136
61

A HUMAN FACTORS EVALUATION OF CURRENT TOUCH ENTRY TECHNOLOGIES

by

Edward A. Baggen

Dissertation submitted to the faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Industrial Engineering and Operations Research

APPROVED:

H.L. Snyder, Chairman

J.S. Greenstein

P.T. Kemmerling

A.M. Prestrude

W.W. Wierwille

A HUMAN FACTORS EVALUATION OF

CURRENT TOUCH ENTRY TECHNOLOGIES

by

Edward A. Baggen

Committee Chairman: Harry L. Snyder

Industrial Engineering and Operations Research

(ABSTRACT)

This research was part of a program sponsored by the Army Electronics Research and Development Command. The program goals are to develop a generic input/output device based on a 4- x 8-in electroluminescent flat panel display coupled with a touch sensitive input device. The primary purpose of this dissertation was to evaluate the six major manufacturers' implementation of the three most common technologies for touch screens. The evaluation was based on operator performance. The three technologies represented in the research were IR beam matrix switches, conductive membrane switches, and transparent capacitance switches. A secondary goal of the research was to establish a link between measurable hardware parameters of any touch sensitive device (TSD) and operator performance. These parameters were then used to build models of operator performance under a variety of conditions.

The primary goal of technology evaluation was approached through two experiments based on two generic types of tasks typical of current and expected TSD applications. These experiments compared six different TSDs from different manufacturers across varying conditions of lighting, viewing angle, and touch target dimensions.

The secondary goal of TSD performance modeling was accomplished through careful measurement of many image quality and touch sensing characteristics of the six TSDs and subsequent construction of stepwise linear regression models of user performance. These models were built using the performance data collected in the first part of the evaluation.

Results from the performance comparison revealed that across tasks and conditions, one device of the IR beam technology was found to be the best performer. Another device of the same technology was equivalent in reading aspects of performance but inferior with respect to touch inputting performance. This performance difference was hypothesized to be due to differences between these two devices in touch sensor parallax.

The linear regression modeling effort resulted in the identification of several hardware parameters which are important to TSD user performance. Additionally, models of performance under specific conditions were developed which accounted for most of the variation observed in the performance data.

ACKNOWLEDGMENTS

I would first like to express my appreciation to the members of my dissertation committee, Dr. Albert Prestrude, Dr. Joel Greenstein, Dr. Walter Wierwille, and Prof. Paul Kemmerling. Each provided me with good advice and technical support during this project. Moreover, each member of my committee has added a different and valuable dimension to my education for which I am very grateful.

I also must thank the members of the research group who contributed to this project. Thanks to , , and especially, , for his technical assistance and guidance in the areas of research that no one tells you about and that I certainly did not foresee. Special thanks to my committee chair and dissertation advisor, Dr. Harry Snyder, for providing insightful advice at difficult times that always seemed so obvious right after he said it. Also, for having always kept this project and my education as a high priority, never missing an opportunity to lend assistance beyond what was expected.

Foremost, I thank , my wife and mentor, for leading the way and leaving plenty of bread crumbs for me to follow. I never would have made it without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Since the mid-1960s, technologies have existed which allow people to interact with a computer by touching a display monitor. The computer side of this interaction consists of the generation on the display of information about viable touch options and the sensing at the display the location of any touch input. The location of a touch input triggers a pre-determined set of computer actions which correspond to the information provided to the operator.

Touch screen interaction is an attractive solution to human-computer interface design for a variety of reasons. For example, a touch screen and its related display serve as a unitary input and output device. This dual function allows the operator to focus attention on one location rather than two. Also, pointing to a desired object is a natural form of human communication which requires little training. No complex set of rules or physical skill has to be mastered for effective communication between the human and the computer.

A further advantage of touch screen devices (TSDs) is their flexibility. Both the information displayed to the operator and the action assigned to a touch location are software determined and can be changed depending upon the application. Thus, the amount of abstraction, interpretation, and memorization of functions required of the operator can be adjusted to the capabilities and demands of the user population for each application of the computer system.

Uses of Touch Screen Devices

Touch screens are in common use today in several settings where flexibility and ease of interaction are primary requirements. The earliest use of TSDs was the air traffic control industry in Great Britain (Johnson, 1967). This application consisted of the use of a CRT to display and input digital flight information on target aircraft. The information was input at fixed switch locations which could be assigned a variety of functions or aircraft ID sequences. The touch screen in this case replaced a keyboard as an alphanumeric input device.

Several evaluations have supported the use of TSDs as flight data display and entry devices in an air traffic control setting (Hopkin, 1971; Johnson, 1967; Stammers and Bird, 1980). Johnson (1967) reported that in comparisons with more conventional keyboards, "...the touch display provides both a faster and more accurate means of communicating between an operator (air traffic controller) and a data-processing system (p. 277)". Stammers and Bird (1980) found experienced controllers preferred the touch entry method of flight data communication over keyboard entry and marking of flight data strips.

Other applications of TSDs include process control, aircraft cockpits, automobiles, education, and libraries (Pfauth and Priest, 1981). In each of these applications TSDs offer advantages over other input devices which make their use profitable. Factors which support TSD use include: (1) high stress environments;(2) untrained user populations; and (3) limited space (Pfauth and Priest, 1981).

All of the applications mentioned above share a common task structure. This structure consists of the two-way transfer of alphanumeric information through the touching of "keys" or touch locations on the display screen. These "keys" may change function or meaning depending upon the task but they generally do not change in size or location on the screen. For example, in one automotive application, menus of possible actions are presented to the operator (Monty, Snyder, Farley, Donohoo, and Baggen, 1985). Menu items are arranged in fixed locations on the screen and a touch of a menu label is used to activate that function. In summary, present applications of TSDs are typically limited to key type inputs most commonly in a menu selection format.

Potential Uses of TSDs

As touch entry technologies become more advanced and microcomputers more powerful, the use of TSDs is likely to expand into other areas of human-computer interaction. Specifically, the control of a cursor to track, move, or place alphanumeric information is a potential use for TSDs.

The interface structure of a placement task is different from that of the key entry type tasks of most present TSD applications. Cursor placement requires the acceptance of touch inputs and the display of touch targets at any point on the display. Some placement tasks may also require tracking of moving targets.

At present, dedicated cursor keys, joysticks, trackballs, lightpens, and mice are used as auxiliary input devices to fulfill the placement function. Because of its limited accuracy due to human finger size and placement capabilities, it is unlikely that TSDs will replace other input devices in most cursor placement applications (Ritchie and Turner, 1980). However, special demands exist in some applications which favor the use of TSDs. In a review of TSD design and application issues, Pfauth and Priest (1981) concluded that TSDs have advantages over other types of input devices when the task or task setting is extremely demanding or when space is limited.

Gaertner and Holzhausen (1980) evaluated the use of a TSD to replace existing trackball and joystick systems in an air traffic control setting. They used a touch overlay on a computer-generated radar map for two tasks previously accomplished by a combination of trackball and keyboard entries. The tests demonstrated that TSDs were effective, accurate, and easily accepted by operators in this application.

These air traffic control tasks require the controller to key in flight instruction and heading information using special function keys. In this situation correctness of target selection and accuracy of command inputs are critical. Time pressure on the operator is also quite high, particularly under heavy traffic conditions.

Gaertner and Holzhausen (1980) recommended the use of a single TSD for both locating targets on the radar map and inputting command information. They pointed out that touching a target is quicker and requires less mental effort than manipulating a control to place a cursor and pressing a key to select that target.

The TSD also provides immediate feedback of the target selection at the place where the operator's attention is focused. Furthermore, inclusion of key functions on the same device allows the operator's attention to remain focused on the primary display during the task. For these reasons, Gaertner and Holzhausen predicted better performance of air traffic control tasks with touch technology.

Other proposed uses for TSDs include control of remotely piloted space vehicles (Slaby, Hartley, and Pulliam, 1985) and command and control of military activity (Davis and Badger, 1982). Some tactical military applications will require touch screen control accurate down to the single display pixel. Waller (1984) reported that Hughes Aircraft Company was developing a system to meet this requirement.

From a review of current and proposed uses of TSDs, it is evident that the strongest asset of touch technology is its flexibility as a multifunction keyboard (Davis and Badger, 1982; Pfauth and Priest, 1981). The ability of touch entry keys to change function between and within applications should insure the continued growth of TSDs in human-computer interfaces. There is also a growing demand for TSDs as relatively high resolution pointing devices, particularly in military (Waller, 1984) and air traffic control (Gaertner and Holzhausen, 1980) applications. However, it is unlikely that TSDs will replace other pointing devices such as light pens, trackballs, etc., except in applications where special circumstances create an advantage for touch. These circumstances might include extreme environmental conditions where dirt or mechanical abuse might damage other devices, limited space availability, and demanding task situations in which attention cannot be diverted from the primary display.

Touch Technologies

Several technological approaches to touch screen sensing have developed over the past two decades. Design solutions to the touch screen problem have ranged from conductive crosswires inlaid on the display surface to pressure-sensitive strain gauges mounted on the edges of a glass overlay. Of the many early approaches, four technologies have gained marketplace acceptance as touch sensitive interfaces. These four types of TSDs are capacitive film, conductive film, acoustic ranging, and infrared LED beam.

The infrared TSD uses arrays of infrared emitters and detectors positioned at opposite sides of a rectangular frame to sense the X,Y position of objects penetrating the plane of the frame. This frame is positioned around the edge of a CRT or flat panel display so that the plane of the detector array is broken when one touches the surface of the display. This device can be activated by a finger or any other object that breaks the grid pattern of the infrared beams. The resolution of the device depends upon the number and spacing of infrared emitters and sensors, as well as the thickness of the device (e.g., finger) that breaks the beam.

A variation on the infrared beam technology has recently been developed but has not as yet received much attention. This variation uses a single beam and a series of rotating prisms to cover the display surface area with a single beam. The X,Y position of a touch is derived from the time the beam is broken and the positions of the prisms at that time.

