Int. J. Man-Machine Studies (1986) **25**, 73-88

A comparison of menu selection techniques: touch panel, mouse and keyboard

JOHN KARAT, JAMES E. McDONALD† AND MATT ANDERSON‡
IBM Austin, †New Mexico State University and New Mexico State University

(Received 8 July 1985 and in revised form 6 January 1986)

Two studies were conducted to test user performance and attitudes for three types of selection devices used in computer systems. The techniques examined included on-screen direct pointing (touch panel), off-screen pointer manipulation (mouse), and typed identification (keyboard). Both experiments tested subjects on target selection practice tasks, and in typical computer applications using menu selection and keyboard typing. The first experiment examined the performance and preferences of 24 subjects. The second experiment used 48 subjects divided into two typing skill groups and into male-female categories. The studies showed performance advantages for on-screen touch panel entry. Preference ratings for the touch panel and keyboard devices depended on the type of task being performed, while the mouse was always the least preferred device. Differences between this result and those reporting an advantage of mouse selection are discussed.

1. Introduction

The process of doing work using a computer system can be broken down into three steps. The user must understand the goal of the work to be accomplished, formulate a task solution (a series of steps to accomplish a work goal), and carry out the plan. Whether the user is solving old problems with a new system, or is experienced with the computer as a problem-solving tool, the final stages require relating the steps of the solution to the system in a "language" that the system understands.

Given that humans have limited information processing resources, one goal for system designers is to make the process of relating the user's goal to the system as easy and natural as possible. To this end interfaces using menus have been developed from which alternatives are selected, rather than recalled from memory. While menu selection on a computer has traditionally involved typing on a keyboard, several recent workstations have designed interfaces utilizing "more natural" selection mechanisms (e.g. pointing and voice recognition).

If we examine the process of making selections from everyday lists, we find several common selection mechanisms. These include verbalizing the selection ("I would like a cheeseburger and fries."), pointing with the hand to a selection on a list ("Give me this one.") or possibly marking a selection with a pencil. It is interesting to note that selection in a restaurant reverts to pointing under many circumstances (e.g. unpronounceable entrees). One argument for utilizing devices such as touch panels and joysticks in computer systems is that they provide a mechanism which allows the user to point in a relatively natural fashion (i.e. with a mechanism that tracks hand movement).

0020-7373/86/070073+16$03.00/0

The purpose of the current study was to examine three ways of selecting items in a menu-driven computer system. The methods considered were selection by pointing and touching an on-screen item with the hand, selection by moving an on-screen pointer with an off-screen device, and selection by typing an identifier associated with a given display item. The primary motivation for examining these methods is that they can be thought of as representing different degrees of "naturalness" of selection. We were interested in examining the role of such selection techniques in a computer dialogue.

There are likely to be tradeoffs in the consideration of selection mechanisms. Though the mechanics of typing have been studied extensively (e.g. Cooper, 1983), relatively little is known about the limitations of other selection mechanisms in general computer dialogue use. It seems intuitive that some pointing devices are more natural than keyboards for certain applications (e.g. graphics), but little is known about the merits of these devices in applications which currently require the use of a keyboard for input (e.g. text tasks). Selection by pointing has been evaluated in a number of recent studies which carry out movement of an on-screen pointer through manipulation of off-screen devices, such as keyboard cursor keys, joysticks, tracker balls and mice (Card, English & Burr, 1978). These studies generally find the mouse to be the best off-screen pointer manipulation device: thus the mouse was used in this study.

The current experiment was designed to address some points not considered in reported work. First, the tasks commonly used in pointer studies were target selection tasks (tasks in which a subject points to a target randomly positioned by the system on a VDT). The work reported here uses a menu system where users were asked to solve problems using the system. We would like to replicate the results of target experiments in a more realistic computer dialogue. Next, most previous studies have not included touch panels. While there is some recent work which has included evaluations of touch screens (Whitfield, Ball & Bird, 1983), this work is again confined to target selection tasks. Additionally, previous studies have failed to evaluate the most common form of menu selection, typing an identifier associated with an item. Finally, we included a situation in which menu selection was only part of the dialogue. Some of our experimental tasks involved typing information on the system keyboard.

A SIMPLIFIED THEORY OF SELECTION

While the structure, the terminology used, and a host of other factors undoubtedly affect the use of a menu system, the current work focuses on the selection process. This is done to provide a manageable problem domain for applied cognitive psychology. We treat the selection process as an event consisting of several components. This technique has been found useful in modeling human cognitive behavior (Newell & Simon, 1972; Card, Moran & Newell, 1983).

There are many well-learned "natural" selection procedures. If someone makes a menu selection in normal human interaction, it is not generally necessary to instruct them in how to do it (although someone did at one time). For human-computer interactions, a menu can help the user to decide *what* to do, and *how* to do it. The user may have to be instructed in a "new" selection method (i.e. telling them that the machine can't hear them or doesn't know where their finger is pointing). For a new user there may be cognitive costs associated with learning the new method. If the selection method is not a simple extension of well-learned procedures, the user must devote attention to figuring out how to convey the selection to the system. On the other hand, studies have shown that well-learned methods tend to be performed very rapidly and with little cognitive effort (e.g. Shiffrin & Dumais, 1981).

