

Empirical Validation of a Computer-Mediated Coordination of Interruption

Sonja Gievska
Prof John Sibert
The George Washington University

Department of Computer Science
The George Washington University
Washington DC, USA
Email: sonjag,sibert@gwu.edu

Abstract

User's cognitive state is extremely fragile and sensitive to external interruption, especially when a user is performing a cognitively-taxing task. Given that interruptions frequently occur during human-computer interaction, and given the relative scarcity of human attention and their vulnerability to interruptions, interruption-resistance is needed to be considered when designing user interfaces. The main objective of this research has been to explore the possibility of facilitating user's performance in a context of interruption, by employing a context-sensitive computer-mediated coordination of interruption. A novel Interruption Taxonomy, which synthesizes various kinds of interruption-related context information, served as a basis for implementing graceful coordination of interruption. An exploratory user study was conducted to experimentally evaluate the adequacy of the proposed approach for mediating interruptions and the utility of the Interruption Taxonomy.

Keywords

Context-sensitive user interfaces, human interruptions, coordination of human interruptions, interruption taxonomy

INTRODUCTION

Disruptive effects of interruptions have been studied across disciplines, especially in psychology. The results of these experiments have shown that people make more mistakes, hesitate and delay in making decisions, have difficulties remembering, and in general are less effective when exposed to interruptions (Bailey et al. 2001), (Cellier and Eyrolle 1992), (Cutrell et al. 2000), (Gillie and Broadbent 1993), (Trafton et al. 2003). Persistent and repeated interruptions may affect a person's attitude toward the entire situation, very often leading to a non-optimal behaviour (Kirmeyer 1988), (Tulga and Sheridan 1980). Emotional arousals, such as discomfort, stress, subjective feeling of failure, dissatisfaction, and stress may furthermore degrade human performance (Bailey et al. 2001), (Kirmeyer 1988).

As interruptions naturally occur during any communication including human-computer interaction, even a small reduction in their harmful effects can have significant benefits. Computer outputs are very often generated in isolation from the context in which the interaction occurs. Having an option to control whether or not system messages will be permitted to intrude on users is not much of a choice. The quality designers should aim for is a tacit and sensible ways of interrupting people to reduce the disruptive effects of interruptions.

By filtering and mediating the information flow, especially messages that may disturb the user and intrude on the user's current endeavor, user interfaces can do a better job in supporting users in their activities. Employment of theoretically sound methods for coordinating potentially disruptive interruptions can be employed in a number of systems and domains, especially critical in information-intensive and safety-critical tasks. Reliability-seeking systems used in domains such as: critical-care, nuclear plants, emergency, military, and aerospace, could greatly benefit from a user interface that filters messages and coordinates interruptions.

The theory and traditional user-interface design guidelines do not address the interruption problems entirely. Recommendation and empirical results still do not generalize to wide majority of application domains and systems. There is a lack of a general framework to guide interface designers in developing more tacit and graceful interaction that can leverage the strengths and support the weaknesses of humans in presence of interruptions. This research had two related objectives: (1) to develop taxonomy that captures enough context knowledge to support coordination of interruptions during human-computer interaction (HCI), and (2) effectively demonstrate the utility of the conceptual model in terms of its expressiveness and effectiveness.

INTERRUPTION TAXONOMY

As a basis for the framework a new Interruption Taxonomy is outlined to categorize a variety of traceable information needed to exhaustively describe the context space. The choice to include a category was based on whether the previous research had pointed to the importance of such a choice and the relevance to the goal of this research. The selection was

also driven by the need to identify and represent the common themes, and to interrelate a variety of theories and experimental results that have been published. Interruption-related information is categorized according to context: Task Context, User Context, and Environment Context. A graphical representation of the taxonomy three-dimensional space is presented in Figure 1.

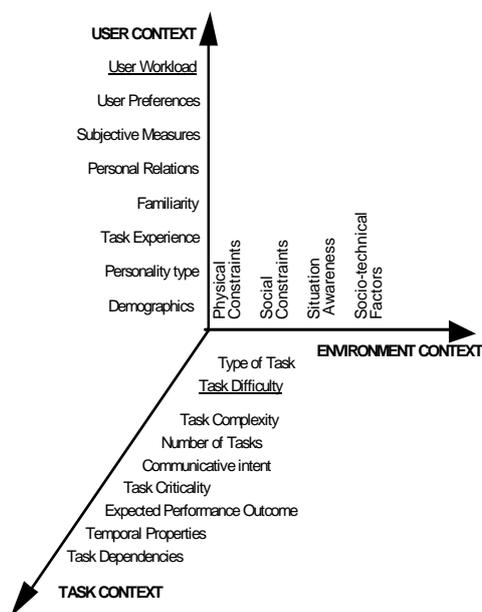


Figure 1. Interruption Taxonomy.