Capacitive TSDs rely on the capacitive quality of the human body to locate touches on a display surface. An electric potential is generated across a conductive film deposited on a glass surface (ideally the display surface). This potential is changed by contact with a human finger. The coordinates of the touch are computed by sampling of the potential across the screen in two orthogonal directions and locating the X and Y samples with the largest potential difference (Ritchie and Turner, 1975). Resolution of this device is limited by the sampling power of the device driver and the uniformity and predictability of the electric field produced on the display surface.

Acoustic ranging TSDs operate on the same principle as sonar. Acoustic waves are generated from two orthogonal edges of the display surface and, because of the acoustic properties of glass, surface waves are propagated across the display. A finger or other fairly dense object will cause a part of the surface wave to be reflected back toward its source creating a return signal on sensors placed there. The time between transmittance and return of the signal at each sensor is proportional to the distance that the wave has traveled. Therefore, the coordinates of a touch can be determined by the time intervals of the returns from the two orthogonal transceivers (de Bruyne, 1980).

Conductive membrane TSDs are similar to capacitive devices in that electric potentials are created across transparent conductive layers in front of the display. The major difference between conductive and capacitive devices is that conductive film devices require two layers of conductive material while capacitive devices use one. These two layers are generally separated by small nonconductive bumps on the back of the outer conductive film. A pressure from a finger or other object deforms the outer layer and makes a contact between the two layers, thereby creating an electrical contact. This change in the electrical signal can then be transformed into coordinates on the display surface (Greenstein and Arnaut, 1987). Conductive film TSDs are usually matrix devices with a pattern of conductive and nonconductive areas on each layer. Each conductive area is addressed individually so that a short at one area is immediately transformed into a touch location on the display surface. Another type of conductive film TSD (referred to as an "analog" device) converts analog voltages from the TSD into coordinates.

Because of differences in underlying technological bases, each type of TSD has certain unique characteristics which may prove to be advantageous or disadvantageous under specific task requirements. Logan (1985) has compiled a list of characteristics of each of the four major technologies which he thinks are important to user acceptance and performance. Logan cited resolution type, parallax problems, stylus limitations, durability problems, optical clarity, tactile response, and environmental sensitivity as parameters by which the four types of TSD vary significantly.

Other advantages and limitations of each technology have been identified (Schulze and Snyder, 1983). Infrared TSDs, for example, have an upper limit on touch resolution, determined by the spacing of individual LEDs and detectors, of about 1/8 inch between touch centers. Infrared panels also have a problem with false touch recognitions due to the fact that any object which breaks the beam pattern will activate a touch whether or not that object actually touches the display surface.

A related problem with infrared devices is that parallax occurs between touch recognition points and display targets when the object used to activate the touch is not perpendicular to the display surface. This phenonmenon is particularly troublesome when curved-surface CRTs are used as the display. The curved surface of the CRT requires that each infrared beam (which must travel a straight line) increases in distance from the display surface from the center to the edge of the display. Therefore, parallax is not constant across the display with CRT-driven infrared TSDs.

On the other hand, infrared beam TSDs enjoy some significant advantages over other technologies. For example, infrared beams do not interfere with the optical qualities of the display. They also have a high level of linearity and stability of touch locations. Acoustic ranging devices share the advantage of not producing any visible interference with the display surface, unless they are applied to a glass overlay rather than the display surface itself.

Both capacitive and conductive film TSDs require additional solid surfaces to be placed between the viewer and the display and therefore change the optical quality of the display. Conductive devices require two conductive surfaces to be added while capactive devices require only one extra film layer. Both conductive and capacitive film devices are also susceptible to reduced reliability caused by mechanical wear from extended usage. In the case of capacitive film, uneven wear at common touch points changes the electrical field at those points and produces error in assignment of X,Y coordinates. Conductive films are susceptible to cracking of the second conductive layer from frequent deformation at touch locations.

Capacitive film TSDs have the unique characteristic of only being activitated by the human body or another object of similar electrical qualities. This characteristic greatly reduces their susceptibility to accidental or spurious touch recognitions. Another characteristic unique to capacitive and "analog" conductive devices is that in the past they have been susceptible to drift of X,Y touch locations because temperature and humidity changes affect the electrical characteristics of the film. Recent design modifications by the major producers of capacitive TSDs have addressed this drift problem.

A characteristic unique to conductive film devices is that the outer film deforms when touched, thereby providing some tactile feedback. Also, the touch weight required to produce a touch signal can be designed to the needs of the application by changing the flexibility of the outer film.

The fact that acoustic ranging devices depend on a surface wave on the glass surface of the display limits the range of environments in which this device can be used. Scratches on the glass surface and foreign particles adhering to the glass surface will interrupt the waveform and produce a return signal indicating a touch. Thus, extremely rough or dirty application environments are not recommended for this type of device.

Variables Influencing Human-TSD Interactions

Identified in the TSD research literature are several variables which affect user performance and preference. Display declination or viewing angle, for example, has been shown to affect both performance and preference (Beringer and Peterson, 1983, 1985). In a relatively high resolution search and touch task, Beringer and Peterson (1983) found that increasing the angle that the display is declined away from the viewer increases touch error in the vertical direction. Vertical error was also found to be affected by the vertical position of the touch target on the display. Y-touch error was greatest at the higher declination angles and at the top or most distant edge of the screen. Beringer and Peterson (1983, 1985) hypothesized that this combination of error results could best be explained by parallax inherent in the infrared device used in the studies and by reach distance to the target.

Weiman, Beaton, Knox, and Glasser (1985) found that horizontal viewing angle affects the number of errors in key-entry type TSD tasks using both an infrared and a capacitive film TSD. These findings indicate that parallax created by the spatial separation of the touch active plane and the phosphor plane in TSDs regardless of technology has an effect on touch accuracy.

Image quality. Given the same display as an image source, different types of TSD overlays have varying effects on the quality of the image as seen by an observer. Pfauth and Priest (1981) have listed the effects of touch sensitive overlays on image quality as important factors in the selection of TSD devices. However, only one study has been conducted which attempted to measure TSD image quality and relate it to operator performance and preference. Schulze and Snyder (1983) conducted three experiments, each based on a different operator task. In Experiment 1, the operator's task was to locate and point to a target symbol in a random field of detractor symbols. Variables included in this experiment included CRT phosphor, video polarity, dot matrix character size, density of detractor symbols, and TSD technology.

Experiment 2 utilized an airplane seat assignment task in which subjects assigned passengers to seats on a graphic representation of a commerical airliner. The task consisted primarily of following a hierarchy of menu selections. Independent variables in Experiment 2 included CRT phosphor and CRT technology.

Experiment 3 consisted of a city information directory task. Subjects proceeded through hierarchies of information on civic events by choosing menu categories with touch key entries. In this experiment video polarity, CRT phosphor, and TSD technologies were varied at the same levels as in previous experiments.

The tasks in Experiments 2 and 3 were selected as representative of current applications of TSD technology. Unfortunately, as with many real-world type tasks, they did not lend themselves easily to the rigorous experimental control required of this laboratory comparison. For example, each trial of Experiments 2 and 3 consisted of many short decision steps on the part of the subject. Although each trial had the same number of steps, each was unique in the actual decisions and selections made. In these experiments there was no way to guarantee that each trial was equally difficult. The complexity of these two tasks may have produced enough extraneous variability in results to obscure hardware-related differences in performance.

In each experiment, Schulze and Snyder collected task time and number of touch errors as performance measures. They also had subjects fill out subjective rating scales with measures of legibility, usability, aesthetic quality, and other qualitative judgments. In addition to performance and preference, several measures of image quality were taken of each combination of CRT and TSD. These measures were all based on luminance modulation throughput of the system and on luminance variation across the display.

Data from each performance, preference, and image quality measure were initially analyzed individually. After separate analyses of each measure, image quality measures were correlated with each performance and subjective measure. The effects of interest to the present proposal are the main effects of TSD technology on performance and preference, particularly in Experiment 1, and the correlations of these results with image quality measures.

Each of the three experiments showed significant main effects of TSD on some performance and subjective measures. However, the ordering of devices in terms of these measures was different for each experiment as well as between measures. The inconsistency of findings within and between tasks indicates that other factors partially confounded with TSD technologies, but not controlled within this study might have influenced performance and preference. Given the mixed results within performance and preference measures, it is not surprising that the correlations between performance and image quality as well as between subjective ratings and image quality did not provide a consistent pattern across measures or tasks.

A point of procedure from Schulze and Snyder's (1983) study has importance to the present study. In all three experiments, display targets were arranged so that the center of the target always corresponded with the center of a touch location on the TSD. Additionally, the active touch location for a correct touch in each task was set to be greater than the minimum location size for the TSD with the lowest touch resolution. This procedure insured that all three tasks in Schulze and Snyder's study were equivalent to the key-entry type task described above.

Parallax. Much of the literature on TSDs points to parallax as a source of errors in operator inputs (see, for example, Beringer and Peterson, 1985; Beaton and Weiman, 1984; Logan, 1985; Stammers and Bird, 1980). At present, there is no agreed upon geometric or operational definition of parallax as it pertains to TSD interface. Such a definition must be made before measures of parallax can be taken and their effects on performance evaluated.

One issue in defining TSD parallax is whether the same measure should cover the cases of both infrared and conductive film TSDs. Parallax errors occur in an infrared device when the end of the object used to point at the target does not match the location of the part of that object which is breaking the invisible infrared beams. This is visually quite different than the case of film TSDs in which the end of the pointing object is at the touch surface but is not able to penetrate to the display image plane.

Both these phenomena are the result of parallax between the touch plane and the image plane; however, we do not know if they are perceptually similar enough to be covered by a single label. The only evidence related to this issue is the finding of Weiman et al. (1985) described above in which infrared and capacitive devices produced similar touch errors in off-angle viewing conditions. This issue as well as the general issue of accurately describing the effects of parallax on TSD performance are areas requiring more research.