With a "natural" menu selection the user:

Carries out the decision process, then indicates the choice.

For a "new" selection process, the user must:

Carry out the decision process, figure out how to indicate the choice, then indicate the choice.

With experience, the additional step required to figure out how to indicate the choice becomes well learned. That is, the procedure for indicating a choice becomes automatic, and the individual does not have to devote additional resources to it. The exact mechanisms for procedure automation are not considered here (see Shiffrin & Schneider, 1977, for a detailed discussion): only the distinction between new and relatively well-known procedures is needed. Even though people do learn new procedures, a desirable system design goal is to limit device-specific knowledge (knowledge required to operate specific system components). This makes learning through the integration of system procedures into existing knowledge easier by making computer dialogue consistent with "natural" dialogues.

This objective might suggest an advantage for direct pointing, but the issue is still complex. Card *et al.* (1983) have shown that formal task analysis can be used to aid in making design decisions. Their task analysis breaks behavior down into a number of cognitive cycles and motor components, and can be used to predict operator performance. If we assume users are familiar with all three selection methods, it is possible to carry out such an analysis and reduce selection for each to a single thought cycle and a single physical movement. Though the distances moved by the hand may vary for the three methods, the effect on task time would be small, and the analysis would suggest little difference between the methods. However, this analysis does not seem to capture all elements needed to insure "ease of use" for a range of users.

While a keystroke-level analysis (Card *et al.*, 1983) might work for well-learned tasks, it is not clear that it can be extended to more complex tasks, or even to simple tasks which are less practised. One problem with using such a model in a cognitively complex domain is estimating with any certainty the amount of cognitive processing or the number of cognitive cycles required for an activity. Thought cycles are hypothesized by Card *et al.* (1983) as corresponding to a processing time of about 100 msec. While it might be possible to agree on the number of cognitive steps of this magnitude required for fairly simple well-learned tasks, cognitive science has not developed sufficiently to provide detailed models for much of normal cognitive behavior. It is not clear that pointing with a mouse requires the same amount of cognitive processing as finger pointing, or if switching input devices increases processing demands. We could provide estimates of cognitive load for switching between input devices, but we would certainly want to verify empirically such estimates before accepting them as valid. Since a large percentage of task time might be spent in simply making selections, it is important to carefully consider the tradeoffs involved.

For this study two experiments were run. The experiments differed primarily in the amount of practice provided, and in the classification of the subjects. The second experiment was included to address issues which arose in the analysis of data from the first study.

Experiment 1

METHOD

Subjects

Twenty-six subjects (25 females and one male) were obtained from a temporary employment agency and ranged in age from 18 to 57 years with self-reported typing rates ranging from 30 to over 100 words per min. Forty-two per cent of the subjects had obtained college degrees, 50% had some college but no degree, and the remaining 8% had completed high school. Most of the subjects (83%) reported some word-processing experience, but only four had any programming experience. Two subjects failed to complete the experimental tasks and were excluded from the analyses. All reported analyses are based on data obtained from the 24 subjects who successfully completed the experiment.

Apparatus and task environments

An IBM Personal Computer (PC) served as the experimental controller. Three independent input devices were attached to the PC (a touch panel, a mouse, and a keyboard) such that each could be used to select options from menus presented on the "task display" (a 13-in IBM Color Monitor). The input devices were an Elographics analog-membrane touch panel, a Mouse Systems optical mouse, and a standard IBM PC keyboard. The "instruction display" (a 12-in IBM Monochrome Monitor) was used to present instructions, application descriptions, and individual menu tasks to the subjects. Subjects sat facing the two displays with the keyboard located between the subject and the displays throughout the experiment. The touch panel was attached to the front of the task display during all input device conditions. The mouse was located on a metal pad next to the subject's preferred-hand side of the keyboard during the mouse-device conditions, and was moved away for keyboard and touch panel conditions. The keyboard was used as a menu selection device for the keyboard-device condition, and for entering text into "fields" during all three conditions.

Two computer-based systems were developed for the experiment and simulated on the PC. The two menu task environments (applications) simulated a computer telephone aid and a personal appointment calendar. Both of the applications appeared to the subject as hierarchically organized menus from which the subjects selected options. Additionally, half of the tasks that the subjects were asked to perform required the entry of typed information into the system (such as names, comments or telephone numbers). Thus, subjects performed tasks (e.g. calling someone from a directory or changing an appointment) which required selecting options with a device and typing into selected fields from the keyboard. The system responded to subject input by appearing to perform the specified actions (actual telephone connections were not made nor did an appointment data base exist).

