UCRL-JC-108993
PREPRINT

Does the User Interface Make Interruptions Disruptive? A Study of Interface Style and Form of Interruption

Nancy A. Storch

This paper was prepared for submittal to the
Computer Interface '92 Striking a Balance 1992 ACM Conference on Human Factor on Computer Systems
Monterey, CA May 3-7, 1992

Lawrence Livermore National Laboratory

DISCLAIMER

Does the User Interface Make Interruptions Disruptive? A Study of Interface Style and Form of Interruption

Nancy A. Storch
Lawrence Livermore National Laboratory
University of California
P.O. Box 808
Livermore, CA 94550
(510) 422-8942
E-Mail: STORCH3@LLNL.GOV

UCRL-JC--108993

DE92 011295

Forward

This paper describes the experiment and reports the findings of a research project which looked at how unexpected interruptions disrupt computer user performance and whether there is a significant effect due to the style of user interface. In order to conduct the experiment, a conference room was converted into a temporary usability testing laboratory configured with borrowed office furniture and timing, eye-tracking, video, and computer equipment. The experimental testing took place in August, 1991. The use of human subjects was reviewed and approved by the Institutional Review Board. The formal name of the project was "Developing Advanced Assessment Methods and Metrics to Study the Effects of Interruptions Relative to Graphical or Character-Based Computer User Interfaces". This work was jointly funded by the Advanced Technology Center in Administrative Information Systems Department and Applications Development Department.

I want to express my deep appreciation to the twenty computer scientists and systems analysts who volunteered to participate as subjects. Their enthusiasm and interest were remarkable. It was not easy to be a subject, with the headtracking optics clamped on one's head, doing boring data entry work.

In addition, there are a number of people whom I want to recognize for their special encouragement and support. Professor Stuart Klapp of California State University, Hayward, guided the analysis. Jose Velez, Applied Science Laboratories, lent us an eyetracking system and taught us how to use it. Tom Bennett, NASA Ames Research Center, lent us a headtracker out of his virtual reality system. The following people are LLNL employees. Tim Sharick, Advanced Technology Center, provided the nurturing environment which allowed for the success of this project. This included the space, computers and Montana State student Leif Nelson. Leif was my key assistant during the test sessions, and also wrote the HyperCard data entry interfaces and installed his modified ProtoTymer. Tom Morado and Don Tucker, Technical Information Department's video services, configured all of the video equipment and were readily available to handle problems during the experiment. Applications Development Department provided Debbie Streets who very capably recruited and scheduled the volunteers. Eric Henson, Small Systems Support, saved me weeks of time reducing the eye fixation data by writing EXCEL macros. Mimi Alford, Lawrence Livermore Television Network, used her creativity to edit scene and eye video session tapes into something suitable for presentation.

Others at LLNL who have encouraged me and assisted in the review of this document are: Bill Banks, Derek Hendry, and Eugene Schultz.

Abstract

This experiment examined the influence of the style of the computer user interface on the extent to which performance was decreased by task interruption. Performance on data entry of a personnel database was compared using two different styles of interfaces, under three different forms of interruption. Ten subjects were given a graphical user interface with a mouse and screen buttons; and ten were given a character-based interface with tab and function keys. Interruptions came at unexpected times. Each subject was interrupted three times, once with each form of interruption: telephone call, visitor, and on-screen message. During each interruption, the subject was asked to answer a simple question. The number of fields entered minus the number of fields in error, per minute, was measured during the two-minute post-interruption period. Eye motion data was analyzed in terms of the amount of time and the average durations of eye fixations on the screen. An exit questionnaire solicited the subjects' perceptions of their performance.

*This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

Compared to the control, the screen interruption was disruptive, the telephone interruption was not, and the walk-in visitor interruption produced intermediate performance. Performance was worse in the first minute than in the second minute immediately following the interruptions. Lower performance was directly related to looking more at the screen. The duration of the interruption was weakly correlated to the performance. Subjects' perceptions of their performance did not correlate with performance measures.

The disruptive effect of the screen interruption compared to the lack of disruption after the telephone interruption was a surprise. This result is contrary to the commonly held belief that telephone interruptions are a problem to computer users. It may be important to avoid disruption caused by screen windows popping up in design of multi-window user interfaces. Trends in office automation have the screen being used for more and more functions, with demands on screen displays being further aggravated by information available through integrated networks.

The disruption of the screen interruption may be due to the abruptness of the on-set of the interruption which prohibited the subject from completing the current action. The message box locked out the main task until the subject responded to the question. This differed from the onset of the walk-in and telephone interruptions which did not lock out interaction with the main task. Another possible explanation for the disruption of the screen interruption compared to the telephone interruption may be the similarity in sense modality. This work is seen as preliminary to further study.