Touch device resolution and task resolution requirements. There are two related issues concerning touch resolution in TSDs. The more researched of these issues deals with the size of touch targets for best performance under varying task conditions. The related issue which has not been addressed in the literature concerns the minimum touch point resolution requirement for a device, given varying task requirements.

Beaton and Weiman (1984), Weiman et al. (1985), and Gaertner and Holzhausen (1980) have all conducted investigations into the best size of touch-sensitive targets in a key-entry type of application. These studies agree that a touch key size of approximately 20 mm^2 produces a low level of input errors and is small enough for most multi-key applications.

In the key-entry type of application, touch key locations on the display can be predetermined and positioned to coincide with the boundaries of the minimum resolvable touch locations of the touch sensors. Thus, for this application, resolution of touch sensing below 20-mm centers would not enhance performance and might contribute to unnecessary expense and/or slow response time of the device.

Relatively low resolution devices using fairly large touch active keys are suitable for most present and foreseeable uses of TSDs. However, special applications, such as military command-control and commercial air traffic control include tasks which may have higher resolution requirements. For example, the target selection task in air traffic control described above probably has different touch target size and sensor resolution requirements than a fixed location key application.

With a fixed-position target, the sensor resolution requirement is dependent upon the minimum target size which, in turn, is dependent upon the operator's touch accuracy limits and the density of keys required by the task. If touch target location is random and variable, then the relationship between touch target size and sensor resolution is no longer simple. In this case, sensor resolution is more dependent upon the requirements of the task. For example, in a battlefield command map application, the task may be to pinpoint one small graphic symbol closely surrounded by other potential targets. Even this relationship is subject to mediation by software solutions such as zoom or cursor movement by touch keys. These solutions, while practical, would be slower and less direct than the fine resolution touching of targets.

When a display target is congruent with the boundary of a touch location, a touch anywhere within the boundary of the target will lead to a correct or accurate touch input at the centroid of the target. On the other hand, when display targets are randomly located on the screen, the relationship between the boundary of the display target and the centroids of touch locations is less direct and can result in some touches outside the targets being hits and others inside the target being misses.

When the resolution of the touch panel is equal to or exceeds that of the display, then the one-to-one relationship between display target boundaries and accurate touches will again hold. At lower touch sensor resolutions, the area of all touch locations that have a centroid which falls within the boundary of a display target (resulting in a hit) will vary depending upon the location of that display target. Figure 1 illustrates this relationship between touch resolution and hit area variation.

From the above discussion, it can be seen that the effects of touch sensor resolution on task performance can be expected to be quite different depending upon whether random or fixed display target locations are used. To date, the only studies to be conducted with relatively high touch accuracy requirements have been done with fixed target locations (Beringer and Peterson, 1983; 1985) and

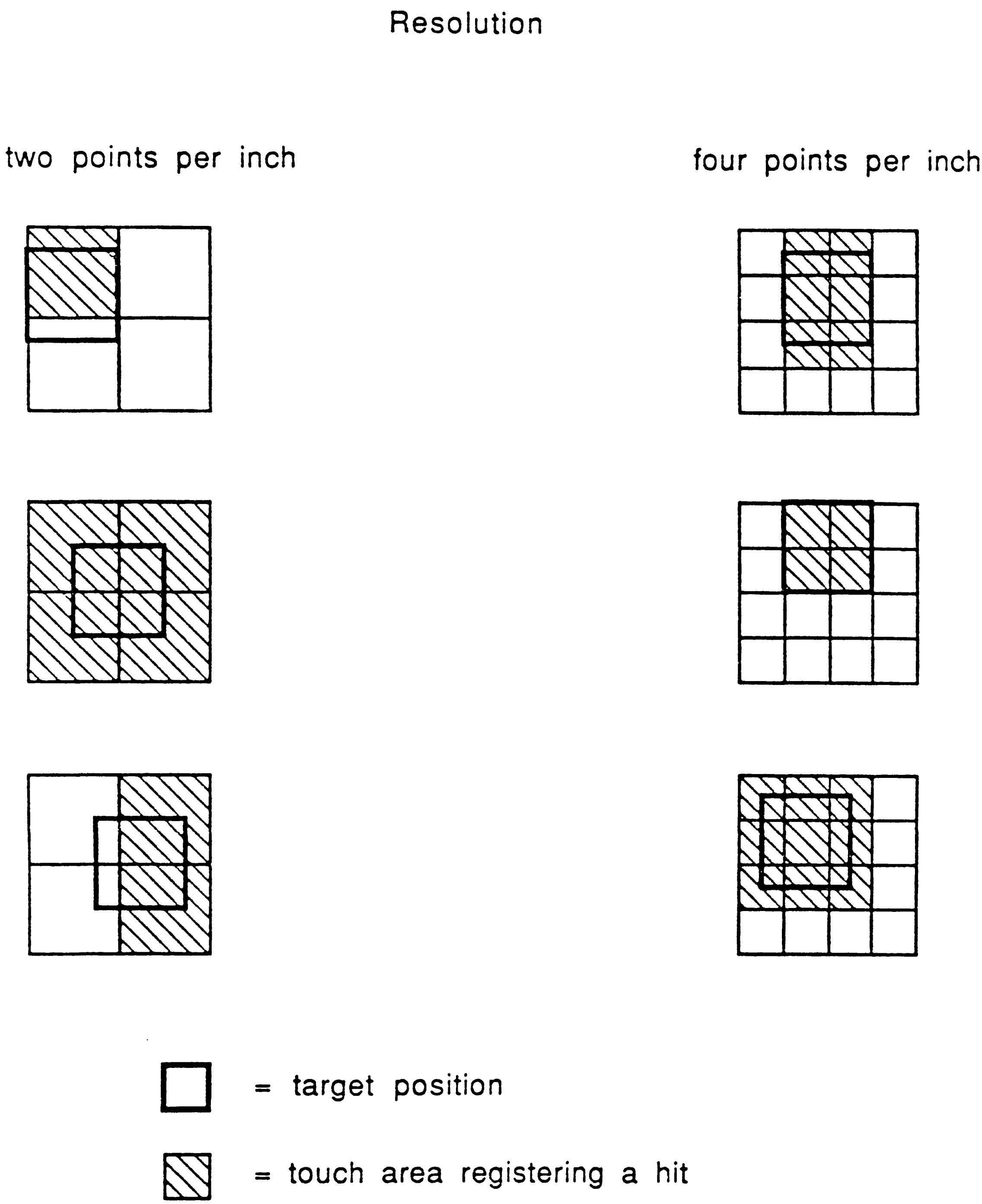

Figure 1: Effect of Touch Resolution on Hit Area.

thus do not address the resolution requirements of target selection type tasks.

Touch response characteristics. Several touch panel response characteristics have been identified which have an impact on user performance and preference. These include feedback of touch recognition, touch device reliability, tactile pressure required for a touch activation, and response time of the device from touch inputs.

Visual feedback of touch registration has been shown to reduce touch errors and to lower training time requirements (Beringer and Peterson, 1985; Weiman et al., 1985) for TSD operators. Hopkin (1971) found that reducing TSD reliability to 99.7% caused operators to reject TSDs and prefer keyboard entry in air traffic control tasks. Logan (1985) has pointed out that the feel of contact at the display surface may affect operator performance but no study has confirmed this statement to date. Pfauth and Priest (1981) and Hopkin (1971) have pointed out that system delays, especially when no feedback of acceptance is given, disrupt the natural advantages of TSD operation.

Touch sensing reliability, feedback, and response time are specific to the particular approach to and implementation of the TSD interface taken by a manufacturer. The quality of components, for example, can have a great impact on both reliability and response time.

Touch pressure, however, is more directly linked with and subject to the limitations of a TSD interface technology. The four technologies described previously vary considerably in the amount of contact required for touch acceptance. An infrared device, for example, does not require any contact with the device for a touch to be registered. A conductive film device, at the other extreme, requires a deformation of the plastic overlay in order to register. This difference among technologies has not been previously investigated in terms of its effects on operator performance or preference.

PURPOSE OF THIS RESEARCH

In summary, the review of research literature on TSDs has provided five important points.

1. The use of TSDs is expanding into a wide variety of applications. These applications will cover an equally wide range of environmental conditions under which the TSDs must perform.

2. The variety of different applications can be divided into two basic categories of task. These two task types are fixed location key entry and random location target selection.

3. Four different technological approaches to TSDs have evolved and survived. Each has become a viable entry into the TSD marketplace. Each of these four technologies has its own advantages and disadvantages with regard to meeting the requirements of expanding applications.

4. Several important TSD parameters which affect human performance have been identified. These parameters include image quality, parallax, touch sensor resolution, and touch weight.

5. Limitations of the four technologies cause them to vary with respect to the five parameters listed above. Therefore, it is likely that TSDs based on different technologies will exhibit different operator performance.

These five statements point out three previously unaddressed or unresolved research issues, including:

1. What effects do the different technological approaches to TSD design have on operator performance?

2. What contributions do image quality, parallax, device resolution, and touch force make to operator performance, independent of technological approach?

3. Based on measurable TSD parameters, can models of performance be generated which will predict performance across a range of tasks and task environments?

The present research is designed to address these issues. Previous work compared different technology TSDs with respect to operator performance and preference on different tasks (Schulze, Beaton, and Snyder, 1984; Schulze and Snyder, 1983). Subsequently, the performance and preference measures were correlated with measures of image quality of each device. The present research replicated portions of the previous study and extend them into new task and environmental settings. The present study also generates a more detailed model of human-TSD interaction, including factors other than image quality. From this enlarged model better predictors of performance and a better understanding of the relative contributions of image quality and other relevant parameters were achieved.

The present study is similar to Schulze and Snyder's design in that it is also based upon a direct comparison of TSD devices of different technologies. It again correlates image quality metrics with human performance on TSDs. The present study differs from Schulze and Snyder's by using only simple tasks that can be more adequately controlled and by including both key-entry and target selection type tasks. The independent variables manipulated in this study represent a range of task settings in terms of lighting, viewing conditions, and task demands. The goals of this study were to evaluate and model human performance with different TSD technologies under conditions that generalize to most present and projected applications of TSDs.

