Human–Computer Interaction — INTERACT'99
Angela Sasse and Chris Johnson (Editors)
Published by IOS Press, IFIP TC.13, 1999, 295-303

Coordinating the Interruption of People in Human-Computer Interaction

Daniel C. McFarlane

Naval Research Laboratory, Code 5513
Washington, D.C. 20375, USA
+1-202-767-2116, mcfarlane@acm.org

ABSTRACT People have cognitive limitations that make them sensitive to interruption. These limitations can cause people to make serious mistakes when they are interrupted. Unfortunately, interruption of people is a side effect of systems that allow users to delegate tasks to active background processes, like intelligent software agents. Delegation carries the costs of supervision, and that often includes being interrupted by subordinates. User interfaces for these kinds of computer systems must be designed to accommodate people's limitations relative to being interrupted. A theory-based taxonomy of human interruption was used to identify the four known methods for deciding when to interrupt people. An experiment was conducted with 36 subjects to compare these four different design approaches within a common context. The results show important differences between the four user interface design solutions to the problem of interrupting people in human-computer interaction (HCI).

KEYWORDS interruption, human-computer interaction, user interface design guidelines, experiment, mixed-initiative dialogue, intelligent software agents, coordination

1 INTRODUCTION

Advances in computer technologies have increased the practicality of building systems that allow people to perform multiple activities at the same time. However, people's cognitive capabilities have not increased. It is possible that these technological advancements carry unfortunate side effects that conflict with people's unchanging cognitive limitations.

The telephone is a familiar example. This technology allows people to do several things concurrently and is useful for isolated conversations. However, in real work environments it also allows people to have several concurrent dialogues that become intermixed over time. This kind of multitasking is useful and natural, however it also introduces the unfortunate side effect of causing people to be interrupted. Telephone users must accept a certain amount of interruptions as the unavoidable cost of doing several things concurrently. Intelligent software agents cause this same problem. These systems can be assigned to do useful things in the background while their human users work on other tasks. However, whenever an agent must initiate an interaction with its user, it must first interrupt them from whatever else they are doing.

Interruption of people is problematic because people have cognitive limitations that restrict their ability to work during interruptions. For example, an interruption of a commercial airline crew before takeoff contributed to their subsequent crash of the plane. A Northwest Airline crew was preparing to fly out of Detroit. They began their pre-flight checklist, but were interrupted by an air traffic controller with new taxiing instructions and a warning about wind shear. After the crew finished talking to the controller they made the mistake of not resuming their checklist. They took off without checking the status of the airplane's flaps. A flight emergency occurred shortly after takeoff because the flaps were in the wrong position. The crew mistakenly interpreted the problem as wind shear and crashed the plane (NTSB 1988).

It is essential to discover user interface design solutions that accommodate peoples' limitations and allow them to be interrupted safely.

2 BACKGROUND

Researchers have observed that interrupting people affects their behaviour. This is the basis of a classic effect from psychology called the Zeigarnik Effect (Van Bergen 1968). This effect was first identified in 1927 and describes a finding that people have selective memory relative to interruption, i.e., that people are able to recall the details of interrupted tasks better than the details of uninterrupted tasks. Results from many studies of the Zeigarnik Effect have produced somewhat inconsistent results. However, two findings seem universal: (1) interrupting people affects their behaviour, and (2) the interruption of people is a complicated process.

Work has also been done to compare different user interface design approaches for solving the problems associated with interrupting people. It was found that interaction design affects people's ability to successfully resume previously interrupted tasks. Two examples are interaction logic approaches for calculators (Kreifeldt & McCarthy 1981) and backtracking control for database access (Field 1987). Other work has been done to begin to identify which aspects of human interruption cause people to make mistakes (Czerwinski et al. 1991, Gillie & Broadbent 1989, Cellier & Eyrolle 1992).

A set of new interdisciplinary theory-based tools provides a general definition and taxonomy of human interruption (McFarlane 1997, McFarlane 1998). This taxonomy identifies eight major dimensions of the problem of human interruption that are exposed in the current literature. The third factor from the taxonomy, "Method of Coordination," is a critical aspect of human interruption that has not yet been directly investigated.