The two applications differed in structure and complexity. The depth (number of selections needed to reach a given action) and breadth (number of options available from a given state) of the two systems' menu hierarchies were different. The number of menu selections for the required tasks ranged from one to eight for the telephone system tasks, and from three to eleven for the calendar tasks. The calendar tasks included: reviewing daily schedules, checking past and future appointments, scheduling new appointments, and modifying existing appointments. The telephone tasks involved calling persons listed in various directories and modifying those directories by changing the contents of entries, deleting entries, or creating new entries.

Experimental design

All subjects used all three input devices to perform tasks within both applications. For the first experiment subjects were classified into three device order groups. The order in which the subjects used the input devices was based on a Latin-Square, with three device orderings (keyboard/touch panel/mouse; touch panel/mouse/keyboard; and mouse/keyboard/touch panel). There were several within subjects variables which resulted in a 3 (device) by 2 (application) by 2 (task type) within-subjects design. Application order (telephone/calendar or calendar/telephone) was counterbalanced across subjects. This resulted in a total of six groups (within each of the three device order groups half of the subjects used each application order) to which the subjects were assigned in a pseudo-random fashion. Within each condition, the order of presentation of individual tasks and applications remained constant across devices.

A set of practice trials preceded each set of tasks. During an experimental session, subjects used the first input device to perform 25 practice trials, 10 tasks in the first application, 25 additional practice trials, and then performed 10 tasks in the second application. This procedure was repeated for the two remaining input devices. In total, subjects completed two sets of practice trials (one set before each application) and two sets of application tasks (10 tasks in each task environment) for each of three devices.

Procedure

Subjects received practice with the appropriate input device prior to performing tasks within an application. Practice consisted of selecting a target "button" which appeared in random locations on the task display. Subjects were instructed to select the target as rapidly as possible while keeping selection errors to a minimum. Targets were outlined boxes (14×13 mm) with a single letter inside. On each practice trial the screen would go blank, then the target box would appear. The position of the target box and the box label ("a" through "z") were randomly selected on each trial.

Selection with the touch panel involved touching anywhere inside the target with the subject's finger. Selection with the keyboard involved typing the letter which appeared inside the target box. Selection with the mouse involved positioning the screen pointer inside the target and pressing any button on the mouse. Movements of the mouse resulted in equal movements (1 : 1) of the pointer on the task display. Target selection time was measured from the appearance of the target on the display to the selection of the target with the input device. For the various devices errors were defined as typing an incorrect letter for the keyboard, touching the screen outside of the box for the touch panel, and pressing a mouse button while the pointer was outside the box for the mouse.

Following the practice trials subjects read general task instructions along with a brief description of the upcoming application (no attempt was made to develop a complete instruction manual for the applications). The first menu panel of the appropriate application appeared on the task display and the task descriptions were presented sequentially on the instruction screen. The successful completion of one task (detected by the simulation system) initiated the presentation of the next task description and returned the application to the first menu panel. Subjects learned the telephone and calendar systems by attempting to complete the assigned tasks. No additional assistance was provided.

Subjects performed the same 20 tasks using each of the three input devices. Visually, the menus were identical for each input device condition. A single prompt line at the bottom of the screen displayed "Select your choice". (Since the tasks followed practice with the current selection method, we did not expect or observe confusion with how selections were to be made). All tasks required menu selections, and half of the tasks also required text entry into "fields". In order to enter text into fields, subjects first selected the target field on the task display. After selection the field was "activated" by the system, and visually identified by a prompt line. Subjects then typed the required information into the selected field, using backspace correction if necessary, and deactivated the field when finished by pressing the RETURN key. Task completion time was measured from the presentation of the problem on the instruction screen to completion of the entries to satisfy the goal stated in the problem.

RESULTS

Preference data

Subjects were asked to indicate preferences between the input devices for the experimental tasks. After all tasks were completed, subjects were asked to answer a series of questions comparing all pairs of input devices for different task types (practice, selection only, and selection with typing), and applications. There was no significant difference between the preference ratings for the two applications (calendar and telephone), or between the practice and selection-only tasks. For this reason data from these groups is combined. Figures 1 and 2 summarize the results for the 24 subjects in Experiment 1.

Comparisons were made using the first choices of subjects for the experimental tasks. For all tasks the subjects preferred the keyboard or touch panel to the mouse (Chi-square = 9·2, 10·3, and 18·2 for practice, selection only and selection with typing tasks, df = 2, all probabilities less than 0·01). The keyboard and touch panel were given similar preference ratings for selection-only tasks. For tasks with both selection and typing, subjects preferred the keyboard over the touch panel (Chi-square = 4·2, df = 1, p less than 0·05). The difference between stated preference based on task type, though not significant (Chi-square = 2·6, df = 1), indicates some desire by subjects not to switch input devices.

Performance data—practice tasks

In total, each subject completed two sets of 25 practice trials for each of the three input devices. Equipment problems caused practice data from five of the subjects to be lost. For the remaining 19 subjects a 3 (device) by 2 (set) within-subject analysis of variance for the mean of each set of practice trials was performed. Because of the lost data it is necessary to assume that the device order does not affect the performance in this analysis, an assumption that seems reasonable for the target selection task.

