language. In this respect it is similar to Trillium[2]. However we are working in a complex engineering environment and need to interface the user part of the tool to other parts of the system and we also need to document the system formally. The designer can use a design notation to specify the users dialogue with the system. The design is based on the Model-View-Controller paradigm of the Smalltalk-80 system and is modified to a model-view-dialogue object notation. Where the model is the computational part of the system, the view is the the physical layout of user interface objects and the dialogue is the conceptual dialogue the user carries on with the system in terms of choices and actions. This dialogue is independent of the style or layout of the system. The designer can produce a text description of the dialogue and then construct a layout of the view of the dialogue using the prototyping tool.

The design notation should also allow the dialogue to be analysed formally and ensures a consistent notation is used through the different stages of system development. Consistency is particularly important when we come to evaluate the system as it would be useful if we could express the evaluation results in terms of the user interface objects in our notation and feed them back for the redesign of the interface.

4. Related Reading

1 B Shneiderman, "Direct manipulation: A step beyond programming languages", IEEE Computer, 16 pp.57-69 (1983).

2 D A Henderson, "The Trillium user interface environment", CHI '86 Proceedings, pp.221-227 (1986).

3 ECLIPSE/UI/AI/FS, The Applications Interface 1.1, Alvey ECLIPSE document.

4 ECLIPSE/UI/FDL/FS, The Format Description Language (FDL) 1.2, Alvey ECLIPSE document.

5 J K Jacob, "Using formal specifications in the design of a human-computer interface", CACM 26(4) pp.259-264 ACM (April 1983).

6 AI Wasserman, P A Pircher, D T Shewmake, M L Kerstein, "Developing interactive information systems with the User Software Engineering Methodology", IEEE Software Engineering, SE-12(2) pp.326-345 IEEE, (Feb. 1986).

7 P Reisner, "Formal grammar and human factors design of an interactive computer system", IEEE Software Engineering, SE-7(2) pp.229-240 IEEE, (March 1981).

8 E Edmonds, S Guest, "The SYNICS2 User interface manager", Human-Computer Interaction -Interact '84 (B Shackel ed.) pp.375-378 (1985).

9 T P Moran, "The command language grammar: a representation for the user interface of interactive systems", Int.J.Man-Machine Studies 15(1) pp.3-50, Academic Press (July 1981).

10 B J Cox, "Object Oriented Programming", Addison Wesley (1986).

11 J M Carroll, M B Rosson, "Usability specifications as a tool in iterative development" IBM Research report RC 10437, IBM Watson Research Center (March 1984).

12 B Shackel, "Human Factors and usability", ACM German Chapter Conference - Software Ergonomie '85, Stuggart University (September 1985).

David England

David England is a Research Associate at the Department of Computing, University of Lancaster, UK. His first degree is in Mechanical Engineering from the University of Salford, his masters degree is in Information Technology Systems from the University of Strathclyde. He first caught the CHI bug through involvement with a Adaptive Dialogue project during his masters degree. His current thesis work forms part of the larger Alvey Eclipse software engineering project.

EXPERIMENTUS INTERRUPTUS

GRAEME E. FIELD

Abstract

Increased use of information databases has not been reflected in research in the field of HCI. Experiments that have been conducted often relate to abstract tasks rather than to tasks that users will actually perform. Twelve subjects completed realistic tasks in a nested hierarchical experimental design that used two conditions, Selective Retreat or Restricted Retreat, with two task types, Simple or Complex, that were either Interrupted or Not Interrupted. Results indicated that: the provision of a trace of selections significantly enhanced navigation;

active search time was very variable; and Interruptions *do* affect users behaviour in the post interruption period. The small number and type of subject in this pilot study necessitates a cautious interpretation of the results. However, an interesting area of research with direct application to everyday use of computer systems has been broached. Some of the implications of this study and suggestions for future research are discussed.