The "Method of Coordination" is the technique used to decide when to interrupt people. The taxonomy identifies the four known ways of coordinating user-interruption: (1) immediate, (2) negotiated, (3) mediated, and (4) scheduled. No comparison of the relative utility of these four design approaches exists in the current literature. Instead, previous research focuses only on the separate individual solutions without comparing the alternatives.

3 APPROACH

The approach of this paper is to discover the relative strengths and weaknesses of the four different "Method of Coordination" solutions. An experiment was performed that compared all four methods for coordinating user-interruption within a common context.

A fictitious example can illustrate the four user interface design approaches for determining when to interrupt people. Suppose that a person is performing two tasks concurrently: (1) indirectly driving a car by supervising a robotic driver, and (2) conversing with another human passenger in the car. Whenever the robotic driver must initiate an interaction with its human supervisor it must first interrupt them from their conversation. An "immediate" solution would have the robot interrupt the person at any time in a way that insists that the person immediately stop conversing and interact with the robotic driver. A "negotiated" solution would have the robot announce its need to interrupt its supervisor, and then support a negotiation with the user. This would give the person control over when to deal with the interruption. A "mediated" solution would have the robot not directly interrupt its supervisor, but instead contact the person's PDA (personal digital assistant) and request interaction with the person. The PDA would then determine when and how the robot would be allowed to interrupt the person. A "scheduled" solution would restrict the robot to interrupt its supervisor on a prearranged schedule such as once every 15 minutes.

Driving errors are more serious than conversational errors. Therefore a successful user interface design for a robotic driver would have to ensure people's performance on the supervised driving task regardless of the side effects on other activities. However, there is not enough design knowledge available in the current literature to say which "Method of Coordination" would be best for this problem, and different people have surprisingly different intuitive answers.

Prior studies have looked at topics related to each of the four "Methods of Coordination." One cost that has been identified for the "immediate" solution is that people experience a troublesome initial decrease in performance called automation deficit when they try to resume interrupted tasks (Ballas et al. 1992). A few authors have investigated ways to help users more easily resume interrupted tasks. For example: awareness of backgrounded tasks can be heightened with sonification; reminders can prepare people to resume interrupted tasks (Davies et al. 1989); and tools can be devised to help people quickly review interrupted tasks when resuming them (Field 1987).

The "negotiated" solution is an attempt to exploit people's natural ability to negotiate changes in their activities. Clark (1996) says that in normal human-human language usage people have four possible responses to interruption: (1) take-up with full compliance, (2) take-up with alteration, (3) decline, or (4) withdraw. Some papers have investigated usefulness of presenting interruption in ways that allow people to ignore them if they choose. Katz (1995) found that there are overhead costs related to negotiating interruptions, and that users sometimes prefer immediate interruption solutions when that overhead cost is not justified.

The “mediated” solution is an attractive but controversial approach. Delegating the interruption problem to a mediator begets a new task of supervising the mediator (Kirlik 1993). There are five main approaches for mediation: (1) predict people’s interruptibility (Miyata & Norman 1986); (2) implement intelligent user interfaces for supervision tasks; (3) automatically calculate users’ cognitive workload for dynamic task allocation; (4) apply human factors techniques for supervisory control; and (5) use cognitive models to guide interaction.

The “scheduled” solution is an attempt to give a degree of reliable expectation to a user about when they will be interrupted. In many ways, scheduling times for unexpected activities transforms interruptions into normal planned activities. Time management training has been found to have a positive effect on people’s ability to manage interruptions.

These four approaches for determining when to interrupt people have not been compared before. A conservative first question is, “Does it matter which coordination method is chosen as a solution to this user interface design problem?” If this question can be answered that, “Yes, it does matter,” then the relative strengths and weaknesses of the different solutions can be compared.

The main hypothesis of this paper is therefore that the particular method for coordinating user-interruption that is implemented in a user interface will affect user’s performance on interrupt-laden computer-based multi-tasks.

4 METHOD

Four different user interfaces were built for a common computer-based multi-task. These four interfaces are representative implementations of the four known solutions for “Method of Coordination.”