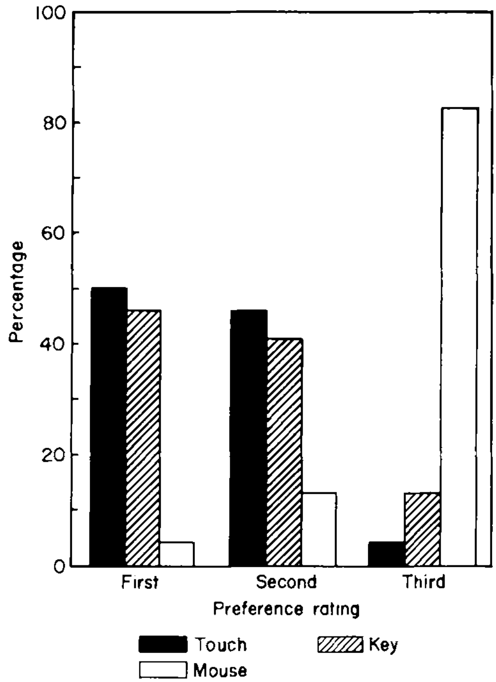

FIG. 1. Selection-only preferences for input devices. Experiment 1.

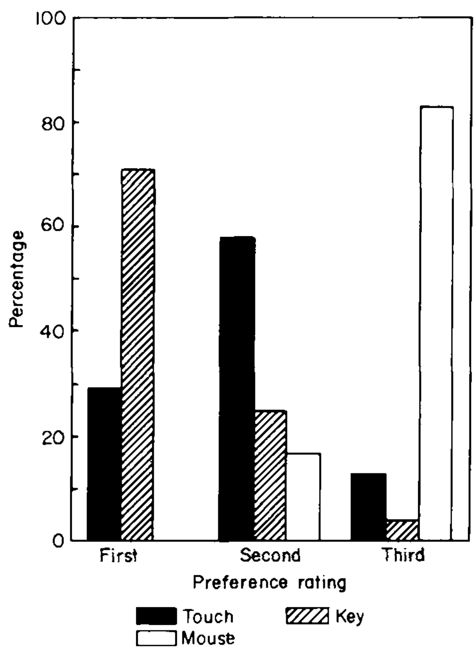

FIG. 2. Selection with typing preferences for input devices. Experiment 1.

Main effects for device ($F(2, 36) = 54{\cdot}9$) and set ($F(1, 18) = 36{\cdot}9$), and the device by set interaction ($F(2, 36) = 10{\cdot}2$), were all significant. Subjects were faster making selections with the touch panel and keyboard (an average of 0·8 and 1·1 sec per selection respectively) than with the mouse (2·7 sec). All devices showed improvement for second set over first, showing a strong practice effect. The difference between the mouse and the other devices was greater on the first set (means of 1·0 and 1·2 for the touch screen and keyboard respectively and 3·3 sec for the mouse) than the second set (means of 0·7 and 1·0 sec for touch panel and keyboard and 2·0 sec for mouse).

Performance data-menu tasks

Subjects completed the application tasks three times, once with each input device. Figure 3 summarizes the total time to complete the 20 menu tasks for the three repetitions. Each block includes 10 tasks within each application. Note that subjects using one device for the first block used a different device on the second and third blocks.

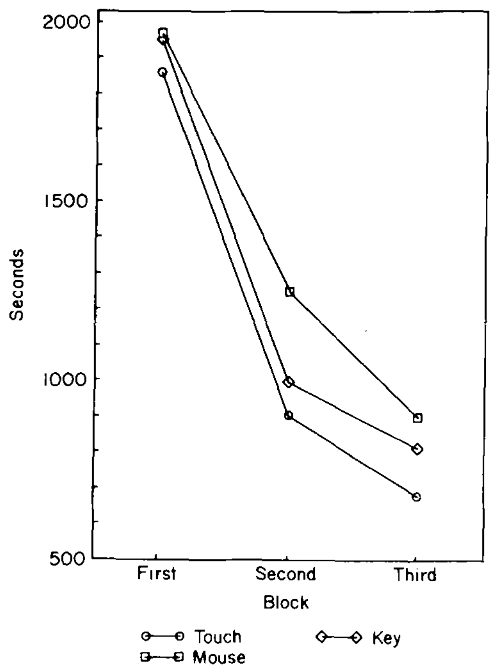

FIG. 3. Menu task performance. Experiment 1.

There is a clear tendency to complete the task sets faster with experience. On the average subjects took over 30 min to complete the first block of 20 tasks (10 calendar and 10 telephone tasks), and less than half that time to complete the tasks on the third repetition. This can be attributed to experience with the applications. Remember that the actual applications were new to the subjects, and the first block times include considerable trail and error learning.