Over recent years there has been a marked increase in the public use of information databases, such as Prestel or Telidon, and in the more specialised, but often simpler, banking and shopping systems. These systems often have menu-driven interfaces, which are considered to be an appropriate interface for novice users (Shneiderman, 1987). The reasonable assumption that HCI researchers would wish to ensure that these systems are; easy to use, rewarding, enjoyable and efficient, would tend to suggest a wide and increasing pool of experimental knowledge of user's behaviour with such systems. Unfortunately, this does not appear to be the case.

One area of menu-based system research that has received some attention, is the so-called 'depth-breadth' tradeoff (see eg Miller, 1981; Snowberry, Sisson & Parkinson; 1983). A major problem with this type of study however, is the artificiality of the experimental situation, and the corresponding lack of generalisability to the use of real information systems. Typically subjects have been asked to find a target word using menus of varying breadth (number of options) and depth (number of levels). In practice however, users of information databases are looking for rather more complex information than a target word. They often require data pertinent to a series of problems such as; train departure times, what's on at the movies and the names and details of restaurants in a particular area of town. An experimental situation that addresses some of these problems of artificiality has been described by Apperley and Field (1984). They devised an information database for a fictitious city called Carlton, and asked subjects to navigate this database to find the answers to a series of questions that users of a real database might wish to solve. Subsequently, Field and Apperley (Note 1) found that subjects who used an enhanced interface, with a Selective Retreat facility, were advantaged over subjects who used a Standard menu system. Amongst other results, the Selective Retreat group accessed significantly fewer screens but used the same amount of time to solve the five problems.

A common feature of many user's interaction with information systems is the unasked for, but never-the-less inevitable, occurence of interruptions. It has never been shown experimentally whether the interruption of a task affects a user's post interruption activity. It is conceivable that users simply resume and complete their task as if the interruption had not occured.

The principle questions asked in this pilot study relate to the effects, if any, of the interruption of users as they attempt to solve Simple or Complex tasks, using systems that either have or do not have a Selective Retreat facility. Another consideration is the possibility of using Interruptions as a dependent measure in the comparison of two systems.

Method

Subjects The subjects were eleven male and one female volunteers, mean age 24.6 years,currently employed in the Department of Electrical Engineering at Imperial College, London. All subjects had had extensive experience with computers.

Apparatus The experiments were conducted using a BBC Model B micro-computer with a Prestel-like information database of a fictitious city called Carlton. All subjects had access to the same information although there were two possible system displays. One display featured a Selective Retreat (SR) facility (Apperley & Field, 1984), which allowed subjects to retreat to any previously selected screen, within a sequence starting from the Main Menu. The other display, the Restricted Retreat (RR) condition, permitted retreats to the previously viewed screen, by keying 12, or to the Main Menu , by keying 11. All keystrokes, the time at which the keystroke was made, and the number of the screen that was accessed, were recorded automatically by the system.

Procedure Subjects read a sheet explaining the experiment. The Experimenter answered any questions and then stated: "There are two points I'd like to emphasise; firstly, would you please answer the questions in order, and, secondly, *all* questions can be answered using the database, there are *no* 'trick' questions without answers."

Design The experimental design was a nested hierarchical, random groups design. Each subject performed in the Selective Retreat or Restricted Retreat condition, and completed 5 Simple Tasks of which 2 were Interrupted and 3 were Not Interrupted and 5 Complex Tasks of which 2 were Interrupted and 3 were Not Interrupted.

Questions A series of ten questions, each of which could be answered using the Carlton database, was devised. Some questions were designated

'Complex' and others 'Simple' on the basis of the perceived number of initial entry responses. Thus Question 4 (Simple) "How long is a session of the film 'Breakout'?" was considered by the Experimenter to have only one possible entry selection from the Main Menu, 'Leisure', whilst Question 7 (Complex) "What is the telephone number of the Carlton Estate Agents office?" was considered to have three possible entries from the Main Menu, 'Business Information', 'Buying and Selling' or 'Accomodation'.