4.1 Subjects

Thirty six subjects were compensated for participating in this experiment (18 males and 18 females). Subjects completed an entrance questionnaire before beginning the experiment, and most reported that they had substantial experience with computer-based tasks. These 36 subjects, however, had highly diverse backgrounds on many other dimensions: age (mean 24.7, min. 18, max. 47), race, years of college education, amount of video game experience, level of typing skill, and self-reported vulnerability to the negative effects of interruption.

4.2 Design

A single-factor, within-subjects, Latin square design was chosen as an appropriate design for this experiment. Six treatments were devised: four experimental and two base case control treatments. Each of the four experimental treatments represented one of the four methods for coordinating interruption identified in the Taxonomy of Human Interruption (McFarlane 1997, McFarlane 1998).

Each treatment condition used a different version of a user interface (the independent variable). The computer-based multi-task was not varied between treatment conditions. Subjects’ performance (the dependent variable) on the multi-task was observed and recorded under the six treatment conditions.

All subjects received all six treatments. However, each subject was assigned to one of six groups that defined the counterbalanced ordering (digram-balanced) of the presentation of the six treatments. The presentation of each treatment was divided into two contiguous trials to avoid the confounding influences of fatigue and boredom. Male and female subjects were randomly assigned to groups, but with three males and three females in each group.

Each subject performed a total of 24 trials of the computer-based multi-task. Each trial was 4.5 minutes long, and there was a brief rest period with a masked screen between each session. Rest periods were a minimum of 25 seconds each. Therefore the total time for a subject to complete the experimental task was about 2 hours. For all subjects, the first 12 trials were practice (~1 hour) and the following 12 trials were experiment (~1 hour). Subjects received the same counterbalanced ordering of trials on practice trials as they did on experimental trials.

4.3 Multi-task

An interruption-laden computer-based multi-task was created as an appropriate testbed for this experiment. This multi-task itself represents an important contribution to the study of interruption. It was very carefully contrived to be both well controlled and appropriately complex.

The background literature identifies many subtle sources of possible influence on people’s behaviour during interruption. It was judged that a fully realistic multi-task would not allow for the control of irrelevant confounds, and, therefore, would not afford the collection of valid or reliable observations. The use of a realistic multi-task would degrade the internal validity of the experiment and muddle the results. An abstract multi-task had to be found that would isolate just those issues relevant to the hypothesis of this paper.

The interruption of people during human-computer interaction is a high-level interdisciplinary topic. The interdisciplinary background literature also shows that interruption is a complex process that involves many subtle low-level mechanisms of human cognition. These individual mechanisms, however, are not the focus of this experiment. It was judged that a simplistic and antiseptic task, typical of those used in studies about low-level topics of human cognition, would be inappropriate for this experiment. The use of such a simplistic multi-task would not allow the observation of the high-level effects, and would degrade the external validity of the experiment. A reasonably complex experimental multi-task had to be contrived to elicit people's behaviour at an appropriately high level.

It is possible to investigate the process of interruption at the level of user interface design without fully understanding the many subtle low-level cognitive mechanisms involved. In this experiment, the interesting, but irrelevant, smaller effects were ignored and isolated from the high-level effects by imposing pure noise into several aspects of the human-computer interaction. This intentional noise equalizes the many smaller effects across the different treatment conditions, and allows the direct observation of relevant high-level behaviour.

An abstract multi-task was chosen. It is a simplified model of a class of common real world multi-tasks. Examples of people performing multi-tasks from this class are 911 emergency dispatch operators and aviation radar operators. A Naval aviation task, for example, requires an operator to identify and maintain tracks of radar images as they appear and change over time. These identification and tracking subtasks do not occur at conveniently spaced times, but can often overlap. These overlapping subtasks can only be attended to one at a time, but the operator must also maintain a concurrent awareness of all subtasks. The operator must also be available for arbitrary interruptions by their leaders for direct requests of information.