Introduction

Interruptions have been studied since the classic experiments of Zeigarnik and Ovsiankina in the 1920's. Zeigarnik (1927) studied the effects of interruptions on recall. She showed that if people were interrupted during some tasks, but allowed to continue with others, then the interrupted tasks were recalled more often than the uninterrupted ones. Ovsiankina (1928) showed that subjects in such experiments, if left to their own devices, would actually try to complete the interrupted tasks.

Modern research in interruptions has looked at how easily people can resume what they were doing when the interruption finishes. Kreifeldt and McCarthy (1981) were comparing Reverse Polish Notation and Algebraic Notation interface designs of calculators. They found interruptions to be disruptive for two groups of subjects, each group solving problems with a different style calculator. The interruption was writing down multiplication tables for a minute. The length of time of the interruption and non-interrupted task were roughly the same. Kreifeldt and McCarthy suggested that the similarity of the interruption to the main task might be a factor in the noticed slow down after the interruptions.

Gillie and Broadbent (1989) conducted a series of experiments to explain why some interruptions are disruptive and some are not. Subjects were presented with a series of 12 separate problems, each of which had an associated list of five or seven items. The subject was to locate items given on the lists. When the subject found the middle item on the list, the interruption occurred, the screen was cleared and the interrupting task appeared. When the interruption was over, the subject continued with the primary task. The length of time for a subject to complete the series of problems without interruption would have been about 22 minutes. Two experiments used short interruptions of 30 sec and two used long interruptions of 2.75 min. When the interruption was simple mental arithmetic (dissimilar to the main task) and the subjects could review position in the main list before proceeding to the interrupting task, for both long and short interruptions, the interruptions were found not to be disruptive. Since the main tasks of Gillie and Broadbent were more difficult than the calculator tasks of Kreifeldt and McCarthy, the result suggests that the memory load at the time of an interruption is not a crucial factor in determining whether or not it will be disruptive.

Interruptions were disruptive for another Gillie and Broadbent experiment under similar conditions. In this case, the interruption task was complex and of short duration. The interruption task was still simple arithmetic, but the numbers were coded as letters. The complexity of the interruption task was the determining factor in the disruption. Interruptions were also found to be disruptive when the interruption task was similar to the main task and the subjects were not given time to review the position in the main list before going on.

The issue of similarity of interruption task to main task was further explored at NASA/Johnson Space Center Human-Computer Interaction Laboratory, where experimenters (Czerwinski, Chrisman, Schumacher and Rudisill, 1991) examined the performance of subjects who were monitoring space station information and interrupted by a task which replaced the information on the computer screen. For least disruption, the experiments indicated that information displayed by an unexpected screen interruption should be in a format dissimilar to the tasks which were interrupted, but that this may not be necessary when an operator is forewarned that an interruption is going to occur.

The purpose of the present study was to find out how the disruptiveness of an interruption may vary from one user interface to another, and whether some forms of interruption are more disruptive than others. The main task was data entry of a sample personnel database. The interruption tasks were chosen to be dissimilar to the main task. These (within subject variable) were requests to reply to simple questions.

All questions were given to all subjects and had three forms: a telephone call, a walk-in visitor with a question, and an on-screen query. Two styles of interfaces (between subjects variable) were used. Half of the subjects were given a graphical user interface and a mouse to select data entry fields and the other half of the subjects were given a character-based interface and use of the tab key to select data entry fields. Records were entered either with a button press or by hitting a function key, respectively. Performance was measured in terms of number of correct fields entered per minute. A limited amount of the eye motion data was analyzed, comparing fixation durations on-screen and off-screen.

The prediction was that performance after any interruption of the mouse-based graphical user interface would have greater degradation than performance after any interruption using the keyboard character-based interface. The reasoning behind this was that the graphical user interface requiring mouse positioning is a more complex interaction, and thus more susceptible to interruption. No prediction was made regarding differences between the forms of interruptions, although it is commonly believed that telephone interruptions are detrimental to getting work done.

Method

Subjects. A memo was sent to approximately 300 computer scientists and systems analysts from the sponsoring LLNL departments asking for volunteers. 20 people responded who also met the screening criteria of using a Macintosh weekly and having typing speed between 20 and 50 words/minute. The volunteer subjects were not paid. Although they knew that aspects of usability were being studied, the subjects were unaware that the purpose of the experiment was to study the effects of interruptions. Subjects were randomly assigned to user interface groups, interruption sequences and question sequences.

The first three subjects showed noticeable subject fatigue. The time spent on the data entry task was thereafter reduced from 24 to 16 minutes for the rest of the subjects. This seemed to solve the fatigue problem. Because of this change in the experiment, the test sessions of the first three subjects had to be excluded from the analysis. One other subject was excluded from the analysis because of not feeling well enough to complete the session. This left 16 subjects, 9 men and 7 women whose performance was analyzed. They were divided: 9 subjects using the character-based interface (4 men and 5 women), and 7 subjects using the graphical interface (5 men and 2 women). Eye motion data was lost for two subjects, a man and women, both using the character-based interface when, for unknown reasons, their raw data files could not be reduced to eye fixation data.