METHOD

General Approach

The underlying goal of this study was to discriminate among design alternatives on the basis of human user capabilities and requirements and to generate usable data for the development of specifications and guidelines for future designs. Although there are many alternatives, the most direct and often most meaningful way to assess the impact of a design is to measure user performance during interaction with the system of interest and to compare it with performance on an alternative system. Performance measures have a high face validity in that they often measure variables that are of direct interest to those concerned with application of the research. Also, some performance measures have been shown to be quite sensitive to changes in human-system interface.

Two shortcomings of direct performance based comparisons between design alternatives are (1) they are expensive and time consuming to conduct, and (2) they often do not generalize well to designs or modifications to designs that are not evaluated in the comparisons. Because of the cost, it is unreasonable to expect that each time a new or modified design becomes available, a full scale user performance-based comparison will be conducted. Therefore, it is important that when a performance-based evaluation is conducted, some connection is made between user performance and generalizable system parameters which can be used by designers, engineers, and/or consumers.

This research has as its core a user performance-based comparison among six TSD design alternatives. Additionally, several measures of physical parameters known to affect user performance were made on each device. Finally, the results of performance and device parameter measures were combined to produce models that predict user performance with TSD systems based on critical hardware parameters.

Performance Evaluation

Visual interaction with a touch panel display is an important part of the human-machine interface of interest in this study. Past experience in this laboratory (Abramson, Mason, and Snyder, 1983; Abramson, Mason, and Snyder, 1984; Snyder and Taylor, 1979) has shown readability and legibility tasks to be particularly useful for human-display interaction evaluations.

Readability refers to the ease of comprehension of the meaning of printed material. Readability has been shown to be influenced by a variety of factors including, contrast, clarity, letter and word separation, font, use of capitals and lower case letters, and length of words.

Legibility refers to the ease of recognition of familiar letters or symbols independent of context. Legibility generally implies less cognitive processing of information than reading but is also quite sensitive to image quality variation.

The present research included two tasks designed to assess display readability and legibility. In addition to assessing the effects of the visual display characteristics of touch panels, these tasks included components which were sensitive to touch input characteristics of the devices. By requiring screen touch responses to reading speed and alphanumeric search tasks, both visual and touch characteristics of each device were reflected in user performance measures.

Performance evaluation consisted of two separate experiments. The first evaluated the effects of changes in the visual environment on TSD user performance, focusing primarily on the contribution of image quality to human-TSD interaction. The second experiment investigated the tactile aspects of human-TSD interaction by varying touch resolution requirements of the operator's task.

Experiment One design. In Experiment One, user performance was assessed in a 6 x 3 x 2 factorial within-subjects design. Figure 2 provides a graphic representation of this design.

Experiment One independent variables. Three independent variables were manipulated in Experiment One by combining six touch panels with three levels of ambient lighting and two levels of viewing angle. The resulting matrix generates 36 different experimental conditions.

Panels. The touch panel variable was manipulated by providing observers with two examples of each of three major touch panel technologies described in an earlier section of this report. Capacitive, conductive film and infrared TSDs were evaluated. One "off the shelf" touch panel was purchased for evaluation from each of six different TSD manufacturers. Each vendor was informed of the type of testing and measures that were to be taken and subsequently submitted a design/device felt to be most applicable to this type of evaluation.

Infrared devices were provided by Electro Mechanical Systems, Inc. and Carroll Touch, Inc. Micro Touch, Inc. and Interaction Systems, Inc. provided capacitive TSDs while Sierracin, Inc. and Elographics, Inc. submitted conductive film devices. All were purchased specifically to fit a 3.75 x 7.5-inch flat panel EL display.

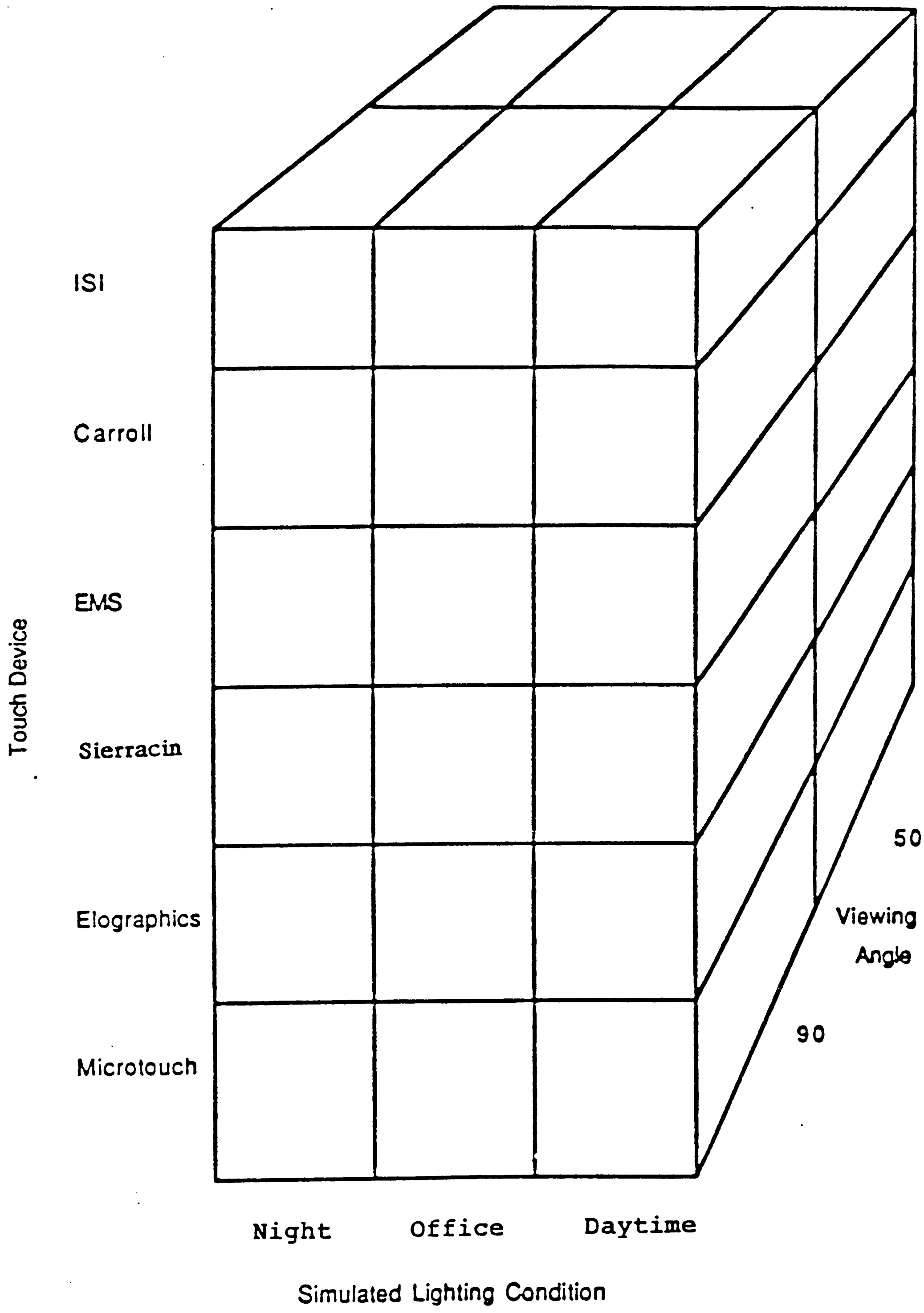

Figure 2: Experiment One Design.

Lighting. Ambient lighting was controlled during the experiment and was varied in the following manner. One lighting condition consisted of a low ambient level with very little light other than the display present in the room.

The additional light was provided by two ceiling-mounted floodlights on dimmer controls. These lights were positioned approximately three feet to the left and right of the subject and directed into the corners of the room at the ends of the wall in front of the subject. This lighting condition allowed the experimenter and participant enough light to see features in the room other than the display. The area of the room behind the subject remained darkened so that very little reflected light reached the subject's eyes from the surface of the display. The illumination of the wall and table area directly in front of the subject (behind the display) was maintained at a constant level of 10 lux. This condition simulated the inside of a vehicle or outside night-time operation.

A second lighting condition had overhead fluorescent room lighting providing an ambient level of illumination at the display of 1291 lux. This condition simulated a normal indoor office setting. The third "daytime lighting" condition included, in addition to normal office lighting, a relatively high intensity diffuse source directed at the panel. This diffuse source consisted of 60- x 122-cm panel of four 40-watt fluorescent light tubes mounted in a white reflector housing. Between this light panel and the subject, at a distance of approximately 50 cm from the fluorescent tubes, was a translucent screen made up of four layers of white mylar. This screen acted as a diffuser, providing a constant luminance background field. The luminance output of this field was approximately 630 cd/m^2 with a maximum variation of 40 cd/m^2 between the dimmest corner and the center of the screen.

The glare screen was positioned above and behind the subject's head as she/he faced the display. The display was positioned at eye level and tilted back 15 degrees from perpendicular. This configuration insured that as long as the subject's eye position was maintained within 1 ft^3, the reflected image of the glare screen filled the display without interference from the subject's head. The luminance of the EL panel display produced by the reflected light of the glare screen, measured from the subject's head position, was approximately 270 cd/m^2. This condition was not intended to simulate precisely direct sunlight conditions, but provided useful information on the susceptibility of each panel to image degradation caused by glare. Viewing angle. The two levels of viewing angle presented to subjects in the study were (1) Straight-on view, in which the subject's line of sight was horizontally 90 deg to the surface of the display, and (2) a 50 deg horizontal condition in which observers viewed the display from a line of sight 40 deg horizontally offset from the perpendicular condition. The inclusion of this variable in the design simulated conditions in which the touch panel may be a secondary display or where multiple users must interact with the device. For both conditions, the display remained tilted back at 15 deg in the vertical dimension.