The multi-task is a dual-task (a two-task multi-task) composed of a continuous game task and an intermittent matching task. The game task is modelled after a video game by Nintendo Corporation called "Fire" that was originally released in 1980 & 1981 as a version of the Nintendo Game & Watch product series. The matching task is modelled after the matching tasks used in experiments of the Stroop Effect. The dual-task is conceptually simple and yet can be very difficult for people to perform. The results of pilot studies confirmed that this dual-task elicits the kind of human errors associated with the interruption phenomenon.

The game task required subjects to control the movement of cartoon style stretcher bearers. The object was to direct the stretcher bearers to catch other game characters as they fell from a building. Each falling character had to be successfully caught and bounced three separate times at three different locations (total time to save each jumper was 16.9 sec.). If a character was missed at any of the three bounce points, then it was lost. See Fig. 1.

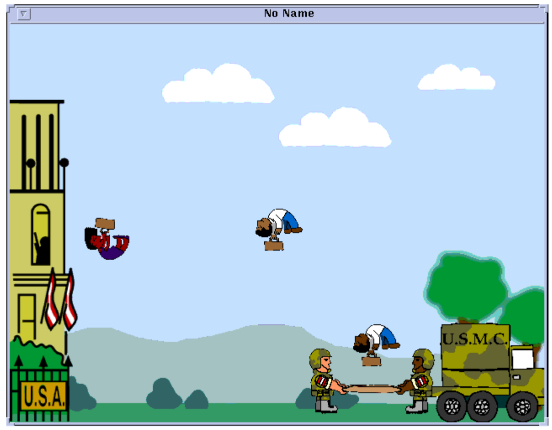

Figure 1. Game Task.

This game task is both continuous and discrete. It is a single continuously running game, however saving each individual jumping character was a completely independent discrete subtask. Errors made while performing one subtask do not automatically cause errors on other subtasks. This subtask composition allowed observations of peoples' behaviours to be easily broken down into discrete units. It was the interactions of several randomly intermixed subtasks that required dynamic problem solving. This arrangement allows the overall complexity of the game task to be conveniently manipulated.

The level of difficulty of the game had to be contrived so that it was complex enough to attack subjects' vulnerability to interruption, but simple enough not to cause subjects to despair of performing well. Through testing with pilot subjects, it was discovered that 59 game subtasks per trial was appropriate.

The second task of this dual-task was the interruption task. This task was an intermittent graphical matching task loosely based on the textual matching tasks reported in investigations of the Stroop effect. The interruption task required subjects to make matching decisions either based on colour or shape. Subjects were presented with a coloured shape at the top of the window, and instructed to choose one of the bottom two coloured shapes according to the matching rule displayed in the center. The matching rule instructed subjects to either "Match by shape" or "Match by color." See Fig. 2.

This matching task is conceptually simple, but deceptively difficult to perform. Pilot studies found that people were not able to automate this task through overlearning even after 2 and a half hours. Each individual matching task required a short focus of attention. It was discovered through pilot testing that 80 matching tasks per trial were appropriate.

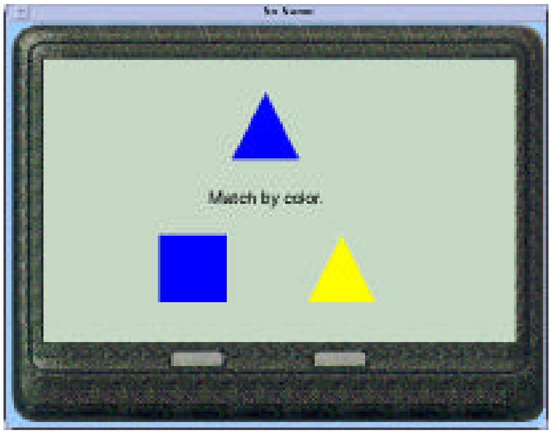

Figure 2. Matching Task.

Subtasks were independent. The graphic nature of the matching task corresponded with the graphic nature of the game task. Matching subtasks had to be done one at a time from a first-in-first-out queue, so there were no interruptions of interruptions.

Each of the 864 trials in this experiment (36 subjects * 24 trials each), provided an unpredictable multi-task. The schedules for game and matching subtasks were created with constrained randomization schemes that produced unique (unlearnable) schedules for each trial that were kept constant in the frequency domain.