Apparatus. Figure 1 gives an overview of the video and digital data collection which was synchronized with a SMPTE (Society of Motion Picture and Television Engineers) time code generator. The eye tracking was done with an Applied Science Laboratories 4100H head mounted system. A Polhemus 3SPACE TRACKER was used to track the head motion. The eye-tracking system EYEHEAD software combined head position with eye position to determine the point of gaze regardless of head motion. The eye tracking system was supported by: a 80386 33Mhz PC with 80387 co-processor, color VGA monitor, 200 Mbyte hard disk, and B/W video recording of subject's eye view with point-of-regard cursor superimposed. The SMPTE 26-bit digital BCD output was cabled to the 16-bit eye tracker event input signal (dropping non-important bits) in order to time synchronize the video eye tracking images with the digital eye tracking data. Color video of the subject from the side was also recorded. All video was Genlocked to the vertical sync of the B/W eye video in order to maintain synchronization to the SMPTE time code.

Procedure. Interruption sequences of the three different forms were randomly generated and randomly assigned to the subjects. One of three questions was also assigned to each interruption (see Appendix A). The subjects used a Macintosh IIfx with a 19-inch color monitor. The data entry application user interfaces were written in HyperCard and the data entry fields appeared the same on the screen. The graphical user interface operated with a mouse and screen buttons; the character-based interface with tab and function keys. The LLNL modified version of ProtoTymer (Miller and Stone, 1989) was used to: run the sessions, capture and time-tag the mouse clicks, button presses, function keys, and data entry fields, and export the results for performance analysis.

Subjects were tested individually in an office mock-up with Macintosh, telephone, chair, pad and pencil, ruler, paper holder, wall calendar, wall clock and wall pictures. The Macintosh was fixed, the chair was adjustable, the telephone was located to the right of the Macintosh to prevent the subject from inadvertently knocking the eye-tracking optics (located on the left side of the headband) when answering the telephone. Each subject read a sheet of instructions, watched a demonstration, was introduced to the mock office and told to consider it his/her own, and then was given a practice sheet of personnel data and as much time as the subject wanted to learn and ask questions. When the subject was ready, the headband holding the optics for the eye tracker was fit on the subject's head and the subject was lead through the eye calibration procedure. Each subject was given the same data entry sheets, with personnel data arranged in rows, and told to work at his/her own pace for about 20 minutes until signalled that the session was over. The subject was alone except for the the walk-in interruption. In order to enter the data correctly, the subject

