and task type ($\underline{F}(2,26)=3.70$, $p=0.038$) with no significant effect for repetition. There were no significant interactions. See Figure 1 for an illustration of the mean times to view the notification for interrupt position and task type.

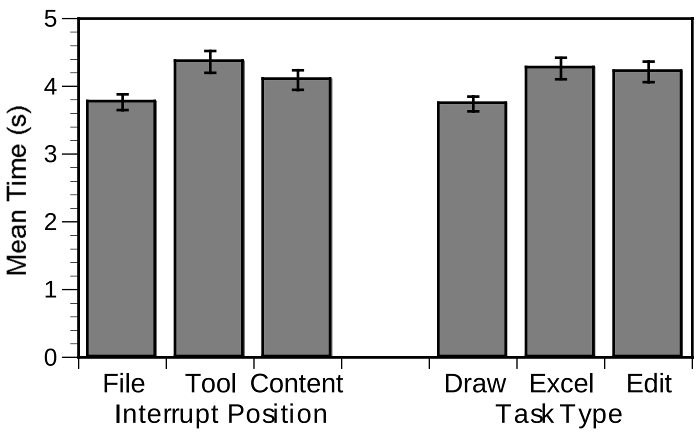

Figure 1. Mean time to view IM for each interrupt position and task type (± SEM).

DISCUSSION

The results of Experiment 1 revealed that when users were interrupted with an IM at the beginning of the task (e.g., when they were in the File menu), they were faster at disengaging from the primary task and switching to the IM than when interrupted at other temporal positions. In addition, when notifications were received during drawing, subjects more easily disengaged from the task at hand to attend to the IM. The Excel and Editing tasks required the user to perform a significant amount of searching, sorting and typing. The Drawing task appeared to be more lightweight in nature in that users simply had to paint and drag and drop objects into their proper spatial locations in order to match a pre-existing drawing. It could have been the very nature of the drawing activities that made them more amenable to task switching.

EXPERIMENT 2—IM AT DIFFERENT PHASES OF WEB TASKS

In a second experiment, we again probed the cost of interrupting users with messages during different phases of a computing task but included more dependent measures. We also set out to determine if different stages of a computer task, classified more broadly than by specific software interactions, 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 the decomposition of problems described in Miyata and Norman (1986).

METHOD

9 advanced users of Microsoft Office 2000, aged 30-56 years old participated in the study.

TASK AND PROCEDURE

Each task comprised two parts: a targeted web search task, and a cursory analysis of the graphic design quality of the target site. We divided the search task into 3 phases. In the first phase, which we call 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 re-view this information and mentally construct three search terms to be used in a Boolean search (a & b & c) before leaving this window. Participants then moved to what we term the *execution phase* in which they typed their search terms into an AltaVista query window in Internet Explorer v. 5.0. Upon receiving the search results, participants entered the portion of the task we refer to as the *evaluation phase*, where they reviewed the results and selected the best match to their target. Participants 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 most likely designer: 1) a student or hobbyist; 2) an upscale professional web design firm; or 3) a small company's IT department. They then began the next trial.

While participants were performing the search task, they were sometimes sent an IM (including an audio alert and a flashing label in the MS Windows Taskbar at the bottom of their screen) using MSN's Messenger Service v. 1.0. When this occurred, they were required to open the IM and reply, "OK," before going back to the search task. Messages occurred in one of the three phases of the search task (planning, execution, evaluation) or did not occur at all (no message). Half of 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 perusing.

DESIGN

The design was a 4 (phase of computing task IM occurred, including no-IM) x 2 (IM was either relevant or not) x 3 (replication per condition) within subjects repeated measures. All independent variables were tested within subjects and presentation order of the trials was randomized. There were 24 (3 replications per condition) trials per session, and the session length was approximately 1.5 hours long.

RESULTS

Of the timing measures collected in this study, three dealt exclusively with IM trials: time to switch from the search task to message, total time spent on message before returning to the task, and time to resume the search task after leaving the message. Analyses of these measures omitted trials without messages (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 by subtracting out the 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.

A doubly multivariate repeated measures analysis of variance 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, $F(2,16)=17.23$, $p<0.001$. Post hoc analyses showed that the time to switch to the message was significantly slower when the IM arrived during the execution phase than either other phase (see Fig. 2). 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, $F(1,8)=39.69$, $p<0.001$, and $F(1,8)=11.31$, $p<0.01$, respectively. The total time spent on messages and time to resume the search task were both longer when the IM was irrelevant than when it was relevant (see Fig. 3). Finally, for overall time spent in the evaluation phase, there was a significant main effect for the interrupted phase, $F(3,24)=5.75$, $p<0.004$. Post hoc tests showed that when the IM 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—see Fig. 4).