4.4 Treatments

Subjects performed the multi-task with one-handed keyboard key presses on an isolated group of six keys of a common extended computer keyboard. See Fig. 3. The "Home" and "End" keys were only used in the "negotiated" treatment condition.

Figure 3. Keys used for performing the experiment.

Whenever a matching task was in the foreground, it appeared in the same window as the game task and totally obscured the view of the game. The game task continued to run without possibility for pause regardless of whether subjects could see it or not. In all treatments, except "Negotiated," once the multi-task was switched to the matching task subjects had to perform all queued matching tasks before they could resume the game task in progress. Whenever a user completed the last queued matching task, the multi-task switched back to the game task. The following are descriptions of the six treatment conditions.

Treatment 1: "Game Only" base case implemented the game task with no matching tasks.

Treatment 2: "Match Only" base case implemented the matching task with no game task.

Treatment 3: "Immediate" treatment condition presented matching tasks directly whenever they occurred regardless of the state of the game task.

Treatment 4: "Negotiated" treatment condition gave subjects control over when they would handle interruptions. When a matching task occurred, its arrival was immediately announced with a flash of a blank matching task for 150 msec and then the game task display resumed. Subjects had to decide when to begin the queued matching task. Subjects could use the "Home" and "End" keys at any time to bring the queued matching tasks to the foreground or push them to the background.

Treatment 5: "Mediated" treatment condition dynamically calculated a simple function of subjects' workload that measured how many jumping diplomats were on the screen. Interruptions were automatically held until workload metric was low.

Treatment 6: "Scheduled" treatment condition held all interruptions and only switched from the game task to the matching task on a prearranged schedule of once every 25 seconds.

4.5 Apparatus

All subjects performed the computer-based dual-task on a laptop computer (166MHz Pentium CPU, Windows95) using the built-in monitor (1024 X 768 pixels, 16 bit colour) and an external extended keyboard. The computer-based dual-task was displayed in a single 640X480 pixels window in the top left corner of the screen. The laptop was raised 4.75" above the tabletop in front of subjects to create a comfortable viewing angle. The experimental software was implemented with double-buffered frame animation running at 20 frames a second.

4.6 Procedure

Subjects participated one at a time and were required to pass a standard test for normal colour vision. They then received written instructions that described the multi-task and all treatment conditions. Subjects were able to refer to these instructions throughout the experiment. Each trial was preceded with an on-screen messages announcing which treatment condition would be next, and reminders to avoid fatigue, and that the game and matching tasks were equally important. Detailed interaction data were unobtrusively recorded by computer throughout the experiment.

5 RESULTS

The hypothesis asserts the existence of a causal relationship between coordination method used and subjects' performance on the multi-task. Six different measures of subjects' performance were chosen as appropriate for testing this hypothesis: (1) number of jumpers saved on the game task ("jumpers saved"); (2) number of switches between game task and matching task in both directions ("task switches"); (3) number of matches done wrong ("matched wrong"); (4) percent of matches done wrong of those attempted ("% m. wrong of done"); (5) number of matches not done ("matches not done"); and (6) average time from the scheduled onset of each matching task until it was actually completed or the trial timed out ("avg. match age").

Note that three performance measures, task switches, matches not done, and avg. match age, are not "traditional" experimental dependent variables because their value was not free to vary under subjects' direct control (except in the negotiated condition). These performance measures are appropriate here, however, because these limitations on subjects' performance are directly linked to the application of the different treatments and therefore illustrate how the four treatment conditions differentially affect subjects' behaviour.

The data from the practice trials were not included in these analyses. Fig. 4 contains bar charts for the six performance measures investigated here. These graphs show the data for 32.4 total hours of human performance on the experimental multi-task (36 subjects * 6 experimental treatments per subject * 2 trials per treatment * 4.5 minutes per trial). Note that these error bars reflect the total variance contained in the data and include some graphical distortion because of the inclusion of outliers.