Intrinsic characteristics of a task make some interruptions more destructive to task performance than others; therefore a considerable amount of attention has been devoted to task characteristics and their interdependencies. Task-related context knowledge is crucial in disambiguating the meaning and the relevance of the interruption task in regard to the current user endeavour. User-related taxonomic categories are included to support representation of and reasoning about a particular situation as the user views it. Finding solutions to the problems associated with interruptions that are effective with respect to some objective criteria (i.e., task characteristics) is necessary but not sufficient. The inclusion of Environment Context dimension attributes substantially greater sensitivity to the system, namely, the ability to adapt to a social setting, physical and organizational constraints, or the particularity of a given situation.

Augmenting user interfaces with a variety of contextual information, and appropriate reasoning machinery can help a system decide when and in what manner to interrupt user's ongoing endeavour. Inferential techniques must be applied to provide context-sensitive actions that are responsive to a particular state of interaction. We have adopted an approach based on Bayesian Belief Networks (BBNs) to represent the causal relationship between different pieces of information and to integrate rules for how to use, maintain, and reason with interruption-related knowledge. BBNs have proved suitable for making predictions and decisions from incomplete and uncertain data (Horvitz et al. 1999), (Jameson et al. 1999). The Bayesian network constructed for selecting the timing of the interruption is shown in Figure 2. Most of the nodes in the network drawn as oval boxes correspond to the taxonomy categories. The decision on the most appropriate timing of interruption (i.e., Interruption Timing) depends on inferring the state of several important variables: Interruption Relevance (A), Sensitivity to Interruption (B), Individual Differences (C), Environmental Conditions (D) and Urgency of interruption (E). The circled areas in Figure 2 represent the parts of the graph that relate each of these variables with the relevant taxonomic categories.

It is our hope that by beginning with the identification and categorization of interruption-relevant context information, this research has in some small way explicated the complex and multidimensional aspects of human interruptions and has attempted to present a way to capture it. The taxonomy could be used as an improved basis for designing graceful human-interaction that can leverage the strengths and support the weaknesses of humans in presence of interruptions.

EXPLORATORY USER STUDY

To demonstrate the scope and the utility of the Interruption Taxonomy, an exploratory user study was conducted. Our final concern was to substantiate enough empirical evidence to show that the Interruption Taxonomy could be used as a basis for coordinating interruptions that will increase user performance in multitasking environments. One of the objectives was to demonstrate the expressiveness of the Taxonomy by applying it in a particular interruption environment. In addition, we examined the effectiveness of mediating interruptions compared to condition with no interruption control (i.e., interruptions were presented immediately at random-generated points).

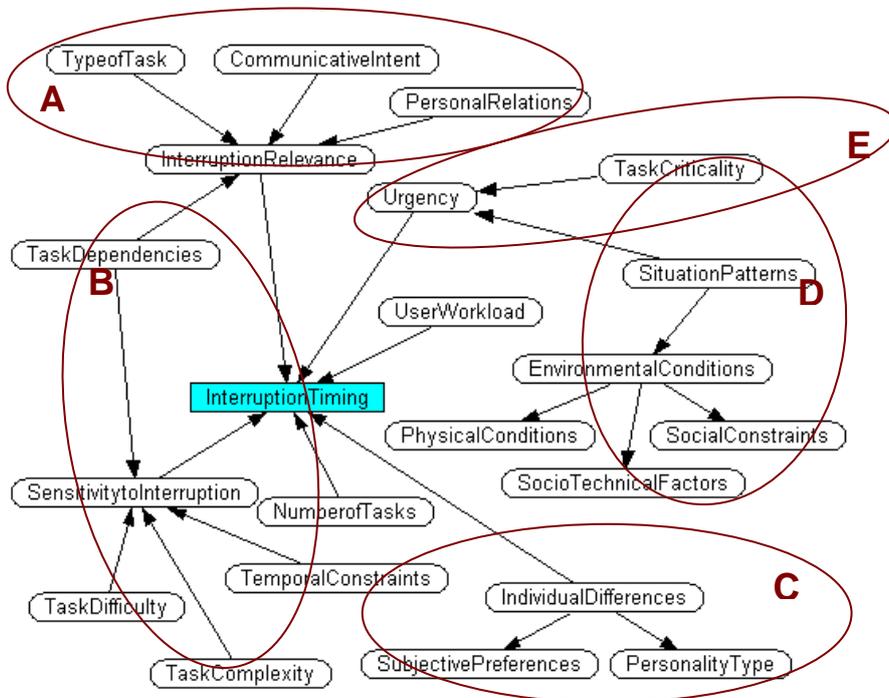


Figure 2. High-Level Dependencies Between Taxonomic Factors for Inferring the Timing of Interruption

Experiment

The general research question of the experiment was “Are there any differences in user performance metrics when mediated coordination of interruption is employed?” Stated as a null hypothesis, we rephrase the question as:

H.1: Mediated coordination of interruption would have no affect on user performance in face of interruption.