The times to complete the menu tasks were subjected to an analysis of variance. Since the distribution of the task completion times was skewed, the analysis was carried out on the log-transformation of the raw completion times (although the analysis on the raw data yields very similar results). Presentation order of the devices was a between subjects variable with eight subjects in each of three device order groups (keyboard/touch panel/mouse, mouse/keyboard/touch panel, and touch panel/mouse/keyboard). There was no significant difference between the presentation order groups ($F(2, 21) = 1{\cdot}37$). Each subject used three devices, two task types (selection only and selection with typing) and two applications. There was a large within subjects order effect ($F(2, 42) = 343{\cdot}85$), with subjects completing the tasks faster on each repetition.

Several of the main effects are not of particular interest here. While there is a significant difference between performance on the two applications ($F(1, 21) = 153{\cdot}4$, with the telephone task faster), the two task types ($F(1, 21) = 776{\cdot}8$, with selection-only faster), and an interaction between application and task type performance ($F(1, 21) = 230{\cdot}0$), no attempt was made to control or equate these factors. The focus was on providing varied task types and applications, without studying their relative difficulty.

There was a significant difference in performance for the three input devices ($F(2, 42) = 17{\cdot}1$). Additional analysis showed a significant difference between the touch panel and mouse performance. No significant difference was found for comparisons of the keyboard with mouse or touch panel. Additionally, there was no significant interaction between task type and device, application and device, or task by application by device. Finally, there was no interaction of device with the within subjects order ($F(2, 42) = 2{\cdot}72$). In other words, the device effect did not vary over task repetitions.

DISCUSSION

Both the subjective evaluations and the performance data suggest that the touch panel and the keyboard are better menu-selection devices than the mouse. Performance measures in the applications showed an improvement of approximately 10% for touch panel over keyboard and 10% for keyboard over mouse. The difference was similar for selection-only and mixed selection and typing tasks. For practice tasks, which were simple target-selection tasks, the performance difference is much greater (the mouse took over three times the time required to make selections by touch panel). However, it should be kept in mind that actual selection represents only a part of the application dialogue (it is estimated that selection time composed approximately 12–25% of the task time for this experiment).

Several new considerations resulted from this work. Given that the subjects in the first experiment were all skilled typists, and we assume also skilled pointers, it was perhaps not surprising to find an advantage for typed and hand-pointing. The mouse was a new device for all subjects (performance gains over the practice trials were greater for the mouse than for the other devices). The performance results were slightly different than those one might expect based on the preference data. Even though the subjects stated a preference for keyboard entry in mixed entry tasks, the performance data showed an advantage for touch panel input.

Given the published research reporting advantages of the mouse as an input device, the failure of the mouse in Experiment 1 was disturbing. Three possible sources for this failure to replicate previous work were considered in designing the second experiment. Since Experiment 1 used only skilled typists, it may have been that this biased the study in favor of the well-known device (the keyboard). Experiment 2 tested both skilled and non-skilled typists in a controlled fashion. A second possible factor was sex of the subjects. The first study involved mostly female subjects, and there has been considerable investigation of sex differences in spatial processing tasks. The manipulation of the mouse to control the positioning of a screen pointer can be considered a spatial task. Thus, the second experiment examined male and female subjects. The third factor was amount of practice. Even unskilled typists are far more familiar with the keyboard than they are with the mouse. While practical limitations prevented us from developing "mouse experts", we felt that additional information could be obtained by extending the amount of practice with each selection device.

Experiment 2

METHOD

Subjects

Sixty-five subjects (34 females and 31 males) were obtained from a temporary employment agency. The subjects ranged in age from 17 to 50 years with self-reported typing rates ranging from under 20 to over 100 words per min. Subjects in Experiment 2 had not taken part in Experiment 1. Forty per cent of the subjects had obtained college degrees, 49% had some college but no degree, and the remaining 11% had completed high school. Most of the subjects (65%) reported some word-processing experience, but only eight had any programming experience. Seventeen subjects failed to complete the experimental tasks and were excluded from the analyses. Of the subjects failing to complete the experiment, eleven failed to finish due to equipment problems. All reported analyses are based on data obtained from the 48 subjects (24 females and 24 males) who successfully completed the experiment.

Apparatus and task environments

An IBM personal computer (PC) served as the experimental controller. The experimental setup was identical to that used in Experiment 1. The practice portion of the experiment was modified. In Experiment 2 targets were generated to provide four distances from a previous target, so that targets could be treated as pairs (in Experiment 1 target location was randomly selected). This was done to provide control over target distances between trials. Additionally, the mouse to pointer movement ratio was changed to 1:2, so that pointer movements were twice the distance the mouse was moved. This ratio is the same as that used by Card *et al.* (1983).

Experimental design

As in Experiment 1 all subjects used all three input devices to perform tasks within both applications. In this experiment subjects were classified as male or female, and within each group were classified as having low or high typing skill. The typing skill classification was made with the aid of information obtained from the agency providing the subjects. Subjects were divided into high-skill (over 40 wpm) and low-skill (under 40 wpm) groups. The result was mixed design with the same within-subjects factors as Experiment 1 nested within sex of subject and typing skill as between-subjects factors (2 by 2). The within-subjects design was identical to Experiment 1, with 12 subjects in each between-subjects group.