Since this experiment is the first to compare the four methods for coordinating user interruption, it was important to attempt to maximize the validity of the results. Nonparametric statistical tests were employed. The decision to use nonparametric tests avoids potential confusions about the validity of parametric analyses. For example, it may be argued that the data do not have consistency of variance between conditions, because the different experiment conditions did not give subjects equivalent kinds of control over all kinds of multi-task performance. The nonparametric Friedman two-way analysis of variance by ranks test with correction for ties was selected (denoted by F_r), with its methods for post-hoc comparisons. This test is calculated on within-subject ranks, and is useful for analyzing within-subjects effects for interval level data. The Friedman test is therefore immune to the biasing influences of abnormal variance or outliers.

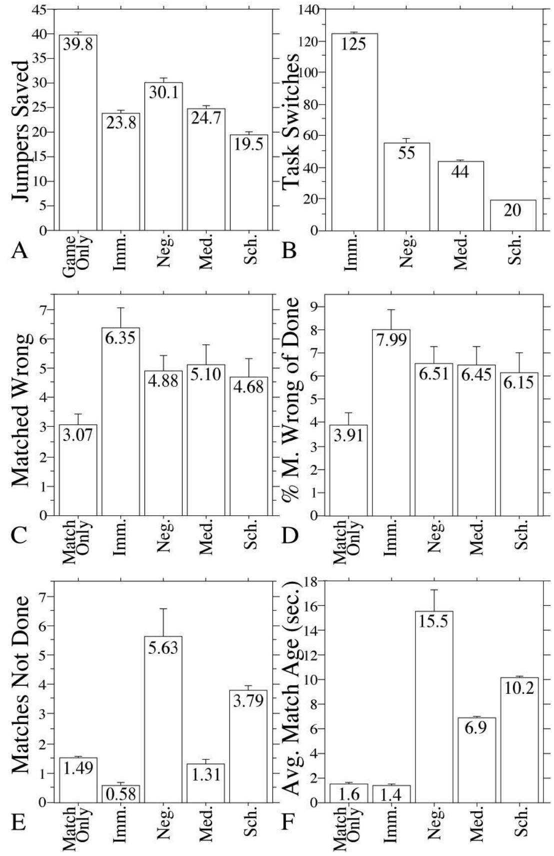

Figure 4. Average scores for 6 performance measures on experimental data. The bar charts show the mean with error bars that depict one standard error. There were 59 jumpers on the game task, and 80 matches on the matching task. Trials lasted 4.5 minutes. The abbreviations are: Imm. (immediate); Neg (negotiated); Med. (mediated); and Sch. (scheduled).

5.1 Overall Effects of Interruption

There must be an overall effect of interruption — otherwise a discussion of the differential effects of alternative methods for coordinating interruptions would not make sense. Two base case treatment conditions were included in this experiment so that this assertion could be validated before testing the main hypothesis. Table 1 summarizes the results of the Friedman test to determine whether there are any significant differences between the five relevant conditions for each measure of performance (the four treatment conditions and one base case). For comparison, F_r must be greater than 9.49 for $p < \alpha$ of .05

These results validate the basic assertion that being interrupted affects people's behaviour. The significance of these results permits post-hoc analyses. Table 2 summarizes the results of a comparison of

Table 1 — Comparison to Base Cases

Performance meas.	base case	F_r	p	$p<\alpha$
jumpers saved	game only	120.410	< .0001	YES
task switches	[no appropriate base case]			
matched wrong	match only	39.627	< .0001	YES
% m. wrong of done	match only	32.911	< .0001	YES
matches not done	match only	65.960	< .0001	YES
avg. match age	match only	117.956	< .0001	YES

individual conditions with the appropriate base cases using the Friedman test's post-hoc analysis methods. Each cell reports the results of significance tests with $\alpha = .05$. Fig. 4 can be used to determine the direction of significant pairs.