The main evaluation objective was to provide empirical evidence of the hypothesized performance advantage of mediating interruption, and get feedback on human acceptance and comfort with the computer-mediated human interruptions. In addition to the need to establish whether mediated interruption would have an effect upon task performance, another factor, frequency of interruption, was thought to be potentially important to the success with which tasks could be performed in the face of interruptions. To measure effectiveness, a variety of performance metrics were used. Domains selected for measurement were those reported to be affected by interruption (e.g., correctness, timeliness, completeness, disturbance). Primary task resumption time defined as time from completing the interrupting task (restoring the TS window) to resuming the primary task was chosen to measure the impact of interruption on user’s performance on the primary task. User accuracy (i.e., percentage of correct object classifications) and reaction time on the interrupting task were analyzed to examine the effect of interruption on the interrupting task performance. Several types of errors were also defined for a more exhaustive measure of correctness when performing the interrupting task. Data concerning user’s qualitative assessment of computer-mediated coordination of interruption was also gathered and analyzed.

We would like to note that our exploratory study was also used as a first step in refining the interruption model and putting the emphasis on the task-related taxonomy categories. For this particular experiment the timing of the interruption was mapped and coded to a certain interruption point within subtask (e.g., before or after a particular interaction event). The system keeps records of the times and contexts of all relevant interaction events. Each time the specified situation arises (i.e., sequence of interaction events), the system infers the user’s goals based on the context of her interaction before and at the time of the interruption. The system adjusts in such a way that a user is not interrupted during interruption points sensitive to interruptions, deferring the interruption task for the next appropriate moment.

Participants and Apparatus

A total of 24 volunteers, 12 females and 12 males, were run as participants in this study and were not compensated for their effort. The participants were a mix of students and professionals, between ages of 16 and 41. A background questionnaire has been administered to each participant at the beginning of the experiment. They were asked to rate their proficiency and experiences in areas such as computer use, planning, multitasking, video games, etc.

All practice and experimental trials were run on Macintosh iMac, running Mac OS 9.0.4. The built-in VGA monitor with 1024x768 pixel resolution was used as the display. The interruption mediator was implemented using Lisp 4.4.3.

Experimental Design

The independent variables in the experiment are Coordination Method and Frequency of Interruption, each set at two levels in a full factorial design, with standard ANOVA used for the analyses. Coordination methods were immediate and mediated, and frequency was set to nine and twelve interruptions per 15-minute session. Within-subject difference in task performance between the condition with immediate, and computer-mediated coordination of interruption under different frequency of interruption was analyzed. Data analyses are conducted within-subject and within-context to compensate for the variability caused by differences between participants or other contextual factors. The number and the type of situations in which an interruption occurs were experimentally controlled (within-context), while still preserving the frequency of interruptions randomly distributed across condition. Interruptions were generated according to a certain scheme, which has been designed in such a way that the targeted situations and interruption points could be ascertained. The experiment employed a balanced Latin square design to counterbalance the order effect. Male and female users were randomly assigned to one of the four order groups.

Dual-task Scenario

As a test-bed application, we used a two-task experimental system developed at the US Naval Research Laboratory (NRL) in Washington D.C that has already been used for interruption-related studies (Trafton et al. 2003). Two experimental tasks bear high resemblance to military-like computer games and simulations (Figure 3). There also appear to be motivational attraction of using game-like tasks; users know right away when they are making progress and get satisfaction from this success at a challenging task. The experimental dual-task could also be seen as simplified scenario of people performing resource-allocation (planning) task, while responding to interrupting task generated as random points. The underlying assumption was that once interrupted, the primary TS task is resumable after the point of interruption.

The primary (interrupted) task is a resource-allocation task named Three-Strike (TS). The objective is to attack and destroy three destinations utilizing available resources, ten heavy and ten light tanks and a certain amount of fuel and munitions. On their missions, users are encountering resistance from differing locales and different kinds of obstacles based on a stochastic model of the TS task. User's interactions are limited to mouse point-and-click events with several dialog-box style windows.