Experiment One task. Participants were asked to complete multiple trials of a Tinker reading speed test under the set of conditions described above. Previous work has shown the Tinker test to be quite sensitive to subtle changes in the quality of CRT-generated or AC plasma display generated text (Abramson and Snyder, 1984).

The Tinker test was developed to assess readability, and consists of a set of short paragraphs each of which contains one "wrong" word that does not fit the context of the rest of the paragraph (Tinker, 1955). The following paragraph is an example of one Tinker trial.

"Howard could always be found at the printer's shop in his spare time, for he loved to run errands for the staff. We knew he was a baker at heart." The "wrong" word to be identified in this passage is "baker".

A trial of the task begins with the subject-actuated presentation of one Tinker paragraph or passage. The task ends when the subject touches the point on the touch panel which corresponds to the "wrong" word.

The touch target corresponding to the correct selection was a rectangle of 11 mm vertically by 19 mm horizontally. The center of this rectangle always matched the location of the target word, making the Tinker test a "key entry" type task. Tinker passages are fairly simple so that the primary limitation to reader performance of the task is reading speed and physical interaction with the device, not reading comprehension skill.

Tinker passages have been developed so that the completion of each passage requires the same amount of time of a reader if all other factors remain constant. Each trial takes approximately 7 s to complete. Therefore, a relatively large number of Tinker trials could be completed without the subject suffering from fatigue.

<u>Experiment Two design</u>. Experiment two also consisted of a 6 x 3 x 2 factorial within subjects design. Figure 3 provides a graphic summary of this design.

<u>Experiment Two independent variables</u>. The three independent variables manipulated in this experiment were touch panels, viewing angle, and target size. Touch panel and viewing angle were held at the same levels as in

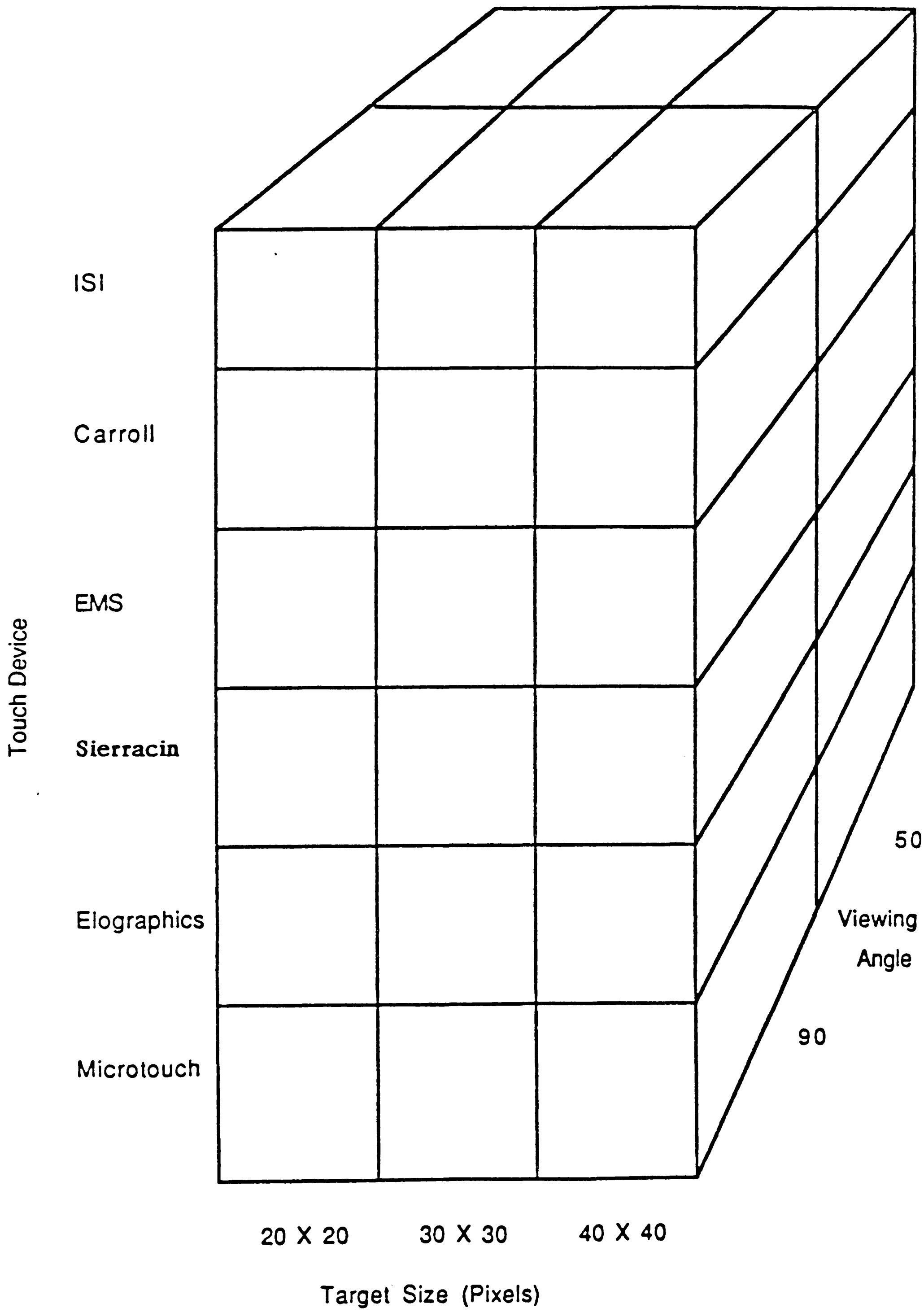

Figure 3: Experiment Two Design.

experiment one. Additionally, the lighting level for all conditions in Experiment Two matched the 1291-lux office condition ("office" condition) from the first experiment. Maintaining these levels across experiments provided a basis for comparison of effects across tasks.

Target size. The size of the target area which registered as a correct response in the Experiment Two task was manipulated as an independent variable. Three levels of target size were selected through pretesting of a range of sizes. Pretesting involved finding the range of target sizes that can be used effectively by an operator. This range was determined by experimentally varying touch target size while requiring subjects to make accurate selections of targets. The range of target sizes selected for possible inclusion in the study was from 6 mm^2 to 15 mm^2. The bottom end of the range was set at the smallest size at which all pretest subjects achieved an 80% hit rate. The upper end of the range was selected by referral to the literature on optimal TSD target size and subsequent testing to insure that all pretest subjects achieved a 100% hit rate at this size. An Elographics resistive film TSD with a limiting resolution of 5 touch points per mm was used in pretesting.

The three target sizes selected for inclusion in the study were 7.5 mm^2, 11 mm^2, and 15 mm^2. These corresponded to 20 x 20 pixels, 30 x 30 pixels, and 40 x 40 pixels on the electroluminescent display.

Experiment Two task. The second task was an alphanumeric search and point task. In this task each trial consisted of a presentation of a display page containing one randomly placed target element among a group of detractor or noise elements. The trial terminated with the subject touching the point on the panel corresponding to the location of the target element.

The location of the displayed target element was randomly assigned, excluding a boundary area within 10 pixels of the edge of the display, and not necessarily corresponding in a one-to-one fashion with touch element boundaries. This random placement made the alphanumeric search task a target selection type task (see Introduction).

Procedure. The within-subjects design employed in both experiments required each subject to complete six 1-hour testing sessions. These sessions were scheduled on six consecutive days at the same time each day. Four subjects were scheduled per day resulting in a three-week period of performance data collection.

On each day of testing, each subject completed the session using one TSD. Order of presentation of the 6 TSDs to the 12 subjects was counterbalanced, thereby requiring two changes of TSD-display hardware per day.

The daily session consisted of blocks of trials of both Tinker and Alphanumeric search tasks. These tasks were separated so that one half of the session consisted of Tinker trials and the other half search trials. The order of tasks was randomly selected on the first day and alternated on subsequent days.

Within each task, seven blocks of 12 trials each were presented. The first block of trials for each task was designated as practice and these data were discarded. The remaining six blocks were presented under the different lighting, viewing angle, and target size conditions included in the study.

Before each block of trials approximately three minutes of rest was allowed while the experimenter set up conditions and the graphics generator produced the stimuli for the next block. This period included a minimum of two minutes during which the subject adapted to the the lighting conditions for the next block. In addition to this adaptation period, extra adaptation was provided prior to high glare blocks. The subject was instructed to focus on the inter-stimulus display, which contained the word **READY** at the center of the screen, for 15 seconds. This procedure provided additional adaptation to the low level of character-to-background contrast produced by the glare.

Each trial of either task was self paced, being initiated by a push button which the subject held in the nondominant hand. Prior to the start of each trial the display screen contained an intertrial message which said either "READY" or "THE NEXT TARGET LETTER IS *", where * was one of the 26 letters of the alphabet. Which message was presented depended upon the task.

At the beginning of a trial the subject's dominant hand rested on a board on the table in front of him or her. This board contained an interrupted beam switch which was activated any time the dominant hand was moved from its start position. Subjects were instructed to leave their dominant or pointing hand at the start position until the target had been identified and they were ready to point to it. When the target had been identified, the dominant hand was used to touch the screen, selecting the target.

At each touch of the screen, the computer responded with a beep. When the coordinates of the touch fell within the target area, the beep was followed immediately by a return to the interstimulus display. Subjects were instructed to continue to make touches until the target disappeared in the event the first touch was erroneous.

Before the first day of testing, each subject received instructions on the use of the apparatus and what responses they were expected to make to the trial stimuli. These instructions are included in Appendix A. At this time, 12 typewritten examples of Tinker passages were also given to subjects and these were read over by the experimenter and subjects with the subject picking out the "wrong word" from each example. This extra practice was provided because pretesting showed a marked learning curve over the first several trials of Tinker passages as subjects became used to the pattern of the task.