Table 2 — Post-Hoc Comparison to Base Cases

Performance measure	Base case	base & Imm.	base & Neg.	base & Med.	base & Sch.
jumpers saved	game only	YES	YES	YES	YES
task switches	[no appropriate base case]				
matched wrong	match only	YES	YES	YES	no
% m. wrong of done	match only	YES	YES	YES	YES
matches not done	match only	YES	no	no	YES
avg. match age	match only	no	YES	YES	YES

5.2 Effects of Different Interruption Coordination Methods

Do the different methods of coordinating interruption affect people differently? Table 3 summarizes the results of the Friedman test to determine whether there is any significant difference between the four experimental conditions for each measure of performance (base cases are not included). For comparison, F_r must be greater than 7.82 for $p < \alpha$ of .05.

Table 3 — Analysis of Experimental Conditions

performance measure	F_r	p	$p < \alpha$
jumpers saved	72.263	< .0001	YES
task switches	87.000	< .0001	YES
matched wrong	17.599	.0005	YES
% m. wrong of done	10.267	.0164	YES
matches not done	53.034	< .0001	YES
avg. match age	78.100	< .0001	YES

The data from all six performance measures support the main hypothesis with statistical significance. It is concluded that the particular method for coordinating interruptions of people implemented in user interfaces affects people's performance. These significant results permit post-hoc analyses. Table 4 summarizes the results of pairwise comparisons between the four experimental conditions using the Friedman test's post-hoc analysis methods. Each cell reports the results of significance tests with $\alpha = .05$. Fig. 4 can be used to determine the direction of significant pairs, with two exceptions. The "Neg. & Sch." pairs for Fig. 4 E & F have significant differences in opposite directions from the graphs. This is possible because the statistics were performed by ranks on a within-subjects effect, but the graphs portray the raw data that includes large individual difference effects (between-subjects) especially for the negotiated interruption solution.

Table 4 — Post-Hoc Analysis of Main Effect

Performance measure	Imm. & Neg.	Imm. & Med	Imm. & Sch.	Neg. & Med.	Neg. & Sch.	Med. & Sch.
jumpers saved	YES	no	YES	YES	YES	YES
task switches	YES	YES	YES	no	YES	YES
matched wrong	YES	YES	YES	no	no	no
% m. wrong of done	no	no	YES	no	no	no
matches not done	YES	no	YES	no	no	YES
avg. match age	YES	YES	YES	no	no	YES

5.3 Interpretation and Guidelines

Being interrupted affects people. Unfortunately, the above analyses reveal that there is no one "best" choice of method for coordinating interruptions for all kinds of human performance. There are instead, trade-offs between the four coordination methods and the different kinds of human performance. The following discussion does not include the base cases.

The negotiated coordination method was very successful in several ways. Negotiation resulted in the best performance on the jumpers saved measure, and it produced matched wrong performance as good as any other coordination method, and it produced many fewer total task switches than the immediate interruption solution. There was, however, a large price to pay in the completeness (matches not done) and promptness (avg. match age) of performing the matching subtasks.

The results from the percent matched wrong of done measure confirm that the negotiation solution did not pay a penalty in increased matching errors. The negotiated solution did not have relatively few matching errors simply because subjects performed fewer total matching subtasks on this condition. Subjects sometimes delayed handling interruptions quite long on the negotiated condition compared to other treatment conditions. This was not a classical speed-accuracy tradeoff. Subjects were not taking more time to make careful matching choices in the "negotiated" condition; they were only taking more time to procrastinate making matching choices.

The immediate coordination method produced nearly opposite performance of the negotiated solution. Immediate resulted in the best performance of any coordination method on the matches not done and avg. match age performance measures. However, these successes were gained at large costs in performance relative to the other coordination solutions on jumpers saved, matched wrong, and task switches performance measures.

The scheduled coordination method resulted in the best performance on task switching measure, but paid a heavy price on all other kinds of performance. The mediated coordination method produced mediocre levels of on all kinds of performance. It was neither the best nor worst for any performance measure.

No single method for coordinating user-interruption is a clear winner. Instead, each solution has its pros and cons relative to the different measures of human performance. These results alone do not have the external validity required for creating formal user interface design guidelines. These results do, however, have some value. They may support the following generalizations: (1) people perform very well when they can negotiate for the onset of interruptions, however giving people this kind of control also means that they may not handle interruptions in a timely way; (2) when people are forced to handle interruptions immediately, they get the interruption tasks done promptly but they make more mistakes and are less effective overall.