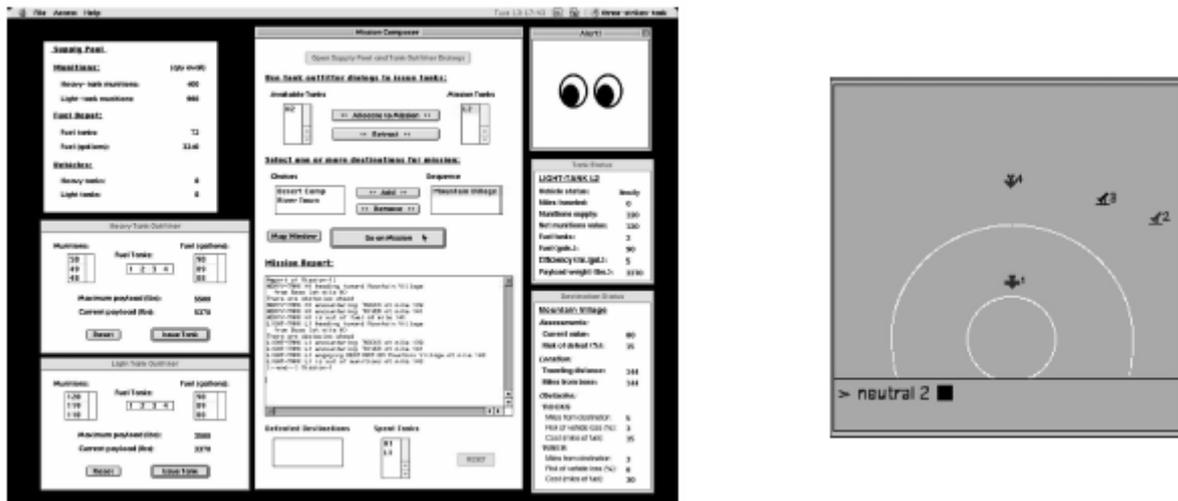


Figure 3. The Interfaces of the Primary and the Interrupting Task.

The interrupting task is Tactical Assessment (TA) task presented to a user at random points while she is performing the primary task. In this task, the user plays the role of a fighter aircraft pilot looking at a radar-screen-like display where three types of objects appear. The objective is to indicate whether the approaching object is hostile or neutral. The decision of the "pilot" is assisted by an intelligent-automated component that colors the objects as red (hostile), blue (neutral) or yellow (when the assessment can not be made). The user is to confirm the hostile/neutral indications or give the appropriate classifications of the yellow objects based on a set of rules. User's interactions consisted of two keystrokes per object using the right-hand numeric key pad. The first keystroke was the object classification (5-neutral, 6-hostile) and the second was the track number that appears next to each object.

At times of interruptions, the primary task is paused, the computer screen clears, and the window of the interrupting task appears on a black background totally obscuring the primary task windows. The interrupting task is considered "finished" when the last flying object is classified, or the last object disappears when the timeout expired. Then, the control switches

back to the primary task. At that point, participant's attention returns to the interrupted task; the primary task is restored to the state at the time of interruption. Participants were instructed to attempt performing both experimental tasks effectively. Participants were allowed to refer to the written instructions throughout the experiment.

Treatments

A short description of the four experimental treatment conditions follows:

Treatment 1: "Immediate 9" treatment condition is an implementation of immediate coordination of interruption, which means that there was no control of interruption. The interrupting task was generated at nine random interruption points, and presented to a user immediately after generation.

Treatment 2: "Mediated 9" treatment condition is an implementation of the mediated coordination of interruption based on the approach that was advanced in this research. The primary task was interrupted nine times at most appropriate interruption points as selected by the interruption mediator. The system adjusts in such a way that a user is not interrupted during interruption points sensitive to interruptions, deferring the interruption task for the next appropriate moment. An assumption was made that it is unlikely to matter if an interruption is deferred until an appropriate interruption point

Treatment 3: "Immediate 12" treatment condition is similar to treatment 1 with twelve interruptions presented immediately to the user at random generated points.

Treatment 4: "Mediated 12" treatment condition employs the same coordination mechanism as described for treatment 2. The primary task was interrupted twelve times at most appropriate interruption points as selected by the interruption mediator.

Procedure

Twenty four participants were run through the experiment. Each participant was instructed and trained on the primary and the interruption task, independently, followed by a 15 min. practice trial in the context of interruptions (dual-task scenario). The practice session lasted for approximately 1 hour including the time for completing the background questionnaire. After short break, the participants concluded four 15-minute trials of the TS task interrupted by a number of times, either 8 or 12, by the interruption task. Subjective exit questionnaires were distributed at the end of the experimental trials to measure the subjective level of stress, motivation, distraction and anxiety experienced by a user during all four treatment condition. At the conclusion of the experiment the participants were debriefed. Each participant spent approximately 2 hours 30 minutes.

Pilot Study

A pilot study with fifteen was conducted to help construct the interruption model for this particular experimental environment, and implement the research version of the interruption mediator. The main objective of the pilot study was three-fold: (1) to select and investigate the relevance of taxonomic factors hypothesized to influence user performance in this particular experimental environment; (2) to construct the BBN-based interruption model; and (3) to gather data for the initialization and training of the constructed BBN, which was used as a basis for implementing the interruption mediator. It should be noted that the pilot study and the formal experiment were conducted with different set of participants.

The first phase in implementing the interruption mediator was to create an appropriate context representation as a basis for mediating interruptions. The construction of the interruption model was tightly-coupled with the Interruption Taxonomy. In order to encourage greater clarity, it was decided to focus on the Task Context dimension. Interruptions can arrive in a number of different scenarios, and it is of crucial importance to know the exact task context in which it occurs so that the mediator can select the most appropriate timing to interrupt the user. All task-context categories were hypothesized to be relevant for selecting the most appropriate timing of interruption in the experimental, fairly complex, task environment.