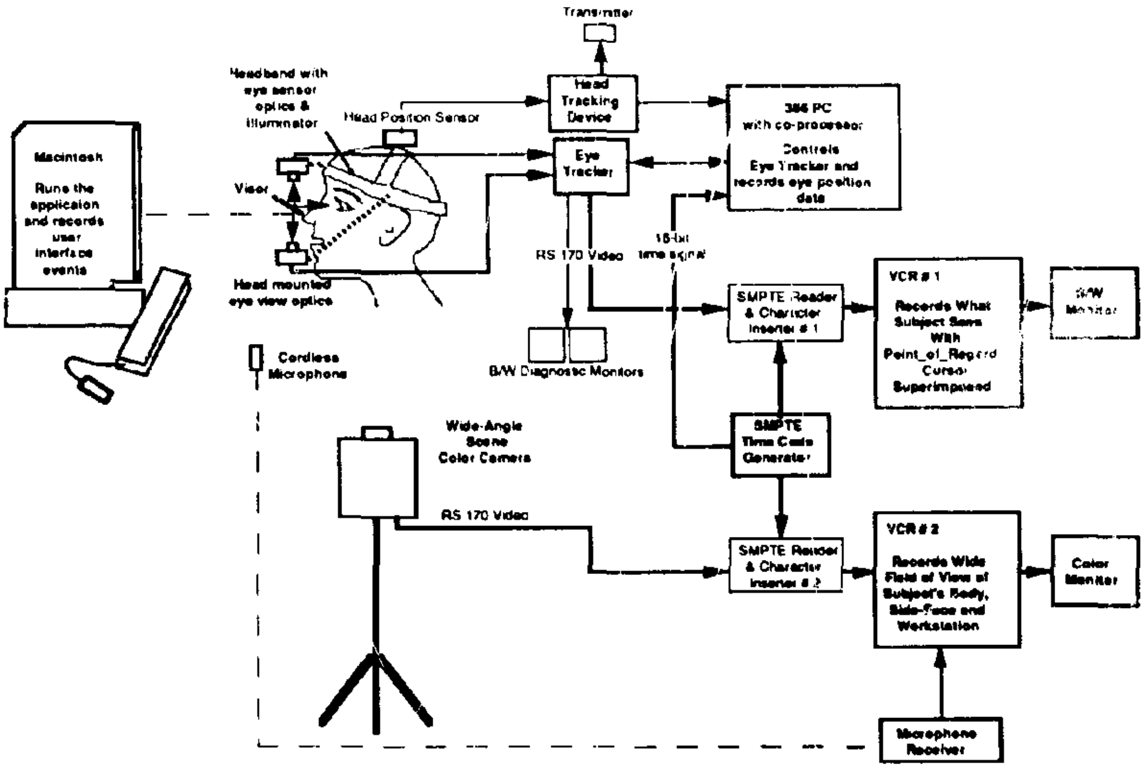

Figure 1. Overview of video and digital data acquisition.

had to look carefully at the data on the sheets because "name", "address", etc. were organized in columns, but not in the same order as the fields displayed on the screen which had to be filled-in.

Each subject's session was scheduled into four segments, with one of each form of interruption assigned to three segments, and the other segment used as a control condition. The length of time of the control condition was four minutes. For each interrupted segment, the subject entered data for two minutes, the interruption occurred, the subject responded to the query, and then returned to entering data for another two minutes. Thus from the end of one interruption to the beginning of the next, the length of time was either four or eight minutes, depending on whether or not there was an intervening control period. The length of time of the interruptions varied depending on the subject's response. The subjects completed a questionnaire (see Appendix B) at the conclusion of the session.

Subjects were separated by a screen from the observation area of the room, which contained video and eye tracking systems. A telephone interruption was created at the appropriate time by buzzing a secretary in an adjacent office who then made the interrupting call. Likewise, she was buzzed when it was time to send an on-screen "Broadcast" message. The walk in interruption was created by the experimenter going around the screen over to the subject on the other side of the room and breaking in with a question.

Eye Tracking. Eye point-of-gaze data were collected by the eye tracking system which illuminates the eye with near-infrared light producing a backlighted bright pupil and corneal reflection. The separation between the pupil and the corneal reflection varies with change in point of gaze but does not vary significantly with eye translation. A change in pupil-corneal reflection separation is approximately proportional to the change in point-of-gaze. The system computes the vector distance between the centers of the pupil and corneal reflection to determine direction of gaze. Fixations (point-of-regard as subject looks at a stationary target in a visual field) are distinguished from saccades (rapid voluntary eye movements used to move from one fixation point to another) and very small involuntary eye movements which occur during fixation. The raw x-y eye coordinate data were smoothed and eye fixations were identified from the filtered data. A fixation is defined when the point-of-gaze remains within a one degree by one degree area for at least 100 ms.

Results and Discussion

Performance. The interruptions affected performance depending on the form of the interruption. The screen interruption was disruptive. The walk-in interruption may have been disruptive. The telephone interruption was not disruptive. The overall error rate was 5.9% of the total entries. The mean number of correctly entered fields (number of fields entered minus number of fields entered in error) per minute appear in Table 1. This represents data from the entire two minute post-interruption period.

Computation of analysis of variance (ANOVA) for mixed design was done with the statistical computer program CLR ANOVA. The effects of each independent variable (style of interface, form of interruption, minute after the interruption) and the interactive effects of combinations of independent variables were examined. The ratio which defines Fisher's F-test represents a comparison of between-groups variability to within-groups variability. The larger the F ratio is, the more certain it is that the independent variable had an effect on the dependent variable. A significance level of $p < .05$ was used.

The statistical analysis separated performance for the first and second minutes, however there were no significant interactions involving minutes. Performance was worse in the first minute than in the second minute, $F(1,14) = 8.966$, $p < 0.01$. That is, when looking at the independent variable, minute, the F ratio is 8.966 where the degree of freedom for minute (numerator) is 1, the degrees of freedom for the error variance (denominator) are 14, and the probability of F occurring by accident is < 1%. The means for the first and second minute were 5.32 and 6.21 correct entries, respectively.

The number of correct entries tended to be higher in the character-based than in the graphical interface condition, but this difference was non-significant, $F(1,14) = 1.833$, $p=0.197$. There were no significant interactions involving the style of interface.

The subjects were slowed by the screen interruptions, i.e., the number of entries was less after the screen interruption than in the control, no-interruption condition, $p < 0.05$. Performance in the other two interrupt conditions did not differ significantly from the control. Performance was significantly less after the screen interruption than after the telephone interruption $p < 0.05$, but performance did not differ significantly between the screen and walk-in conditions (pairwise comparison, Duncan multiple range test).