Apparatus. The apparatus used in performance evaluation consisted of eight major components. Six of these components were linked in a computer controlled stimulus presentation and data collection loop. The other two components were hardware for manipulating independent variables. These components were;

1. PDP 11/55 computer
2. ADAGE 3000 graphics generator
3. PLANAR Electroluminescent display
4. Six touch devices and associated controllers
5. Lab Peripheral System (LPS) (computer add-on)

6. HP 2640 computer terminal
7. Glare producing light panel
8. TSD/display mount and housing assembly

Figure 4 is a block diagram of these components and their interrelationships.

The central element in the computer-controlled portion of the apparatus was the PDP 11/55. This computer contained the program which controlled the sequence of events in the experiments. It also received and stored the data collected from each TSD. The PDP 11/55 also controlled and supported the experimenter's I/O terminal, the ADAGE graphics generator, and the LPS system.

Tinker passages, search task patterns, and all interstimulus display screens were generated by the Adage 3000 graphics generator. The Adage system also produced all the driving signals necessary to display these images on the electroluminescent flat panel.

All of the stimulus and interstimulus presentations to the subject during the experiments were made via a PLANAR electroluminescent (EL) display. This display was a discrete element flat panel device with a 512 x 256 element resolution. Active screen dimensions were 19 x 9.53 cm. The display was monochromatic amber and was refreshed at 60 hz.

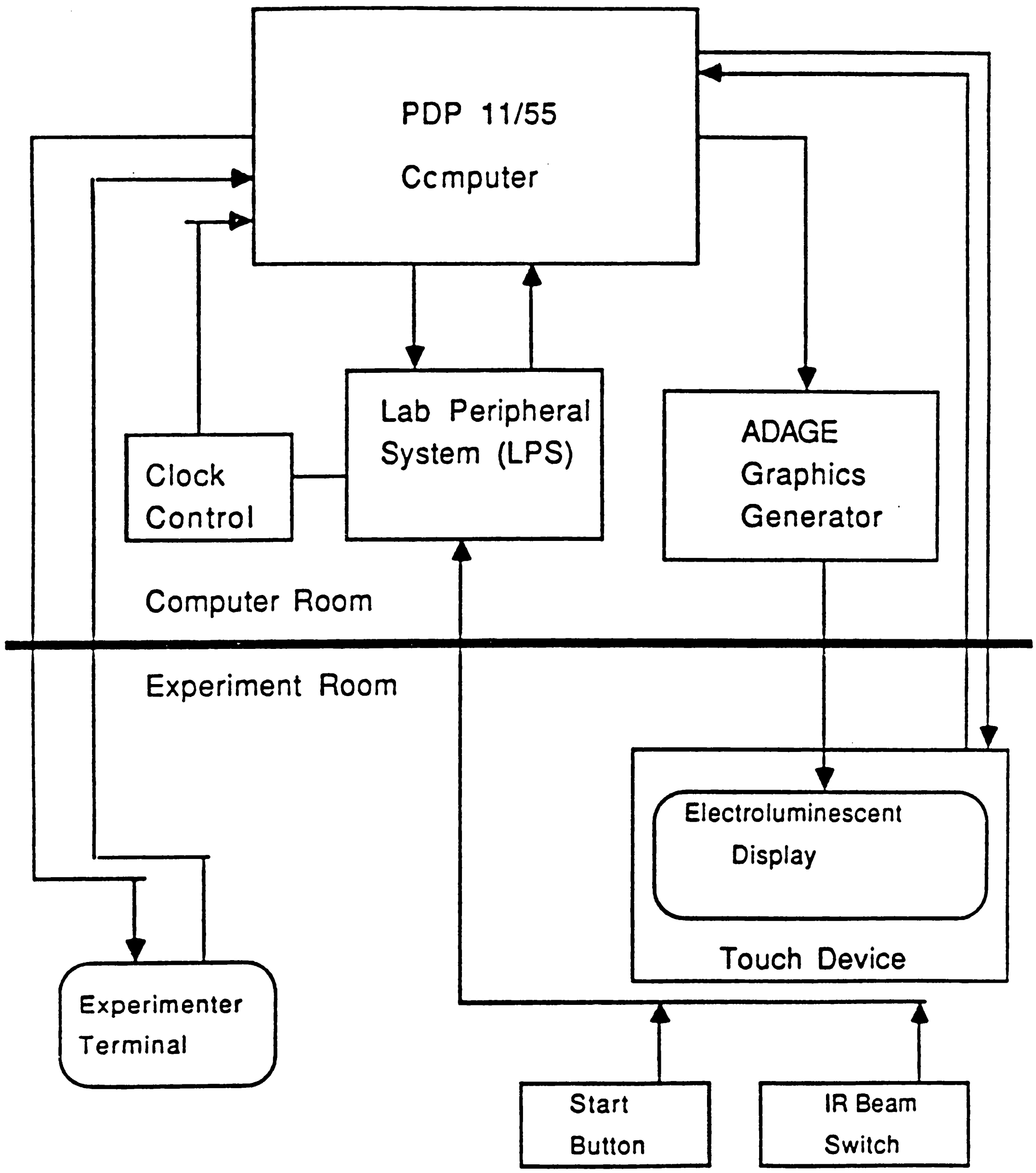

Figure 4: Block Diagram of Experimental Apparatus.

Each of the six TSDs evaluated in the study were mounted on the same EL panel during performance testing. Each TSD with the exception of the Sierracin device had its own controller and was linked with the PDP 11/55 via an RS232 serial interface. The Sierracin device communicated with the PDP 11/55 through a 16-bit digital I/O controller and parallel port that is part of the LPS system.

All six TSDs were set up to operate in a single touch mode in which only one set of X and Y touch coordinates was transmitted each time the device was touched and not again until the touch had been removed. In the case of IR beam devices, the touch was recorded when and where the beam grid was initially broken and could not record another set of coordinates until the object breaking the beam was removed. Subjects were informed before the experiment that only one point was recorded per touch and that they must remove their finger after each attempt in order for the devices to reset. Many of the devices offered other alternatives, such as continuous output or release modes, but for the sake of consistency in the comparisons only single touch mode was evaluated.

The LPS system, in addition to the parallel communication capabilities mentioned above, contains a real time clock and event timer, program controlled electronic relay switches, and analog-to-digital conversion channels.

This system provides sensing and analog device control capabilities to the PDP 11/55. In this study, the LPS provided the clock timing for the time-based dependent measures. The LPS also provided a two-way communication link between the PDP 11/55 and the subject by receiving inputs from the start button and interrupted beam switch and sending feedback via an auditory beeper.

The two noncomputer elements of the experimental apparatus were the light panel used to produce the high glare lighting condition and the the mechanical mounting and housing used to integrate the six TSDs with the EL panel display. The light panel is described in the section on the lighting independent variable.

The mechanical device for TSD/display integration consisted of an aluminum frame, brackets for mounting each TSD, and a plastic and poster board bezel and facade. The frame and brackets held the TSD and EL panel in contact with each other in a fixed relationship such that TSDs could be removed and remounted in the same place. When mounted, the TSD/display unit was fixed at a height 30 cm above the table and an attitude 15 deg from vertical such that the display was at eye height and the top tilted away from the subject.

The bezel and facade assembly was mounted to the frame between the subjects and display. It was flat black in color and insured that the only visible difference from one TSD/display mounting to another was the appearance of the surface of the individual TSD.

Subjects. Twelve volunteers, four female and eight male, participated as subjects in this study. All participants were screened with a Bausch and Lomb Orthorater for 20/22 or better corrected near visual acuity.

Dependent variables. The primary dependent performance measures collected during the performance evaluations were task time and number of touches. Two task time measures were taken for each trial of the Tinker and search tasks. These measures were reading/search time and entry time.

Reading or Search time was defined as the time elapsed from the presentation of the target stimulus until the subject decided on the correct response. Reading time was measured by having the subject rest his/her dominant hand on an interupted beam switch prior to each trial. When the hand was moved to touch the screen, the switch was activated and the elapsed time recorded.

Entry time was measured as the elapsed time between hand switch activation and acceptance of a correct touch by the TSD.

Feedback regarding the accuracy of each subject's response was provided during experimentation by having the screen go blank after the correct target was touched. The target remained in view until an accurate touch was made.

This type of feedback allowed learning to take place and forced the impact of touch characteristics such as parallax and touch resolution into the task time dependent measure, since the timer did not stop until a correct response was made. If no feedback were provided, subjects would not be aware of the accuracy of their touch inputs, and no change in task completion time could be expected among conditions.

Hardware Parameters Evaluation

From the review and discussion of relevant literature presented in the introduction to this dissertation, several physical characteristics of the user-TSD interface were selected for measurement on each of the six candidate TSDs. These measurements fell into two categories; those based on the visual image quality of the display as seen through a TSD, and measures based on the TSD's response to touch inputs from the user. A description of each of these measures of TSD hardware parameters follows.

Image quality. A microphotometric measurement system was used to determine the image quality produced by each

TSD integrated with an electroluminescent display panel. The results of the microphotometric measurements were incorporated into a number of metrics that describe image quality of the touch devices.

Techniques for making microphotometric measures on flat panel displays and transforming them into image quality metrics have been previously established (Beaton, Schulze, and Snyder, 1983; Schulze and Snyder, 1983). Diffuse and specular glare measures have also been recently developed (Beaton and Snyder, 1984; Hunter, Reger, Farley, and Snyder, 1985). The present study employed these measures and procedures and extended their utility by describing their relationship to human operator performance on flat panel TSDs.

The microphotometric system used in this study was produced by EG&G Gamma Scientific and consisted of a GS-2110A electromechanical scanning telemicroscope and a GS-4100 intelligent radiometer. The scanning system was driven by an IBM/XT Personal Computer. The telemicroscope was fitted with a 1X objective lens, a D-46 photomultiplier detector, a D-46 photopic correction filter, and a 0.010 x 1.0 mm scanning aperture.

This scanning system was used to make line spread function (LSF) measurements. To make these measurements, the scanning aperture was moved at 0.010-mm steps across a maximum contrast line produced on the EL display. Luminance as a function of aperture position was digitally recorded on the IBM/XT. This procedure produced a luminance profile of the line that was scanned.