Procedure

Experiment 2 differed from Experiment 1 in the amount of practice completed by subjects during a session (all other aspects of the experiment were identical). For Experiment 2 subjects performed a block of 288 practice trials between each set of application tasks. The target location was varied in such a way to allow for evaluation of target selection time over four selected distances from the previous target. The amount of practice (two blocks of 288 trials each per device in Experiment 2 vs two blocks of 25 trials) was extended to increase proficiency and enable closer examination of device performance.

RESULTS

Preference data

Following completion of all tasks, subjects were asked to fill out a questionaire identical to that in Experiment 1. Similar preference results were obtained in the second experiment. The results for the 48 subjects are summarized in Fig. 4 (Practice), Fig. 5

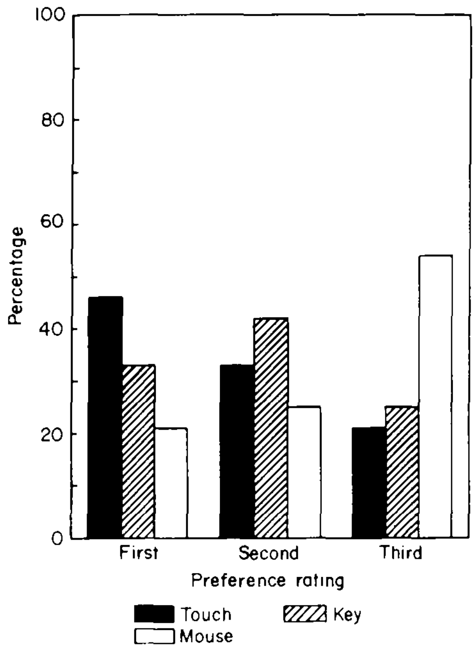

FIG. 4. Practice task preferences for input devices. Experiment 2.

(Selection only) and Fig. 6 (Selection with typing). Again the touch panel and keyboard were preferred to the mouse as an input device (Chi-square = 14·5 and 26·4, df = 2, for selection only and selection with typing tasks), although there was no significant difference in device preference for the practice task. While there was a tendency for the low-skill typists to prefer the touch panel more often than the high-skill typists for all task types, the differences were not significant. There were no significant differences in ratings between male and female subjects.

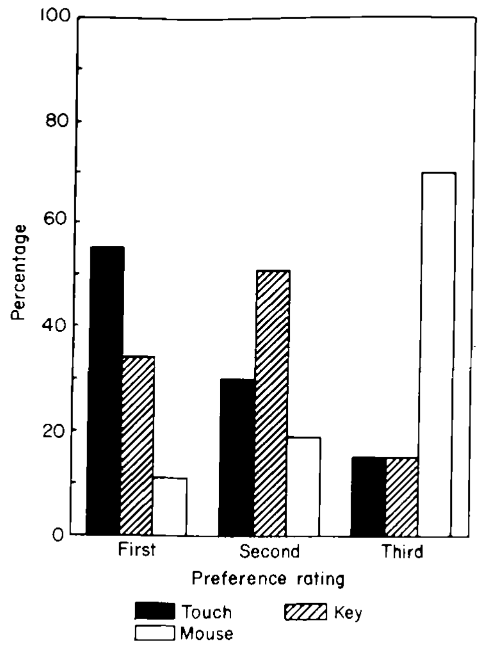

FIG. 5. Selection-only preferences for input devices. Experiment 2.

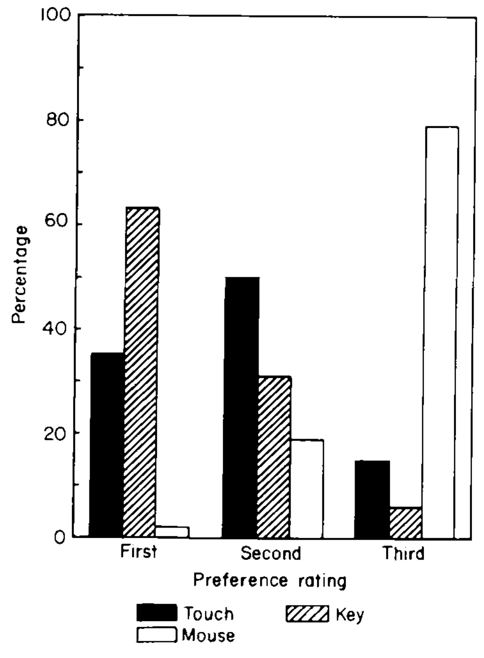

FIG. 6. Selection with typing preferences for input devices. Experiment 2.

In this experiment, there was a significant difference in the device ratings for the different task types. Selection with typing task ratings differed from both practice task ratings (Chi-square = 12·3, df = 2), and selection-only task ratings (Chi-square = 8·7, df = 2). The practice task ratings were not significantly different from the selection-only task ratings. The mouse was more favorably evaluated in the practice task compared with the mouse ratings in the selection with typing task (although it remained the least-preferred device). For the selection-only tasks, the touch panel was more favorably evaluated than the keyboard, while the reverse was true for the selection with typing tasks.