Eye Motion. The subjects' longer duration fixations looking at the screen correlated with decreased performance for the screen interruptions. The eye motion analysis dealt only with the two minute post-interruption period, although data was collected for the entire session. On-screen vs off-screen fixations were analyzed. To be on-screen, a fixation had to be within the boundary of the 19 inch Mac monitor. The off-screen fixations were everything else, including the data entry sheet, the keyboard, the telephone, the visitor, the table and the room. One subject using the character-based interface propped his data entry sheet up in front of the monitor for most of the session. This gave higher on-screen

Table 1. Mean number of correct fields entered per minute during the two-minute post-interruption period.

Style of interface	Form of interruption								
	Screen		Walk-in		Telephone		Mean	Control	
Graphical									
1st minute	4.29		4.71		5.14			5.29	
2nd minute	5.14		5.57		5.29			5.43	
both minutes		4.71		5.14		5.21	5.02		5.36
Character-based									
1st minute	4.78		6.00		6.00			6.33	
2nd minute	6.78		6.56		7.67			7.22	
both minutes		5.78		6.28		6.83	6.30		6.78
Mean (weighted)		5.31		5.78		6.13	5.74		6.16

percentages for him, which were not noticed in the overall averaging. The fixation data that were collected for the two minute control period accounted for fixation durations totaling 20 seconds to 100 seconds per subject. This was due to variation in ability of the eye tracker to follow the subject's pupil, the subject's physical characteristics and eye motion, and in a few cases, movement of the eye tracking optics on the headband during the session. Thus the results are presented in terms of percentages and means, rather than direct measures.

As in the analysis of performance, the statistical analysis of eye fixations showed no significant interactions involving minutes, so the data from the entire two minute post-interruption period was used. The graphical interface users looked at the screen 30% of the time; the character-based interface subjects looked at the screen 22% of the time. The percent time spent looking at the screen by the character-based interface subjects is consistent with the 24.6% time measured by Zwahlen, Hartmann, Rangarajulu and Escontrela (1985) who compared data entry from hardcopy and from screen. Table 2 gives the mean percent of time spent looking at the screen. The trend was to look more at the screen after the screen interruptions, and less after the telephone interruptions. The telephone interruption caused the subjects to look offscreen more than the other interruptions or control, $p<0.05$ (pairwise comparison, Duncan multiple range test). The subjects looked offscreen: 80% after telephone interruption, 68% after screen interruption, 75% after walk-in interruption and 74% during control.

On-screen fixations were compared with performance. Four correlations were done, comparing the percent of time looking on-screen to the number of correct entries, for each interruption during the two-minute post-interruption period. The correlations were strong for the screen and telephone interruptions, but not for the others (correlation coefficients are: $r_{(screen)}(12) = -0.739, p<0.01$, $r_{(telephone)}(12) = -0.758$, $p<0.01$). The conclusion is that looking at the screen and entering correct data are related in such a way that a lower number of entries corresponds to looking more at the screen, especially after a screen or telephone interruption.

The mean times of fixations, on and off the screen are presented in Table 3. The trend was for the fixations of graphical user interface subjects to be of longer duration than those of the character-based user interface. For the control period, on-screen and off-screen durations were about the same regardless of group. After the telephone interruption, on-screen and off-screen fixation durations decreased for both groups. The largest drop (210 ms) was on-screen for the character-based user interface group.

Questionnaire. Subjects' perceptions of performance did not generally match performance measures. The subjects filled-out a questionnaire at the conclusion of the session. The questions, together with the mean values broken down for the graphical user interface (GUI) group and the character-based used interface (CUI) group appear in Appendix B. The two user interface groups answered the questions in a very similar way. The largest difference was 19% where the graphical user interface subjects thought that the walk-in and telephone interruptions reduced their rate of working more so than the character-based user interface subjects did. The CUI subjects thought that the on-screen interruption reduced their rate of working more than the other interruptions (matches performance), while the GUI subjects noticed no difference (contrary to performance). The GUI group thought that the interruptions reduced their rate of working more than the CUI group did (contrary to performance). Both groups did not think that the interruptions affected them very much (mean 4.04 out of scale 1 to 9). The GUI group thought that they made more errors after the walk-in and telephone interruptions (true, 7 to 1 for screen), while the CUI group thought that they made more errors after the on-screen interruption (false, 13 errors each for screen and walk-in/telephone). The GUI subjects indicated that they somewhat more often use keyboard style user interfaces, while the CUI subjects indicated that they somewhat more often use graphical style of user interfaces (means 5.86 to 4.67 out of scale 1 to 9). Both groups indicated that they are interrupted in their work an average of three to four times an hour. Three subjects indicated that they are interrupted an average of over six times an hour.

How eye motion relates to performance. Subject performance after the telephone interruption was the best, and

Table 2. Percent of time eye fixations were on-screen during the two-minute post-interruption period.

Style of interface	Form of interruption				
	Screen	Walk-in	Telephone	Mean	Control
Graphical	34	32	25	30	29
Character-based	30	18	16	22	23
Mean(weighted)	32	25	20	26	26

Table 3. Average length of fixations (in milliseconds) on-screen and off-screen during the two-minute post interruption period.