Interruptions The Interruptions were presented as a screen that asked the subject to perform a particular task, such as completing a numeric sequence or looking up the title of a book in a group of texts alongside the terminal. They occured after the subjects had accessed particular screens, hence the neccessity to emphasise the completion of questions in order. The Carlton System software included an 'Interruption Vector Sequence', which permitted the interruption of a particular screen, only after a set sequence of screens had been previously accessed. This ensured that subjects were not 'accidently' interrupted during their database traversal. When subjects selected a screen that was to be interrupted they saw the entire screen for 0.1 second before it was overwritten by the Interruption screen. After subjects had completed the instructions of their interruptions, they were told to press a particular key, and were returned to the screen that they had last selected.

Analysis Small subject numbers (N=12) precluded the use of parametric multivariate analysis of variance as an overall test of significance. Analyses were primarily conducted using the non-parametric Mann-Witney U test, although the t - test was used when the number of data points generated reached twenty or more.

Dependent Variables The dependent variables of this study were defined as follows. Active Search Time (AST) is the amount of time that each subject is actively engaged in searching the database. Thus time spent dealing with interruptions, time to write down answers and any time used attempting to make invalid selections was subtracted from AST. Number of Selections (NS) is the total number of valid screen selections made by subjects during their database search. Interruption screens were not included in this total. Access of Different Screens (ADS) is the total number of different screens that were accessed by the subjects. Number of Selections per Question (NSQ) is the total number of selections required to reach the target screen for a question. Post Interruption Pause (PIP) is the period of time from the rewriting of the interrupted screen till the subject resumes active search of the database. Time to Target (TTG) is the amount of time required to reach the target screen for a question following an Interruption. Number of Retreats (NRT) is the total number of times that a subject selected a higher level in the menu system in the post Interruption period. Number of Screen Accesses to Target (NSAT) is the total number of selections required in the post Interruption period to attain the target screen and complete the task.

Results

Selective Retreat (SR) vs Restricted Retreat (RR)

Subjects in the SR condition made significantly fewer selections in finding the answers to the questions than subjects in the RR condition (X_{SR}=72.7 sel, X_{RR}=88.7 sel, U=0, p=0.002), and accessed significantly fewer different screens (X_{SR}=40 scr, X_{RR}=43.7 scr, U=4.5, p=0.046). There was no significant difference in AST between the two groups, although the difference in AST was in the expected direction and approached significance (X_{SR}=41.4 sec, X_{RR}=55.5 sec, t =1.75, df=118, p=0.083). The AST was very variable with standard deviations of the same order as the means (s_{SR}=33.5 sec, s_{RR}=52.6 sec).

Simple vs Complex Tasks

Subjects made significantly fewer selections per question to find the answers to the Simple Tasks (Sim) than to the Complex Tasks (Com) (X_{Sim}=7sel, X_{Com}=9.2 sel, t =2.3, df=118, p=0.023). There was no significant difference in AST for finding answers to Simple or Complex Tasks, although the difference in AST was in the expected direction (X_{Sim}=43.9 sec, X_{Com}=53 sec, t =1.12, df=118, ns).

Post Interrupt Activity

There was no significant difference in the PIP between subjects in the SR and RR conditions regardless of task type (X_{SR}=4.7 sec, X_{RR}=6.5 sec, t =1.43, df=41, ns) or after the Interruption of Simple or Complex Tasks across conditions (X_{Sim}=5.2 sec, X_{Com}=5.3sec, t =0.05, df=42, ns).