Performance data—practice tasks

The analysis here considers only the mean time to make target selections without regard to distance between the current and previous targets. Since target positions and distances were controlled for the different devices, this is a reasonable simplification.

In total, each subject completed two sets of 288 practice trials for each of the three input devices. For the 48 subjects a 3 (device) by 2 (set) within 3 (device order) by 2 (sex of subject) by 2 (typing skill) mixed analysis of variance for the mean selection time of each set of practice trials was performed. Data from all trials (including those in which the first response was incorrect) were used. An analysis using only correct trials yields the same pattern of results with slightly faster average selection times. There was no significant difference in overall preformance between the device order groups ($F(2, 36) = 1·38$). As in Experiment 1 there were significant main effects for device ($F(2, 72) = 85·5$) and set ($F(1, 36) = 53·0$), and the device by set interactions ($F(2, 36) = 12·5$), were all significant. Some details of the results did change in this experiment. Subjects were faster making selections with the touch panel (0·56 sec overall) than with either the keyboard or the mouse (1·16 and 1·30 sec respectively). All devices showed improvement for second set over first, demonstrating a strong practice effect. The improvement for the mouse (from 1·48 to 1·12 sec) was greater than that for the touch panel (from 0·60 to 0·53 sec) or the keyboard (from 1·24 to 1·12 sec). There were no significant main effects of sex of the subject or skill, but there was a significant interaction between skill and device ($F(2, 72) = 4·8$). The skilled typists did do better with the keyboard selection than the low skilled typists, averaging 1·00 sec per selection compared with 1·32 sec. It is interesting to note that the low-skill typists did about as well with the keyboard as with the mouse (1·28 sec for the mouse).

One additional note needs to be made concerning the practice data performance. Subjects were not penalized for error trials and were simply told to make selections as rapidly as possible. There are greatly different error rates (typing or pointing errors) for the different devices. For keyboard selection the error rate is about 5%, but the rate is much higher for the touch panel (27%) and mouse (22%). For the touch panel these error rates may have been due to parallax-like problems, but are also associated with a tendency for some subjects to drag their fingers across the screen toward the target (which results in the first touch of the screen being outside the target box). For the mouse they may have been the result of a strategy of holding down the mouse button and "sliding" over the target. If these error rates are real problems with the selection techniques (causing incorrect selections in real tasks), we would expect to see increased performance times for the menu-selection tasks. For these tasks errors will result in incorrect menu presentations.

Performance data—menu tasks

Subjects completed the application tasks three times, once with each input device. Figure 7 summarizes the total time to complete the 20 menu tasks for the three repetitions. Each block includes 10 tasks within each application. Note that subjects using one device for the first block used a different device on the second and third blocks. The pattern of results are the same for Experiment 1 and Experiment 2.

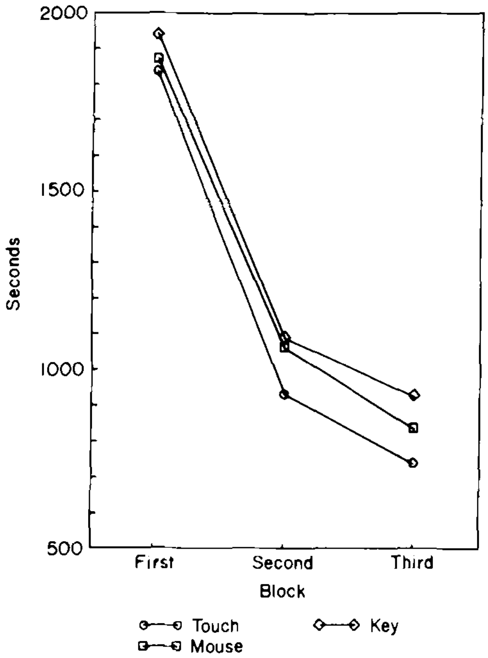

FIG. 7. Menu task performance. Experiment 2.

The times to complete the menu tasks were subjected to an analysis of variance. Because of the positively skewed distribution of the task completion times, analysis was performed on the log transformation of the completion times (performing the same analysis on the untransformed data yields very similar results). Sex of subject, typing skill and order group presentation for of the devices were between subjects variables with four subjects in each of the groups. The device orderings were the same as used in Experiment 1. Each subject used three devices, two task types (selection only and selection with typing) and two applications.

As in the first experiment, there was no difference between the device order groups. Differences for both sex of subject and typing-skill were not significant. Examination of the interactions shows one particularly interesting result. There is a significant interaction between task type, device and skill level ($F(2, 72) = 5{\cdot}3$). While high-skill typists perform better for tasks involving typing with all selection devices, they do not perform better on all selection-only tasks. As might be expected, for selection only tasks typing skill affects only performance when the keyboard is the input device.