Style of interface	Form of interruption					
	Screen	Walk-in	Telephone	Intrpt Mean	Control	Mean
On-Screen						
Graphical	267	256	243	256	252	255
Character-based	258	210	229	237	255	242
Mean (weighted)	263	238	238	248	254	
Off-Screen						
Graphical	234	274	250	252	252	252
Character-based	263	237	232	244	269	250
Mean (weighted)	249	254	241	248	261	

performance after the screen interruption was the worst. Subjects spent the least amount of time looking at the screen after the telephone interruption and the most amount of time looking at the screen after the screen interruption. The fixations were shorter after the telephone interruption (compared to the control) and they were longer after the screen interruption. This relationship probably also holds true for dwell times (continuous fixations at the same location). Zwahlen, et al. (1985) found that dwell times reflect the amount of mental information processing involved. In their studies, long dwell times looking at the reference document were noticed for subjects while doing a data entry task and when looking at the screen for detection and correction of errors in a maintenance task, whereas short dwell times were noticed for keyboard verification. The conversion of the eye motion data from this experiment to dwell times and a more detailed break-down of the off-screen locations could confirm that subjects were slowed by mental processing when they looked at the screen. For the data entry task, subject performance during the post-interruption period decreased in relationship to how much the subjects looked at the screen.

In the analysis, eye motion data was lumped into two-minute post-interruption intervals. The big effect of disruption in eye motion patterns may be seen better by careful study of much smaller intervals of only a few seconds duration. Further study could also confirm the implication that subject's eye pattern disruption due to an interruption continues for a number of minutes after the interruption and is related to performance.

Why some interruptions are more disruptive than others. Why are some interruptions more disruptive than others? One possibility is that the duration of the interruption determines how disruptive it is. The duration of the interruption was defined to be the length of time measured from the onset of the interruption to when the subject resumed the main data entry task. The durations varied from 11 seconds to 1.76 minutes. All subjects were required to respond to the interruptions, but how much and how long a subject responded was up to the individual. The mean durations of the interruptions appear in Table 4.

The duration of the screen interruption was longer than the others, but this was true only for the character-based interface condition. This produced a significant interface-by-interruption interaction, $F(2,28) = 3.827, p < 0.05$. Looking only at the character-based interface condition, the form of interruption is significant, $F(2,16)=5.705, p = 0.014$. The screen interruption is significantly longer than both the walk-in, $p<0.01$ and telephone interruptions, $p<0.05$, Duncan test. The durations of the walk-in and telephone interruptions do not differ significantly from each other.

The issue of whether a longer duration of the interruption leads to a reduced number of entries was explored. If so, then the correlation of duration and entries should be negative. The correlational analysis treated each interruption separately, i.e., there were 16 subjects times 3 interruptions = 48 paired entries to be correlated. This correlation was: $r(46) = -0.259, p<0.05$. The conclusion is that duration

Table 4. Duration of interruptions in minutes.

Style of interface	Form of interruption			
	Screen	Walk-in	Telephone	Mean
Graphical	0.40	0.33	0.52	0.42
Character-based	0.76	0.37	0.47	0.53
Mean (weighted)	0.60	0.36	0.49	0.48

and entries are related in such a way that longer interruptions were more disruptive. But this correlation is not very strong, and there are other factors that may account for the stronger disruptive effect of the screen interruption.

Gillie & Broadbent (1989) found duration not to be related to how disruptive the interruption was. In their experiments both the main and interruption tasks were on the screen, whereas this experiment shows longer interruptions to be more disruptive. It is not clear how to account for this difference. Perhaps it is due to the element of predictability; in this experiment, the interruptions were a surprise to the subjects, whereas Gillie & Broadbent's interruptions always occurred after the middle item in the list was found.

One way that the screen interruption differed from the telephone and walk-in interruptions is that it prevented the subject from completing the in-progress entry. When the screen interruption occurred, the message question came on the screen overlaying some of the data entry fields; the main task could not be resumed until the subject responded to the question. When the telephone rang, the subjects were able to reach over and answer any time within a few seconds; some continued to hold their place or enter data while answering. When the experimenter walked-in, the subjects had a wide range of responses, from completely stopping what they were doing, to continuing the data entry task while listening to the question.

The prevention of completing the in-progress entry may have emotional or other implications that endure into the two minutes of entry activity following the interruption. The classical Zeigarnik effect is an enduring effect do to the prevention of task completion. Memory for a task that is interrupted in this way is enhanced. An interruption occurred while a subject was entering data for 93% of the interruptions. Subjects continued throughout the walk-in and telephone interruptions to complete entry of data fields (75% of all subjects: 100% of the graphical user interface subjects, 56% of the character-based user interface subjects). The largest difference in behavior between groups occurred after the walk-in interruption, where 71% of the graphical user interface subjects completed fields compared to only 33% of the character-based user interface subjects. These data indicate that subjects wanted to work through the interruptions and that the graphical user interface made it easier to work through the walk-in and telephone interruptions.