Both horizontal and vertical luminance-by-distance scans were performed, and upon completion of these measurements data files were transferred to an IBM 370 computer system for analysis. This analysis consisted of a Fast Fourier Transform (FFT) on the LSF, yielding a modulation transfer function (MTF).

The modulation transfer function describes a luminance contrast modulation by spatial frequency curve. Any point on the function represents the contrast that can be achieved by the display while producing a sine wave equivalent luminance intensity pattern of a certain spatial frequency. Typically this function starts with high modulation at low spatial frequencies and drops off at higher frequencies. For a more detailed description see Cornsweet (1970).

The integral of this MTF and a number of other metrics that can be derived directly from this calculation were used to describe the image quality associated with the touch devices. These metrics are defined and further described below.

1: Area Under the System MTF. The luminance-by-distance profile of a line produced by the EL panel and measured through the touch device was subjected to a discrete Fourier analysis to obtain the MTF, after which the following integral was obtained:

$$\text{Area under MTF} = \int_{0}^{Ny} M(f)\ df. \qquad (1)$$

where M(f) is the MTF at spatial frequency f in cycles/mm, and Ny = the Nyquist sampling limit.

2: MTFA. The visually weighted area under the MTF is defined as the area bounded by the system MTF and the contrast sensitivity function (CSF) of the human observer under existing display conditions. The MTFA has been shown to be an effective measure of physical image quality that correlates well with perceived image quality (Snyder, 1973; 1980).

$$\text{MTFA} = \int_{0}^{Ny} [M(f) - T(f)]\ df. \qquad (2)$$

where T(f) is the contrast threshold function of the visual system. This function was represented in this study by

$$CSF(F) = (7.65463 \times 10^{-4})(e^{.166404f})$$

which approximates the 90% population CSF.

3. SSF: The variance or squared spatial frequency (SSF) measure is the system MTF weighted by the square of the spatial frequency, having been integrated over all spatial frequencies. The SSF is defined mathematically as:

$$SSF = \int_{0}^{Ny} [M(f) \quad (f^2)] \quad df. \tag{3}$$

4. EP: The equivalent passband (EP) is the two-dimensional area under the sine-wave response or MTF, and is defined as:

$$EP = \int_{0}^{Ny} [M(f)]^2 \quad df. \tag{4}$$

These four metrics of image quality are based on the image resolving capabilities of a light emitting display at its output. They deal with the ability of a display system to pass information in the form of spatial luminance variations of different frequencies. These metrics can be thought of as measures of the signal of a display.

As with any information transmitting device, electroluminescent displays (especially with TSD overlays) produce a certain amount of noise or spatial luminance variation that is not part of the signal. This noise may be caused by imperfections or irregularities in overlay materials or by patterns of reflective conductors that must be present in some TSDs.

Whatever their cause, measures of the visual noise inherent in each TSD system are important. The following three measures of display noise were reported in this study.

1. WS: The area under the Wiener spectrum (WS) has been established as a meaningful measure of the visual noise inherent in TSD display, (Beaton, Schulze, and Snyder, 1983; Schulze and Snyder, 1983) and is mathematically defined as:

$$WS = \int_{0}^{Ny} |w(f)|^2 \, df. \qquad (5)$$

in which

$$w(f) = \sum_{x=0}^{N-1} f(x)[\cos(2 \quad f/N) - j \sin(2 \quad f/N)]. \qquad (6)$$

Procedurally, the Wiener spectrum was measured with the same microphotometric system described above. Initially, scans were made of a calibrated standard luminance source. Scans of the standard source represented the display component of the display-TSD system. Next, a TSD was placed between the standard source and microscope at the focal distance of the scope and another set of scans was made. For each scan a mean luminance was computed for this scan and subtracted from each observation in the scan, resulting in a complex luminance waveform profile centered at zero. A fast Fourier transform was taken on this waveform and the area under the resulting modulation by spatial frequency function was weighted to produce the Wiener spectrum.

3. RMS Luminance: The root-mean-square (RMS) luminance is another visual noise measure that will be useful in the present application.

RMS luminance was computed from the same luminance profile scans made for the Wiener spectrum. A standard deviation was computed for each of several horizontal and vertical scans taken on each device. A mean of these standard deviations was the RMS luminance for each TSD.

Glare. An analogous measure to the area under the MTF, the area under the reflectance transfer function (RTF), takes into account the first surface reflectivity of the touch device (Beaton and Snyder, 1984). RTF represents the amplitude by spatial frequency function of reflected or specular glare. At each spatial frequency of interest, the relationship is as follows:

RTF = I amplitude / 0 amplitude,

where:

I amplitude = amplitude of the reflected image, and 0 amplitude = amplitude of the luminous slit.

Like the MTF, the RTF is based on fast Fourier transforms of line spread functions measured by a scanning photomicroscope. The primary difference is how the line that is scanned is produced. In the RTF, a scan is first made of a 50 micron by 10 mm slit placed in front of a high intensity, collimated light source. The FFT of this LSF becomes the numerator of the RTF. The denominator of the RTF is acquired by making a scan of the image of the slit, reflected, at an angle equal to the angle of incidence, from the first surface of a TSD. The FFT of this scan is the denominator of the RTF.

The FFT outputs of the luminous slit used to produce the reflected image are divided frequency-by-frequency by those outputs from the FFT of the reflected image. The area under this resultant RTF produces an indicator of the dispersion of the first surface reflections of the TSD.

Signal-to-noise ratios. The four measures of signal (MTF area, MTFA, SSF, and EP) and three measures of noise (WS, RMS luminance, and RTF area) can be combined to yield 12 possible signal-to-noise ratios for each of the touch devices. Each possible combination was computed and included in the potential predictors of TSD performance.

Ambient MTF. A final measure of image quality taken in this study includes the measurement of signal and noise simultaneously. The ambient MTF consists of an MTF based on an LSF taken under lighting and viewing angle conditions of interest. In this case, LSFs were taken with the microscope placed on the line of sight of a hypothetical subject in the study. Scans were made of each TSD/display combination at the two viewing angles and with the three levels of lighting included in the performance evaluation. Thus, the ambient MTF reflects the effects of off-angle viewing and diffuse glare on image quality.

Parallax. Parallax in general refers to the observation that objects at different distances from an observer appear to change in relative position if viewed from different angles. As mentioned in the introduction, no operational definition of parallax in TSDs has been offered in the literature. Therefore, two measures of parallax were included in this evaluation. Each measure shares the property of changing with relative separation of objects as viewpoint is altered. Figure 5 provides a graphic representation of these measures.

Each measure is based on the relationship between the phosphor plane of the EL display and the activation plane

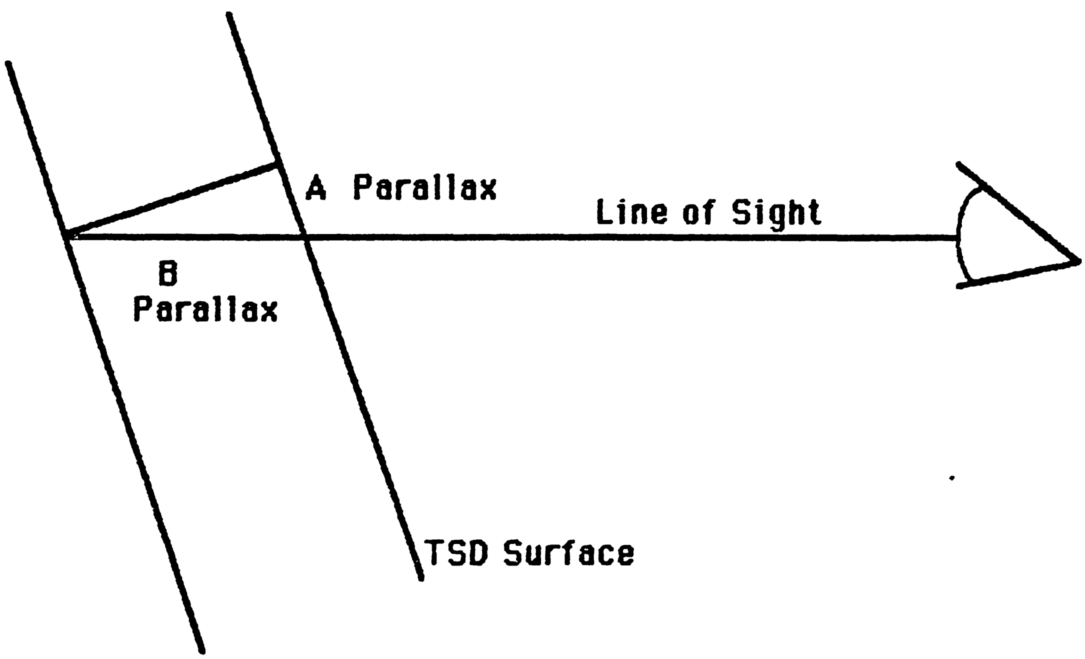

Figure 5: Parallax Geometry.

of the associated TSD. Both measures take into account the relative separation of these planes and the angle from which the display is viewed.

The first measure, called parallax A for lack of a better term, is a distance measured between two points in the sensor plane of the device. This distance is measured between the point on the TSD sensor plane which is directly over the target and another point on the sensor plane at its intersection with the line of sight between the target and the operator. Parallax A is zero when the line of sight is perpendicular to the display and sensor planes; otherwise, it has some positive value which is a function of viewing angle and the distance between the phosphor and sensor planes.

Parallax B was measured as a distance along the line of sight of the observer to the display target. This distance will be bounded by the display target at one end and the intersection with the sensor plane at the other end. This measure of parallax is also a function of the distance between the phosphor and sensor planes of a TSD and the angle at which the device is viewed. It has a value equal to the distance between the two planes when viewed on a line perpendicular to them. The value of parallax B will increase under off-angle viewing conditions.