To describe the context of interaction that can help the mediator determine the sensitivity of particular interruption point and select the most appropriate timing, description of the tasks on a highest level (e.g., editing a document in MS Word, performing the TS task) though important is not very helpful. Therefore, in-depth task description that spans across many levels of abstractions was performed for the primary task to describe different situations in which an interruption may occur. Most tasks, except on the lowest elementary level, were broken down and organized into subtasks, each one with a specific structure of activities, and with a different sensitivity to interruption.

Domain-specific knowledge was used as a basis for coupling and chunking the TS task structure, while subjective and objective measures gathered during the pilot study were used to validate the proposed organization. First, the structure of the primary task was divided into a set of higher-level subtasks based on GOMS-like analysis [Card et al. 1983]. Goal-based categorization has enabled the selection, and categorization of corresponding task-specific taxonomic categories, in particular task type, task complexity, and task dependencies. Five subtasks have been identified as potential cases for mediating interruption: "*outfitting tanks*", "*allocating tanks to mission*", "*planning a mission*", "*carry out a mission*",

and “review the mission report”. A number of corresponding lower-level activities and interaction events were associated with each higher-level subtask. For example, the “planning” subtask consists of three activities, reviewing the map, reviewing the destination status and selecting a destination. They were clustered as one subtask based on the following interaction pattern exhibited by most of the pilot study participants: repetition of these activities before making each destination choice. The categorization of the subtasks was done in the following manner: higher-level subtasks were categorized in terms of task stages, such as planning, evaluation, execution as proposed in (Miyata and Norman 1986). Lower-level subtasks were actions associated with each task stage (e.g., review, make selection, etc.), while interaction events were associated with corresponding types of computer tasks (e.g., scroll, select, type, read, etc.).

Task complexity was associated with every subtask at each abstraction level using the specifics of the experimental setting. Subjective ratings of the cognitive load collected from the pilot study participants and the observed performances were used to cross-validate the proposed values. Pilot study users were asked to rate the relative difficulty level of each subtask. The cognitive complexity/demand scale ranged from 1 (neglective) to 5 (very high). “Planning a mission” was rated as the most complex 4.5, while “allocating a task to a mission” was rated as a subtask with almost neglective cognitive demand 1.3.

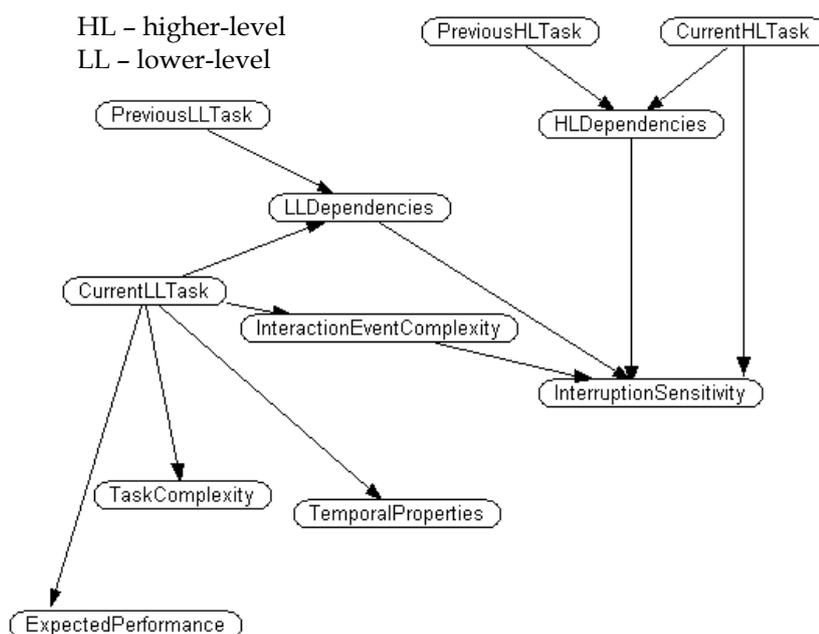


Figure 4. BBN-based Model for Inferring the Interruption Timing in the Dual-Task Experiment

Even though the model of this particular dual-task setting was specified with no dependencies between the interrupted and the interrupting task, it was very important to describe the dependencies between subtasks of the primary task to help in inferring the interruption sensitivity of each particular point. The relationships between higher-level subtasks seemed to only partially describe a particular interruption context. Examination of lower-level relationship was needed, especially for describing transitions between subtasks and cases when a subtask was executed partially or in a few iterative steps. Several functional relationships, such as causal, producer-consumer, cooperation were defined to describe higher levels of associations. Lower-level relationships included categories such as transfer, communication, or shared resources [Malone & Crowston 1994]. Sometimes more than one applied. The strengths of inter-subtask relationships were expressed as corresponding causal probabilities of the nodes of the constructed BBN.

Data available in the form of computer logs, limited personal observations, informal interviews with the participants, and videotapes indicated that sensitivity of particular interruption points can be predicted based on the task-related context representation included in the Interruption Taxonomy. Inspection of the pilot study data have shown that resumption times were sensitive to the taxonomic context categories used to describe the context and the position at which an interruption occurs. Comparison of the resumption times suggested the importance of task complexity and task interdependencies in user’s performance in presence of interruptions.