Although the present data are far from conclusive in identifying the cause of the disruption of the screen interruption, there is a strong indication that this may be happening. Providing the ability to complete the in-progress entry may be a key to designing user interfaces to allow efficient handling of user interruptions.

The similarity in sensory modality may also be a factor in the strong disruption of the screen interruption. This could be another instance of a larger principle that similarity yields disruption and lower performance compared to dissimilarity. Supporting this idea is the work by Gillie and Broadbent which found that interruption of a task with a similar task was disruptive, and Czerwinski, et al which found that subjects had significantly better recall of a primary task when the interruption was dissimilar in format to the primary task's.

Conclusions

The disruptive effect of the screen interruption compared to the lack of disruption after the telephone interruption was a surprise. This result is contrary to the commonly held belief that telephone interruptions are a problem to computer users. Disruption caused by screen windows popping up may be important to avoid in design of multi-window user interfaces.

The disruption of the screen interruption may be due to the abruptness of the on-set of the interruption which prohibited the subject from completing the current action. The message box locked out the main task until the subject responded to the question. This differed from the on-set of the walk-in and telephone interruptions which did not lock out interaction with the main task. Another possible explanation for the disruption of the screen interruption compared to the telephone interruption may be the similarity in sense modality.

The two cases of user interface style were distinguished by the use of a mouse and screen buttons vs the use of tab and function keys. Although, the user interface style was not a significant factor in the disruption of interruptions, further study of different variations of style and input devices may have different results. Performance decreased in relationship to how much the subjects looked at the screen. When given the chance, subjects tended to complete entry of a data field during the interruptions. The character-based user interface subjects entered about 25% more correct fields per minute, and spent 27% less time looking at the screen.

In the workplace, computer routing of messages to the screen and user message utilities may have unsuspected disruptive consequences. More needs to be known about what causes screen interruptions to be disruptive so that user interfaces can be designed which are least disruptive. Independent user interfaces may be necessary when users are required to do two or more tasks at the same time. This work is seen as preliminary to further study.

Future directions for research

A series of experiments are suggested to conclusively determine whether prohibiting the subject from completing an on-going entry enhances the disruptive effect of the interruption. This might involve using two types of on-screen interruptions: one in which the interrupting task inhibits work on the main task, the other in which the interruption is announced but work is not inhibited on the main task. It would be interesting to note whether having each task use a separate screen reduces the disruptive effect. Continue to look at both graphical and character-based user interface styles. Look at both fixed length short and long duration interruptions. Some of the subjects mentioned that the data entry task was boring and that the interruptions were interesting, whereas in their workplace the opposite holds. Follow-up experiments should permit the comparison of the effects of disruption on both the data entry task used in this experiment and a task more like programming, for which interruption of thought may lead to more serious consequences.

Relevance to LLNL

Both interruptions and computer use are a part of the every day work environment. At LLNL there also are a large number of software developers. This gives LLNL the resources to develop custom user interfaces. For projects where user performance is a concern and interruptions are a problem, LLNL can strive to minimize the adverse effects by careful design of the computer-user interface. It is commonly believed that telephone interrupts are bad. The results of this research showed that this isn't necessarily true. Screen interruptions may or may not be disruptive, depending on a number of factors. This research and past research provide information which can be used in the design of better on-line message systems which minimize the disruption of receiving and responding to message interruptions. Scheduling of special times to receive telephone calls and visitors in offices may also improve productivity. In order to recommend specific changes in the workplace, more needs to be known about the factors that determine the disruption of interruptions and how they relate to the work environment.

The departments supporting this research project are most interested in the effects of common workplace interruptions on personnel who spend most of their time interacting with personal computers and/or workstations. The personnel are mainly: computer scientists, systems analysts, and data entry technicians. Typical workplace interruptions are unexpected and include: telephone calls, on-line dialogs with other users, and unexpected visitors.

This research project was also an exploratory step by LLNL software development departments into usability testing. Usability testing is a method commonly used in industry for quality assurance. Usability testing evaluates the usability of computer products according to criteria set forth in the product requirements. The development of successful user interfaces is a process easiest done in a prototyping environment. The iterative steps in developing the user interface are: design, evaluate, analyze, (re)design, (re)evaluate, (re)analyze, etc. As LLNL develops or acquires computer applications and systems which have operational usability requirements, human factors evaluation becomes a necessity and a usability testing laboratory a serious consideration. Higher usability requirements may also be required for computer applications which are transferred from LLNL into the private sector.