The following tentative user interface design guidelines are only informed speculations and ignore concerns about generalizability, however since no guidelines exist they may be somewhat useful. (1) Negotiated solution is best and scheduled solution is worst for accuracy on a continuous task; (2) Scheduled solution is best and immediate solution is worst for causing fewest task switches; (3) Immediate solution is worst for accuracy on an intermittent task; (4) Immediate or mediated solutions are best for completeness on an intermittent task; (5) Immediate solution is best for promptness on an intermittent task.

6 CONCLUSIONS

This paper presents the first empirical comparison of all four known approaches to the problem of coordinating user-interruption in HCI. This topic is an important factor for user interface design of systems that must interrupt their users. The results of this experiment suggest that the "best" solution is strictly relative to the particular kinds of human task performance judged critical to the success of the system.

7 ACKNOWLEDGMENTS

This work was funded by James Ballas at NRL, Michael Shneier at ONR, and John Sibert at George Washington U. Portions of this work were performed as part of a doctoral dissertation at GWU under the direction of John Sibert. Editing performed by: James Ballas, Astrid Schmidt-Nielsen, Susan McFarlane, Derek Brock, and Justin McCune.

8 REFERENCES

Ballas, J., Heitmeyer, C., & Pérez, M. (1992), Evaluating Two Aspects of Direct Manipulation in Advanced Cockpits, *in* P. Bauersfeld, J. Bennett & G. Lynch (eds.), *Proceedings of CHI'92: Human Factors in Computing Systems*, ACM Press, pp.127–34.

Cellier, J. & Eyrolle, H. (1992), "Interference Between Switched Tasks", *Ergonomics* 35(1), 25–36.

Clark, H. (1996), *Using Language*, Cambridge University Press.

Czerwinski, M., Chrisman, S., & Rudisill, M. (1991), Interruptions in Multitasking Situations: The Effects of Similarity and Warning, Technical Report JSC-24757, NASA Johnson Space Center, Houston, Texas.

Davies, S.P., Findlay, J.M. & Lambert, A.J. (1989), The Perception and Tracking of State Changes in Complex Systems, *in* G. Salvendy & M.J. Smith (eds.), *Designing and Using Human-Computer Interfaces and Knowledge Based Systems*, Elsevier Science, pp. 510-17.

Field, G. (1987), "Experimentus Interruptus", *ACM SIGCHI Bulletin* 19(2), 42–6.

Gillie, T. & Broadbent, D. (1989), "What Makes Interruptions Disruptive? A Study of Length, Similarity, and Complexity", *Psychological Research* 50(4), 243–50.

Katz, R. (1995), "Automatic Versus User-Controlled Methods of Briefly Interrupting Telephone Calls", *Human Factors* 37(2), 321–34.

Kirlik, A. (1993), "Modeling Strategic Behavior in Human-Automation Interaction - Why an Aid Can (and Should) Go Unused", *Human Factors* 35(2), 221–42.

Kreifeldt, J.G. & McCarthey, M.E. (1981), Interruption as a Test of the User-Computer Interface, *in Proceedings of the 17th Annual Conference on Manual Control*, JPL Pub. 81-95, pp. 655-67.

McFarlane, D. (1997), Interruption of People in Human– Computer Interaction: A General Unifying Definition of Human Interruption and Taxonomy, Technical Report NRL/FR/5510-97-9870, US Naval Research Lab, Washington, DC.

McFarlane, D. (1998), Interruption of People in Human–Computer Interaction, PhD thesis, George Washington University, USA.

Miyata, Y. & Norman, D. (1986), Psychological Issues in Support of Multiple Activities, *in* D. A. Norman & S. W. Draper (eds.), *User Centered Systems Design: New Perspectives on Human–Computer Interaction*, Lawrence Erlbaum Associates, pp.265–84.

NTSB (1988), Aircraft Accident Report, Technical Report NTSB-AAR-88-05, National Transportation Safety Board.

Van Bergen, A. (1968), *Task Interruption*, North-Holland Publishing Company.