The selection and categorization of the taxonomic factors was followed by the phase of constructing the BBN-based model for the experimental testbed system. The context factors were interrelated, and built into a decision network that decides when the most appropriate time to interrupt the user is. Figure 4 represents a detailed picture of the portion of the BBN graph constructed for the selecting the interruption timing in our experimental environment. In practice, nodes and parts of the general model presented in Figure 3 became quite complex by adding more exhaustive representation of

the context variables and their relations. Dashed links represents relationships to other variables not shown in this portion of the graph.

Current adaptability of the mediator determines which situations require interruption mediation, and selects the appropriate moment for each one of them. This version of interruption mediator maps the appropriate timing of interruption to a certain interruption point within subtask (e.g., before or after a particular interaction event). The results of all participants were aggregated to distinguish effects based on the characteristics of particular situations (task context), rather than of individual participants. This model made predictions about “an average user” performing in a particular interruption context rather than adapting its behavior to a particular user. This allowed a larger training set than if the implementation was restricted to data from a single user.

The prototype version of the interruption mediator was implemented to explore the design space, and identify the limitations and potential adjustments to the proposed taxonomic interruption model. For experimental purposes, it was decided that it would be clearer to implement few well-distinguished situations than treat them all. The complexity of the model was kept reasonable making it manageable for substantial statistical cross-factor analyses. The first version of the interruption mediator coordinates interruptions in response to situations that can be defined on a basis of the domain-specific knowledge, using only task characteristics. Task-based decisions for mediating interruption (as opposed to basing decisions on individual workload assessments) provided means for more controlled environment.

RESULTS

The data from the formal experiment were analyzed with a 2 (immediate and mediated interruption control) x 2 (8 and 12 number of interruptions during 15 min. session) within-subject Analysis of Variance (ANOVA). We analyzed the within-subject difference in task performance between the condition with immediate and computer-mediated coordination of interruption. An effort was made to control the number and the type of situations in which an interruption occurs keeping the frequency of interruptions randomly distributed across condition.

The results have showed that resumption time for the interrupted task is sensitive to changes in method of coordination of interruption. As hypothesized, significant main effect was found for the method of coordination, $F(1, 23) = 48.707, p < 0.00001$. Significant main effect of interruption coordination indicated that participants resume the interrupted task more slowly under immediate coordination condition than under computer-mediated coordination of interruption.

Performance Measures	F	P	P< α
Resumption time			
Coordination Method	48.707	<0.00001	Yes
Frequency of Interruption	5.391	0.0046	Yes
Interaction	0.75	0.4	No
Inter-click Times			
Coordination Method	1.13	0.3	No
Frequency of Interruption	0.262	0.61	No
Interaction	0.6	0.44	No
Disruption Score			
Coordination Method	40.501	<0.00001	Yes
Frequency of Interruption	6.8	0.016	Yes
Interaction	0.165	0.7	No

Table 1. ANOVA Analyses of the Main Effects and Interactions for the Primary Task Performance Measures

Not surprisingly, interruption frequency showed significant main effect, $F(1, 23) = 9.85, p = 0.0046$, suggesting that change in workload imposed by repeated interruptions affects user performance. Significant main effects of interruption frequency indicated that participants had adjusted to higher frequency by resuming the interrupted task faster. It seemed that the objective of the primary task to defeat all destinations was still compelling, so they trade-off their “comfort” and “thoughtful approach” for speed in these demanding circumstances. The analyses of variance indicated no significance interaction between coordination method and interruption frequency, $F = 0.75, p = 0.3$.

No significant effects of coordination method, $F(1, 23) = 1.13, p = 0.3$, or frequency of interruption, $F(1, 23) = 0.262, p = 0.61$, were found for inter-click times. The within-subject differences in inter-click times were in hundredths and tenths of seconds between different treatment conditions. Inter-click times were not expected to be affected by interruptions, though they are important in recognizing the overall effect of interruption on the primary task performance. The overall effect of interruption is also reflected in the within-subject difference between resumption and inter-click times, defined

as a measure of disruption as suggested in (Trafton et al. 2003). A similar pattern was observed for disruption score as what was concluded for the primary task resumption times; statistically significant main effects were found for coordination method, $F(1,23) = 40.501$, $p < 0.00001$, and frequency of interruption, $F(1,23) = 6.8$, $p = 0.016$. Participants exhibited individual differences when performing the primary task, varying inter-click and resumption times, although the times were consistent for each participant across the experimental treatments.

The results concerning the effects of the coordination method and interruption frequency on the performance of the interruption task appear in Table 2. The time needed for object classification was measured in milliseconds and represents the average speed with which a user classified the approaching objects. Results of analysis of variance indicated that object classification times did not significantly differ by coordination method, $F(1, 23) = 0.196$, $p = 0.665$. When using the immediate coordination method, participants required an average of 2.784 seconds to complete each match, while participants using the interruption-controlled version required an average of 2.738 seconds.