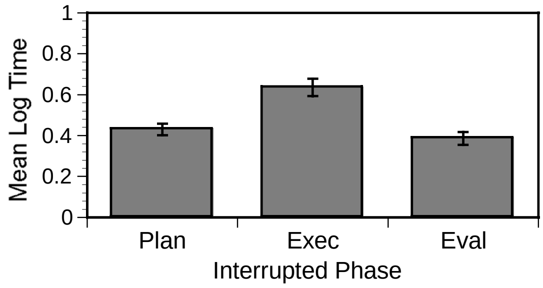

Figure 2. Mean log time to switch from task to notification for each interrupted phase (± SEM).

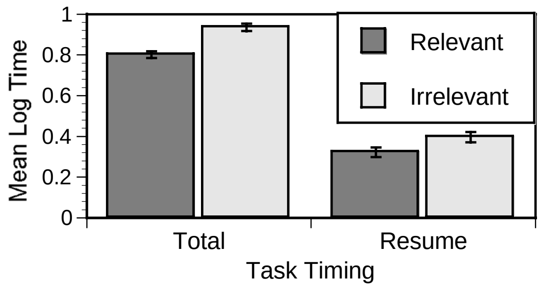

Figure 3. Mean log time attending to IM and resuming search task for each IM type (± SEM).

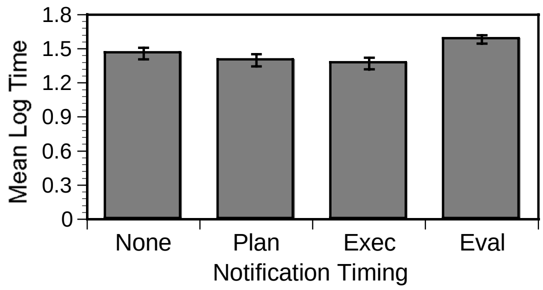

Figure 4. Mean log time to switch from task to notification for each interrupted phase (± SEM).

DISCUSSION

The results of this study can be split broadly into two classes: those relating to switching to the IM itself and those relating to the ongoing web search task. The finding that it took longer to switch from the primary search task to the IM during the execution phase is reminiscent of the notion of *chunking behaviors*. In prior work, researchers (Sellen, Kurtenback and Buxton, 1990) have noted a tendency to delay switching from one task to another until the completion of a subtask (e.g., finishing the input of search words). We informally observed that users often did not attend to a message until they had completed typing their search keywords. The reliably harmful effect of message irrelevance on both reading and search task resumption times suggests that notifications that are unrelated to ongoing tasks take longer to process and it is more difficult to reestablish task context following the interruption.

CONCLUSION

These initial results provide us with some guidelines for the design of policies for notifications in messaging systems. We have demonstrated that the delays associated with an IM disruption depends on the point in a computing task that a user is presented with it. We found that a good time for notifications is early in the task, before the user has become deeply engaged in the task goal; and that notifications arriving during the evaluation, planning, and execution phases were harmful, supporting Miyata and Norman (1986). Moreover, sending an IM while users are typing or interacting with toolbars is significantly more disruptive than during other times. These effects hold not only for the time it takes to switch to the notification (which could arise from the overhead of switching devices, e.g., from keyboard to mouse), but also for the overall task time. In other words, the nature of the task phase when an IM is presented affects not only how quickly a user can disengage from a primary task but also the time to reestablish the context of the previous task. Second, Experiment 2 revealed that IMs that are relevant to ongoing tasks are less disruptive than those that are irrelevant. Finally, the reliable effect of IMs harming the evaluation phase of tasks explored in our studies is intriguing. This result may reflect the time required for users to visually re-orient themselves to where they left off in the search results list, and the concomitant re-scanning of the web search results after the message, or be associated with latencies in users' memory as they attempt to recall why or if a particular result was a candidate target. We are pursuing studies to explore these alternative hypotheses, as well as to continuing to examine ways to remind the user of the primary task after an IM has occurred.

REFERENCES

Broadbent, Donald, & Gillie, Tony (1989) What makes interruptions disruptive? A study of length, similarity and complexity. *Psychological Research*, **50**, 243-250.

Hess, Stephen M. & Detweiler, Mark C. (1994) Training to reduce the disruptive effects of interruptions. In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, 1173-1177.

Kreifeldt, John. G. & McCarthy, Mary E. (1981). Interruption as a test of the user-computer interface. In: *Proceedings of the 17th Annual Conference on Manual Control*, 655-667. Publication **81-95**, Jet Propulsion Laboratory, California Institute of Technology.

McFarland, D. (1999). Coordinating the interruptions of people in human-computer interaction. In: *Human-Computer Interaction—Interact '99*, Sasse, M.A. & Johnson, C. Eds., IOS Press, Inc., IFIP TC 13, 295-303.

Media Metrix (1999). Report on digital media. Dec, 1999. http://www.mediametrix.com.

Miyata, Yoshiro, & Norman, Donald (1986). Psychological issues in support of multiple activities, Chapter in User-Centered Design, Norman, D.A., & Draper, S. W., Ed., Lawrence Erlbaum, NJ, 265-284.

Sellen, A. J., Kurtenbach, G. P., & Buxton, W. A. S. (1990). The role of visual and kinesthetic feedback in the prevention of mode errors. Paper presented at INTERACT '90.