For the remainder of the variables, the results were very similar to those in Experiment 1. There was a significant difference between performance on the two applications

($F(1, 36) = 130{\cdot}4$, with the telephone task faster), the two task types ($F(1, 36) = 1382{\cdot}2$, with selection-only faster), and an interaction between application and task type performance ($F(1, 36) = 408{\cdot}7$). No attempt was made to control or equate these factors. There is a highly significant within subject order effect ($F(2, 72) = 540{\cdot}5$), with performance improving with each repetition. These results are all consistent with those of Experiment 1.

Of considerable interest here is the replication of the device effect. There was a significant difference in performance for the three input devices ($F(4, 72) = 15{\cdot}3$). Additional analysis showed a significant difference between the touch panel and mouse performance, and a difference between touch panel and keyboard performance which was not found in Experiment 1. No significant difference was found for comparisons of the keyboard with the mouse. Additionally there was no significant interaction between task type and device, application and device, or task by application by device (i.e. no difference in the pattern of the main effect of device for the different applications or tasks). Thus, there was no difference in the device effects for different tasks or applications.

General discussion

The performance of subjects using the touch panel was better than that for the keyboard or mouse. While mouse performance did seem to have improved in the second experiment, it remained no better than the keyboard performance. For skilled typists, the level of performance for the keyboard remained above that for the mouse (as in Experiment 1). The mouse performance did not seem to be linked to sex or typing-skill differences of the subject. High levels of practice experience did improve performance, but there were significant improvements for all devices. While improvements were greatest for the mouse, there was no indication that this improvement would cause mouse performance to surpass keyboard or touch-panel performance even after considerable practice. The differences found in the applications tasks were consistent with those that would be expected based on the practice data and the number of selections involved in the menu tasks.

It is difficult to account for these findings within the framework suggested by Card *et al.* (1983). One possibility is that keyboard menu entry and mouse selection require additional cognitive processing compared with finger pointing. This hypothesis is consistent with the notion that touch selection is a highly automated skill for most humans, and that other techniques are less well learned. We do not suggest that these studies render previous findings invalid, only that more attention needs to be given to the nature of the dialogues for which the device is to be used, and the skills of the users. At the very least we must conclude that less "natural" devices such as keyboards, can in some circumstances be preferred and lead to better performance than more "natural" pointing devices such as mice.

This points to a potential weakness in the emphasis placed on the performance of "ideal" users in the Card *et al.* (1983) framework. While some progress may be made in using this approach, we should not expect to be able to use it to predict all aspects of ease of use. It must be remembered that the keystroke analysis emphasizes minimization of motor behavior, and does not adequately consider cognitive behavior except for situations in which the activity is very well learned. There is the added danger that

Card *et al.* suggest that their framework can be used to replace "costly" experimentation. We would argue that until an adequate understanding of cognitive components is reached, we cannot rely on simplistic cognitive models.

Do our results imply that we should build devices with touch panels rather than with mice? As Carroll (1983) points out, such questions deal with issues which are technology-dependent, and difficult to answer in a general theoretical way. While we would prefer to conduct work with general utility, we feel somewhat restricted. The current study supports the statement that when comparing these devices in a menu-selection task, the touch panel is superior.

A final word of caution in interpreting these results. There has been differing opinion on the optimal number of buttons to place on a mouse. Through the use of additional buttons, we can add functional capabilities to the mouse that may be difficult to duplicate with a touch panel. Although it has not been studied, it is not inconceivable that the functions possible with an "enhanced" mouse might justify the added complexity of such a device for certain applications or users. Unfortunately there is no easy answer to questions which ask whether a one, two, three or even a twelve-button mouse is optimal without asking what the device is to be used for. In a field where enhanced function can justify ease of use tradeoffs, we are forced to delay answers to such questions.

References

CARD, S. K., ENGLISH, W. K. & BURR, B. J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics*, **21**, 601-613.

CARD, S. K., MORAN, T. P. & NEWELL, A. (1983). *The Psychology of Human-Computer Interaction.* Hillsdale, New Jersey: Erlbaum.

CARROLL, J. M. (1983). Presentation and form in user interface architecture. *IBM Research Report, RC 10144.*

COOPER, W. E. (1983). *Cognitive Aspects of Skilled Typewriting.* New York, N.Y.: Springer-Verlag.

NEWELL, A. & SIMON, H. A. (1972). *Human Problem Solving.* Hillsdale, N.J.: Prentice Hall.

SHIFFRIN, R. M. & DUMAIS, S. T. (1981). The development of automatism. In Anderson, J. R. Ed., *Cognitive Skills and Their Acquisition*, pp. 111-140. Hillsdale, New Jersy: Erlbaum.

SHIFFRIN R. M. & SCHNEIDER, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, **84**, 127-190.

WHITFIELD, D., BALL, R. G. & BIRD, J. M. (1983). Some comparisons of on-display and off-display touch input devices for interaction with computer generated displays. *Ergonomics*, **26**, 1033-1054.