Performance Measures	F	P	P< α
Accuracy			
Coordination Method	0.023	0.64	No
Frequency of Interruption	1.9	0.18	No
Interaction Effect	0.19	0.66	No
Object Classification Time			
Coordination Method	2.271	0.145	No
Frequency of Interruption	0.106	0.748	No
Interaction Effect	1.93	0.178	No
Number of Incorrect Object Classifications			
Coordination Method	1.245	0.276	No
Frequency of Interruption	2.735	0.112	No
Interaction Effect	0.42	0.524	No
Number of Objects Not Classified			
Coordination Method	12.951	0.0015	Yes
Frequency of Interruption	3.52	0.089	No
Interaction Effect	0.095	0.76	No
Number of Manipulation Errors			
Coordination Method	4.744	0.039	Yes
Frequency of Interruption	0.091	0.766	No
Interaction Effect	1.906	0.181	No

Table 2. ANOVA Analyses of the Main Effects and Interactions for the Performance of the Interruption Task

Several related measures were defined to assess user's accuracy when performing the interrupting task: overall accuracy, incorrect object classifications, missed objects, and manipulation errors. Accuracy was calculated as the percentage of correct object classification in relation to the total number of objects that appeared during 15-minute experimental treatment (eight objects per each trial, nine or twelve trials per treatment). There was no significance effect of coordination method of interruption, $F(1, 23) = 1.675$, $p = 0.215$, and the frequency of interruption, $F(1, 23) = 0.244$, $p = 0.878$ on the accuracy in the interrupting task. Similar analyses were performed on the number of mistakes made by the participants performing the interrupting task. Incorrect object classifications were approximately evenly distributed across all four conditions. Neither coordination method, $F(1, 23) = 1.245$, $p = 0.276$, neither frequency of interruption, $F(1, 23) = 2.735$, $p = 0.112$, showed significant difference.

A more accurate and realistic indicator of user's success on the interrupting task in the dual-task scenario was the examination of two other types of errors. There were two significant main effects found in this domain. The coordination method had significant difference on the following two variables, the number of manipulation errors, $F(1, 23) = 4.744$, $p = 0.039$, and the number of objects not classified, $F(1, 23) = 12.951$, $p = 0.0015$.

Manipulation errors belong to category of unintentional mistakes (omissions), which are related to the effects of interruption (Reason 1990). The comparison of the number of manipulation errors between conditions suggests that selecting more appropriate moments for interruption during mediated condition contributes significantly to providing

omission-free environment. In addition, participants classified more objects i.e., achieved better completeness on the interrupting task in the mediated conditions. In other words, user's success when performing the tactical assessment task depends on omissions (unintentional mistakes) and missed opportunities rather than just the number of correct and incorrect classifications.

Having established that participants had better results with the interruption mediator than without, we can take on the more difficult problem of explaining the contribution of the interruption mediation to the process. One possible explanation for the main effects was based on the extensive experimenter's notes. It was noticed that the rules for playing the interrupting task, which were available to the participants throughout the experiment, were reviewed on more occasions during immediate conditions. The number of incorrect classifications and accuracy did not reveal what was actually happening. The inspection of the notes suggested that participants experienced difficulties in recalling the rules for the interrupting task in immediate conditions and they made more incorrect classifications, which were not recorded by the system. The statement must be qualified by noting that the TA task allows participants to correct their mistakes by deleting their entry and reselecting before the system processes their entry. When noticed, the participants tried to correct their mistakes, which made those trials more difficult and time consuming. They had less time for classifying other objects, so some of the objects were missed. On the contrary, participant seemed to perform much better in the mediated conditions, i.e. they could accurately classify most of the flying objects, and they made less manipulation errors, depending upon the frequency of interruption.

DISCUSSION OF RESULTS

The results of the formal experiment have confirmed the hypothesized performance advantage of mediating interruptions compared to condition with immediate coordination. The analyses of variance have enabled the effect of coordination method and interruption frequency on the primary task resumption time to be quantified and have given statistically plausible results. Users receiving immediate interruptions took significantly more time to resume the primary task than users receiving computer-mediated interruption. The results also indicated that as the level of interruption frequency increases, participants were more likely to trade-off considerate planning with speed, therefore resumption times decreased.

Three out of five performance indices for cognitively simpler interrupting task were not affected by coordination method or varying frequency of interruption. Neither coordination method, nor interruption frequency influenced either user reaction time, or user's accuracy in performing the interrupting task. The coordination method did not significantly affect users' speed. The average time needed to classify an object was consistent across conditions. The lack of differences in these two performance metrics may in part reflect the fact that the interrupting task was short, less cognitively-demanding task consisting of independent and unrelated trials.