There was no significant difference in the TTG, following an Interruption, between the SR and RR conditions regardless of task type (X_{SR}=17.1 sec, X_{RR}=29.4 sec, t =1.25, df=41, ns) or between the Simple and Complex Tasks across conditions (X_{Sim}=21.7 sec, X_{Com}=25.6 sec, t =0.38, df=41, ns). However, there was a significant difference in the NRT to higher level menus between the SR and RR conditions, regardless of task type (X_{SR}=0.5

ret, X_{RR}=6.2 ret, U=0, p=0.01) and between the Simple and Complex Tasks across conditions (X_{Sim}=0.14 ret, X_{Com}=1.25 ret, U=2, p=0.013). There was also a significant difference in the NSAT in the post interruption period, between the SR and RR conditions regardless of task type (X_{SR}=8.7 acc, X_{RR}=19.2 acc, U=0, p=0.003). The NSAT in the post interruption period, between the Simple and Complex Tasks across conditions was in the expected direction and approached significance (X_{Sim}=4.2 acc, X_{Com}=6.9 acc, U=3.5, p=0.07). The four Interruptions had mean lengths (standard deviations) of 79.7(33), 91.7(84), 55(43) and 60(26) seconds respectively. A Kruskell-Wallis one way analysis of variance by ranks showed that there was no significant difference between the lenghts of the Interruptions (H = 0.45, ns).

Discussion

The results of this pilot study indicate that menu-driven database navigation is facilitated by the provision of a Selective Retreat showing users a record of their selections. Users with the Selective Retreat facility made significantly fewer selections, accessed significantly fewer screens and showed a smaller active search time in answering the ten questions than subjects without the Selective Retreat facility. This is in general agreement with the findings of Field and Apperley (Note 1).

The experimenter defined Simple and Complex Tasks were shown to be different, as had been hoped. Complex Tasks required significantly more screen accesses and a longer time to solve than the Simple Tasks.

An analysis of the Post Interruption Activity showed that Interruptions *do* have a significant effect on users post interruption activity. There were significantly fewer retreats towards the Main Menu for the Selective Retreat condition regardless of question type, and in the performance of Simple Tasks for both conditions. Additionally the number of screen accesses to target following Interruption was significantly less for the Selective Retreat condition across task types and for Simple Tasks for both conditions. This interesting result provides experimental evidence that interruptions do actually effect task completion. The realistic tasks of this experiment were still relatively straight forward in comparison with many activities performed on computer systems. It might be supposed that interruptions will have an even greater disruptive effect in these situations.

It is possible that the effects of an Interruption could be used as a dependent variable in the comparative evaluation of a number of systems. This might be particularly useful for systems designed for use in relatively busy communication situations, such as offices. This experiment shows that the provision of a Selective Retreat facility significantly lessens the effects of Interruptions.

It would appear that subjects who used the Selective Retreat facility were more 'sure' of their position within the database, particularly since they used significantly fewer retreats towards the Main Menu in the post interruption period. The Selective Retreat facility provides some prompts for short term memory, and this may be the mechanism that leads to facilitation in the Selective Retreat condition. However, it is also likely that the provision of the Selective Retreat helps subjects to build a 'cognitve map' of the system and increases their contextual knowledge. It is probable that this knowledge of the structure of the system enables them to access the information more effectively. Future work will be directed towards attempting to elucidate the cognitive mechanism that leads to the improved performance in the Selective Retreat condition.

The Post Interruption Pause, that is, the time taken to resume search activity following interruption, was not significantly different for the Selective Retreat and Restricted Retreat conditions regardless of task type, or for the Simple and Complex Tasks across conditions. This result is a little surprising. It might be assumed that the greater the complexity of the task being interrupted, the greater the time that would be required to re-orient oneself before continuing the task. Although the tasks were shown to be significantly different, it is likely that the Complex Tasks were not sufficiently more difficult than the Simple Tasks to show any major effects on post interruption pause. An adequate assessment of task complexity, prior to experimentation, would provide a better basis for delineation of the effects of task complexity.

Although the time required to reach the target screen, following an Interruption, was not significantly different for the Selective Retreat or Restricted Retreat conditions regardless of task type, or for the Simple or Complex Tasks across conditions, the mean times to target were less for the Selective Retreat condition and for the Simple Tasks. The great variability in these times, and the small number of subjects in this study probably accounts for this non-significant result. A similar study using a greater number of subjects is planned, and this should help to shed some light on the question of active search time in the post interruption period.