The construction of the temporary usability testing laboratory for this project required a large effort to bring computers, video and eye tracking equipment together into one system. New frontiers were explored in preparing the experimental design and software for the experiment, as well as in soliciting the volunteers. The process of undertaking this project was a valuable learning experience which can assist LLNL in its future human factors work.

References

Czerwinski, M.P., Chrisman, S.E., Schumacher, B., & Rudisill, M. (1991). Interruptions in Multitasking Situations: The Influence of Similarity and Immediacy, poster presentation. *CHI'91 Human Factors in Computing Systems.*

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

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

Miller, D. P. and Stone, A. C., 1989, ProtoTymer: Human Performance Instrumentation for HyperCard Prototyping. *Proceedings of the Human Factors Society 33rd Annual Meeting -1989.*

Ovsiankina, M. (1928). Die Wiederaufnahme unterbrochener Handlungen. *Psychologische Forschung, 11,* 302-379.

Zeigarnik, B. (1927) Das Behalten erledigter und unerledigter Handlungen. *Psychologische Forschung, 9,* 1-85.

Zwahlen, H., Hartmann, A., Rangarajulu S., Escontrela, L. (1985). Eye Scanning Behavior when Working on Video Display Terminals. In R. Groner, G. McConkie, & C. Menz, Eds.: *Eye Movements and Human Information Processing* . North-Holland, Amsterdam 385-402.

Appendix A.

Secretary preface:

"Hi, this is Lisa from Nancy Storch's office."

Walk-in experimenter preface:

"Sorry for cutting in just now, but"

Queries:

"on Monday, September 10th Nancy is (I am) having a coffee and donut reception for all of the volunteers on this project. We (I) would like to know if you will be able to make it?"

"Nancy is (I am) going to be writing a paper on the results of this project and we (I) would like to send you a copy of the final study. Could I please have your Department and L-code?"

"Nancy has asked me to do (I am doing) an auxiliary survey to the current study regarding PC's in the home. Do you have a computer in your home?"

Appendix B.

• Session ID:______________________ Date:__________

Exit Questionnaire

Please take a few minutes to complete questionnaire and leave it here

Job classification:______________________________

Years of computer experience:

__ less than 1 year __ 5 - 10 years

__ 1 - 3 years __ 10 - 20 years

__ 3 - 5 years __ over 20 years

Please answer the following questions subjectively.

1. Did you notice anything peculiar as you worked? ______

   What? ______________________________

   If so,

   1.1 Did you expect this? ______

   Why? ______________________________

   1.2 Did you do anything about it? ______

   What? ______________________________

2. How easy was it for you to do the assigned task(s)?

   Results: GUI 5.0, CUI 4.56, Overall 4.75

   extremely easy ... not easy

   1 2 3 4 5 6 7 8 9

3. Did the headband hamper your ability to work?

   Results: GUI 4.71, CUI 5.22, Overall 5.0

   not at all ... very much

   1 2 3 4 5 6 7 8 9

4. Did the video taping hamper your ability to work?

   Results: GUI 1.43 CUI 1.11, Overall 1.25

   not at all ... very much

   1 2 3 4 5 6 7 8 9

5. How much did the on-screen message(s)/interruption(s) reduce the rate of your working after the interruption(s)

compared to before the interruption(s)?

Results: GUI 4.71, CUI 3.78, Overall 4.19

not at all very much

1 2 3 4 5 6 7 8 9

6. How much did the **off-screen** message(s)/interruption(s) reduce the rate of your working after the interruption(s) compared to before the interruption(s)?

Results: GUI 4.71, CUI 3.22, Overall 3.88

not at all very much

1 2 3 4 5 6 7 8 9

7. How much did the **on-screen** message(s)/interruption(s) increase your tendency to make mistakes after the interruption(s) compared to before the interruption(s)?

Results: GUI 3.43, CUI 3.22, Overall 3.31

not at all very much

1 2 3 4 5 6 7 8 9

8. How much did the **off-screen** message(s)/interruption(s) increase your tendency to make mistakes after the interruption(s) compared to before the interruption(s)?

Results: GUI 4.0, CUI 2.78, Overall 3.31

not at all very much

1 2 3 4 5 6 7 8 9

9. What style of user interface do you use in your computer work?

Results: GUI 5.86, CUI 4.67, Overall 5.19

graphical only keyboard only

1 2 3 4 5 6 7 8 9

10. How much are interruptions a problem when you are working with a computer?

Results: GUI 4.86, CUI 3.89, Overall 4.31

not at all very much

1 2 3 4 5 6 7 8 9

11. On the average, how often are you interrupted while working with a computer?

Results (normalized to 1 to 9 scale): GUI 4.71, CUI 4.83, Overall 4.78

__ less than once an hour __ 4 - 6 times per hour

__ 1- 3 times per hour __ over 6 times per hour

12. Any comments regarding usability testing?

____

13. Any comments regarding this experiment?

____

Thank you again for all the time you have given us.