Participants' accuracy and reaction time did not change significantly over four conditions. Rather, it is in the domain of unintentional manipulation errors and missed opportunities where dramatic changes occur. The results have shown that the improved performance on the interrupting task during mediated conditions was related to two types of errors, the number of missed objects, and manipulation errors. The lack of a significant main effect for three out of five TA-related performance variables suggests that there is indeed some interference between the interrupting task and the primary task, but its effects are on particular types of errors rather than overall accuracy or speed. In view of the lack of strong evidence concerning the effect of interruptions on the interrupting task it would be useful to address this issue in further research. Several aspects that would seem to be worthy investigating include effects of varying types of interruption tasks with varying complexity, duration, or relationships with the ongoing task. The design space of the taxonomy provides enough possibilities to account for the complexity and relationship between interrupted and interrupting task with the same depth as it was investigated in the experiment presented in the previous section.

User's cognitive state when performing complex task is extremely fragile, and sensitive to external interruptions. The disruptive effects of interruptions pose a serious consideration, especially in time- and safety-critical systems. The results thus far show promising advantages for creating new user interfaces adaptable in face of interruption. The experience gained in this research that mediating interruptions supports users in performing complex tasks, is not likely to be unique, and we may believe that other domains and systems with similar types of tasks would also benefit from applying the proposed interruption model.

The interruption mediator have succeeded in recognizing interruption points sensitive to interruption based on the taxonomic factors that exhaustively describe the tasks and the interaction. By avoiding interrupting people during those, in general demanding cognitive subtasks, the participants were able to resume the interrupted task in much shorter times. The experience gained in this experiment that mediating interruptions support users in performing complex tasks, is not likely to be unique, and we may anticipate that other domains and systems with similar types of tasks would also benefit from applying the proposed interruption framework.

CONCLUSIONS

In this paper, we have discussed the results of an exploratory user study, conducted to examine the expressiveness and the effectiveness of the proposed framework for a computer-mediated coordination of human interruptions in HCI. The

proposed conceptual model is based on the new Interruption Taxonomy and uses Bayesian Belief Networks as a decision-support aid for mediating interruptions. The experimental results support the hypothesized performance advantage of mediating interruptions to lessen the disruptive consequences of interruptions reflected in shorter resumption times compared to condition with immediate coordination method. The initial results of the formal experiment are exciting because this research integrates multiple factors related to task characteristics as a basis for mediating interruptions, and empirically measures the quantitative and qualitative gains of such approach. The presented experimental findings presented are encouraging evidence that context-sensitive mediation is a step further toward facilitating user performance in a context of interruptions.

REFERENCES

- Bailey, B. P., Konstan J. A. and Carlis J. V. (2001) The effects of interruptions on task performance, annoyance, and anxiety in the user interface. *Human-Computer Interaction – INTERACT 2001 Conference Proceedings*, IOS Press, 593-601.
- Cellier, J.-M., and Eyrolle, H. (1992) Interference between switched tasks. *Ergonomics* 35, 1, 25-36.
- Cutrell E. B., Czerwinski M. and Horvitz E. (2000) Effects of instant messaging interruptions on computing tasks. *Ext. Abstracts CHI2000*, ACM Press, 99-100.
- Cypher, A., (1986) "The Structure of User's Activities", In Norman D.A., and Draper S.W., (Eds.), *User Centered Design*, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 243-263.
- Gillie, T., and Broadbent D. (1993) What makes interruptions disruptive? A study of length similarity and complexity. *Psychological Review* 50, Springer-Verlag, 243-250.
- Horvitz, E., Jacobs, A., and Hove, D. (1999) Attention-sensitive alerting. *Proc. of IUI '99*, 305-313.
- Jameson, A., Schafer, R., Weis, T., Berthold, A., and Weyrath, (1999) T. Making systems sensitive to the user's time and working memory constraints. *Proc. of IUI '99*, 79-86.
- Kirmeyer, S. (1988) Coping with competing demands: Interruption and the type a pattern. *Journal of Applied Psychology* 73, 4, 621-629.
- Malone, T., and Crowston, K., (1994), "The Interdisciplinary study of Coordination", *ACM Computing Surveys*, 26, 1, 87-119.
- Miyata, Y., and Norman, D.A., (1986) "Psychological Issues in Support of Multiple Activities" In Norman D.A., and Draper S.W., (Eds.), *User Centered Design*, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 265-184.
- Trafton J. G., Altmann E. M., Brock D. P. and Mintz F. E. (2003) Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal, *International Journal of Human-Computer Studies*, 58, 583-603.
- Reason, J., (1990) *Human Error*, Cambridge University Press, Cambridge, USA.
- Tulga, M.K., and Sheridan, T.B. (1980) Dynamic decisions and work load in multitasking supervisory control, *IEEE Transactions on Systems, Man & Cybernetics* 10, 5, 217-232.

COPYRIGHT

[Sonja Gievska, John Sibert] © 2004. The authors assign to OZCHI and educational and non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to OZCHI to publish this document in full in the Conference Papers and Proceedings. Those documents may be published on the World Wide Web, CD-ROM, in printed form, and on mirror sites on the World Wide Web. Any other usage is prohibited without the express permission of the authors.