There is, as yet, no adequate means of describing the complex patterns of user navigation of information systems. This is an important area of research, since it is the description of a user's navigation that will allow the comparison and summation which is necessary to build predictive models of user navigation. The data capture provided by this experiment may enable us to devise models of navigational style or to derive graphical, algebraic or other mathematical descriptions of user navigation. Work is in progress that will be directed towards this end.

Reference Note

Field, G.E. and Apperley, M.D. (in prep). Context and selective refraction in hierarchical menu structures: An experimental study.

References

Apperley, M.D. and Field, G.E. (1984). A comparative evaluation of menu-driven interface systems. In Interact '84, Proceedings of the First International Conference on Human Computer Interaction, IFIP, London.
Miller, D.P. (1981). The depth/breadth tradeoff in hierarchical computer menus. In Proceedings of the 25th Annual Meeting of the Human Factors Society, 296-300.
Shneiderman, B. (1987). Designing the user interface: Strategies for effective Human-Computer Interaction. Addison Wesley Publishing Company, Reading, MA.
Snowberry, K., Parkinson, S.R. and Sissons, N. (1983). Computer display menus. Ergonomics, 26, 699-712.

AN ADAPTIVE GRAPHICS ANALYZER AS A PREFERENCE-ORIENTED INTERFACE

MAREK HOLYNSKI, ROBERT GARNEAU, MICHAEL LU

Computer graphics presentations should be tailored to the user's needs and on-the-spot applications. The manner in which an architect perceives a picture of a building could be dramatically different from the way an artist views the same structure. An architect must be a conceptualizer who observes the basic elements of design, historical style, materials and engineering techniques. In observing a structure an architect must think about functionality and plausibility of design, while an artist may care little about anything other than appearance. Computer images, therefore, should reflect visual preference, knowledge, application and perception constraints of a particular user or group of users.

A computer system that can filter out all the unwanted or ill conceived images by interacting with the user to select the optimal image or optimal subset of appropriate images could meet this challenge. We have begun to address this problem by applying machine learning techniques to a graphics interface which has resulted in a system we call an Adaptive Graphics Analyzer (AGA).

1. MACHINE LEARNING FOR DETERMINING VISUAL PREFERENCES

In order for the system to learn about user preference, it must have the ability to extract concrete and measurable statistics from a given visual image. The system must be able to assign specific image variable values to each picture generated and realize which values or set of values most influence its appearance. Through a series of images tested with large groups of subjects and analyzed with a rule acquisition package we have determined a set of relevant image variables.

There has been a great deal of preliminary research done to select image variables. Lewis and Holynski (1983,85) did early work in determining the appropriate image variables to construct a "filter" for selecting potentially good images from the population of possible images. They explored three different methods: fractals, structural randomness of primitive elements (the technique used in this research), and grid systems to establish and test image variables. They found that visual form could in fact be quantified using variables such as complexity, regularity, color and order (symmetry).

In a more recent work Holynski, Garneau and Lewis (1986) generated the stimuli as abstract display matrices comprised of controlled combinations of thirty-six primitive elements. The stimuli illustrated the four previously used variables along with three additional variables: balance (image equilibrium), variety (number of different primitive elements that make up a single image) and busyness (amount of display image that contains primitive elements). Two hundred subjects from three different student groups evaluated the stimuli. Standard regression analysis was used to discover which variables were appropriate predictors for user preference. Using an interval scale we found that complexity, regularity, symmetry and busyness had a positive significant effect of approximately the same magnitude on user preference.

These findings were consistent with earlier findings in terms of relationship between regularity and complexity with preference. Although useful, conventional statistical packages do not immediately show how one variable interacts or depends on another. In order to more carefully examine the interrelationship among all the variables, we used