Parallax A and B can be seen in Figure 5 to be the opposite and hypotenuse sides of a right triangle described by the display target, its corresponding touch target, and the point on the line of sight intersecting the touch plane.

Touch characteristics. Touch characteristics measured included touch pressure and panel resolution. Limiting panel resolution (LR) was computed from the published specifications provided with each device and verified in our laboratory. Touch force was measured by placing each device on a balance scale. Weight was then added to the opposing side of the scale while the balance was held stationary by a finger on the TSD. The force required to activate the device was recorded and averaged over several such trials.

Effective resolution. Pretesting of the six TSDs established that a measure of touch resolution based solely on the number of coordinates produced in X and Y was insufficient. Elographics, Microtouch, and ISI TSDs differed from the other three devices in that they did not have mechanically fixed locations for touch coordinates. These "analog" devices each exhibited a certain degree of variability and nonlinearity in coordinates reported from a fixed pattern of touches. Errors in coordinate assignment reduced the effective touch resolution capabilities of these devices below the limiting resolution built in to them. Therefore, a measure of resolution which took into account device variability and nonlinearity was developed. This measure is called effective resolution.

The effective resolution of each TSD was established statistically from a number of measurements taken at fixed locations. Twenty touches at each of 10 locations were made and X and Y coordinates recorded for each touch. Means were computed for X and Y at each location. The difference between the mean value and the design or expected value in each axis at each location was then found. This difference is a measure of nonlinearity at a location.

Variability of touch reports was assessed by computing standard deviations for the X and Y axes at each location. Adding the nonlinearity measure in X and Y at each point to its respective variability measure resulted in an estimate of the effective resolution of the device at that point on the screen. An overall measure of device effective resolution (ER1) was established by computing means in X and Y directions of the ten test locations then averaging these two means.

An additional estimate of effective resolution was computed by using two times the standard deviation of X and

Y at each point as the measure of variability in the computation of ER. This measure was named ER^2.

Performance Modeling Evaluation

Results from operator performance and hardware evaluations were combined in an evaluation of performance models. The purpose of this evaluation was to predict operator performance with linear combinations of hardware measurements.

The operator performance evaluation included three dependent variables: (1) reading/search time, (2) entry time, and (3) touch count. Each of these measures was also included in the modeling evaluation. An additional performance measure was created for modeling by adding reading/search time to entry time to get total task time.

Hardware measures used in modeling included all touch characteristics and image quality measures. Each signal strength and noise metric was included as a single predictor as was every signal-to-noise ratio created from their combination.

The primary step in modeling was to evaluate each hardware measure with simple correlations and single predictor linear regressions. From these evaluations one touch resolution and one parallax measure were selected for further model construction. Limiting selection to one resolution and one parallax measure was done to avoid having many highly correlated predictor variables in models. The parallax and ER measures with the highest correlations with performance measures were included in model selection pools.

Following single predictor evaluations, all remaining hardware measures were subjected to a stepwise linear regression model selection procedure. The goal of model selection was to maximize the fit of hardware and performance measures. Toward this goal, stepwise selections were performed on several alternative data configurations.

Initially, stepwise selection was performed on all four performance measures with all hardware measures as possible predictors. A further attempt to make gains in model fit was to transform time measures of performance into speed measures by computing their inverses. This transformation did not produce any gains in model fit. Finally, models were constructed with subsets of data from each task.

PERFORMANCE EVALUATION ANALYSES AND RESULTS

Data Collection and Reduction

Collection. In both user performance experiments data were collected by the PDP 11/55 computer during each trial. Following each trial, these data were stored on a magnetic disk for later reduction and analysis.

At the start of each trial, as the subject pushed the start button, a 100-Hz clock counter was initialized and started. The value in this clock was read and stored twice during each trial. The first clock reading occurred when the subject's hand was lifted from the start position. The second reading occurred immediately after the final touch was reported for a trial. The values of these readings represent reading or search time and total task time.

Also collected during each trial was the number of touch attempts required before a touch inside the boundaries of the target area was recorded. If no hit of the target occurred within 13 attempts by the subject the trial was terminated. In the case of a computer terminated trial, the touch count variable was stored as 13 and the total task time reflected the time elapsed up to the 13th touch.

The number of these time outs was less than 0.5% of all trials. Time out trials occurred most frequently at off angle viewing conditions in the search task when target size was small. In this condition, the occurrence rate was less than 5% of all trials. Data from time out trials were recorded and analyzed along with all other successful trials.

Certain types of errors by the subject resulted in rejection and subsequent deletion of data from that trial. Trials were rejected if an incorrect word or letter was identified as the target. Another cause for rejection was if the subject moved his/her hand from the start position before he/she was ready to select a target. Finally, trials were rejected if for any reason the subject stopped during the middle of a trial to do something away from the display such as ask the experimenter a question or adjust the height of the chair, etc. Rejections of trials were evenly distributed across conditions; however, some subjects were more prone to committing these errors than others. The percentage of rejected trials was about 1% overall and 2% for the worst subject. Analyses were conducted on cell means only, therefore the loss of observations from a relatively few trials did not result in any missing cells from the experiments.

Reduction. The first modification that was made to the raw data from the performance evaluation was to subtract the reading or search time of each trial from the total task time. This resulted in the two measures of reading time and entry time. After this modification, a mean was computed for reading/search time, entry time, and touch count across the trials in each block of trials. This reduction resulted in a total of 432 values for each dependent measure including one for each subject at each of the 36 unique conditions for each of the two experimental tasks. All subsequent statistical analyses were carried out on the reduced data set made up of these means.

PERFORMANCE EVALUATION RESULTS

Tinker Reading Speed Experiment

Analyses of the Tinker task data began with analyses of variance for each of the three dependent measures: reading time, entry time, and number of touches. Significant findings from the ANOVAS were further investigated with simple effects F tests and Newman Keuls comparisons between condition means. For all ANOVAs, simple effects tests and Newman Keuls comparisons, the criterion for rejection of the null hypothesis was set at $p < .05$.

Dependent variable reading time. Because all F tests were statistically significant, the major finding from an analysis of variance of reading times in the Tinker task was a significant three-way interaction among device, lighting, and viewing angle ($F_{10,110} = 7.16$; $p = .0001$). This interaction was further investigated with a series of simple-effects F tests at each of the three levels of lighting. Two-way tests at each level of lighting showed a significant device by viewing angle interaction at the daytime light level ($F_{5,110} = 25.8491$, $p < .01$) and no significant two-way interactions under office or nightime conditions.

Additional one-way simple effects F tests for differences among devices at each level of viewing angle at office and nighttime light levels revealed no significant findings. ANOVA and simple effects summary tables for reading speed data are presented in Tables 1, 2, and 3.

Newman-Keuls multiple paired comparisons were carried out among the 12 devices by viewing angle combination means at the high glare condition to find the specific source of the significant three-way interaction. Results of these comparisons are summarized in Table 4. These results show that the pattern of significant differences among the six devices at the 90-deg viewing angle differs from the pattern of differences among devices at the 50-deg viewing angle.

At the 50-deg angle, the difference between the ISI and EMS devices was significant as were the difference between Elographics and Microtouch and the difference between the Sierracin and Microtouch devices. The difference between Sierracin and ISI is not significant. At the 90-deg viewing angle this pattern is reversed with the ISI - EMS, Elographics - Microtouch, and Sierracin - Microtouch comparisons all nonsignificant while the Sierracin - ISI difference becomes significant.

The magnitude and direction of the effects involved in the three-way interaction for the dependent variable "reading time" can be seen in Figures 6, 7, and 8 which provide graphic representations of the device by viewing

TABLE 1. ANOVA Summary Table for Tinker Task Reading Times (Experiment 1)

Source	df	MS(sec/100)	*F*	*p*
<u>Between Subjects</u>				
Subjects (Sub)	11	705947.93		
<u>Within Subjects</u>				
Device (Dev)	5	120965.84	9.70	.0001
Dev x Sub	55	12474.02		
Lighting (Lt)	2	2884577.74	43.40	.0001
Lt x Sub	22	66463.23		
Angle (An)	1	381102.929	29.12	.0002
An x Sub	11	13087.20		
Dev X Lt	10	80369.38	16.29	.0001
Dev x Lt x Sub	110	4934.70		
Dev x An	5	34057.72	7.79	.0001
Dev x An x Sub	55	4372.31		
An x Lt	2	227826.46	24.93	.0001
An x Sub	22	9136.95		
Dev x An x Lt	10	18622.43	7.16	.0001
Dev x An x Lt x Sub	110	2602.07		
Total	431			

TABLE 2. Simple Effect F Tests for the Tinker Task Reading Time Measure: Tests of Device by Angle Interaction at Levels of Lighting

Light Level	MS (s/100)	*F*	*p*
Daytime	67261.306	25.849	$< .01$
Office	2051.182	.788	$> .25$
Nighttime	1990.095	.765	$> .25$

TABLE 3. Simple Effect F Tests for the Tinker Task, Reading Time Measure: Tests for Effects of Device at Levels of Lighting by Angle Interaction

Light Level	Viewing Angle	MS (s/100)	*F*	*p*
Daytime	90°	55102.415	21.176	< .01
Daytime	50°	284994.178	109.526	< .01
Office	90°	1898.872	0.7298	> .25
Office	50°	5325.824	2.05	> .05
Nightime	90°	3818.158	1.467	> .10
Nightime	50°	1867.750	0.718	> .25

TABLE 4. Newman-Keuls Comparisons for the Tinker Task Reading Time Measure: Comparing Device by Angle Interaction Means at Daytime Lighting Level

Device	Viewing Angle	Mean	Comparisons
Carroll	90°	6.406	A
EMS	90°	6.787	B
Carroll	50°	7.120	B C
ISI	90°	7.286	B C
EMS	50°	7.362	B C
Elographics	90°	7.582	C D
Microtouch	90°	8.026	D
Sierracin	90°	8.108	D
Sierracin	50°	8.945	E
Elographics	50°	9.143	E
ISI	50°	9.320	E
Microtouch	50°	11.371	F

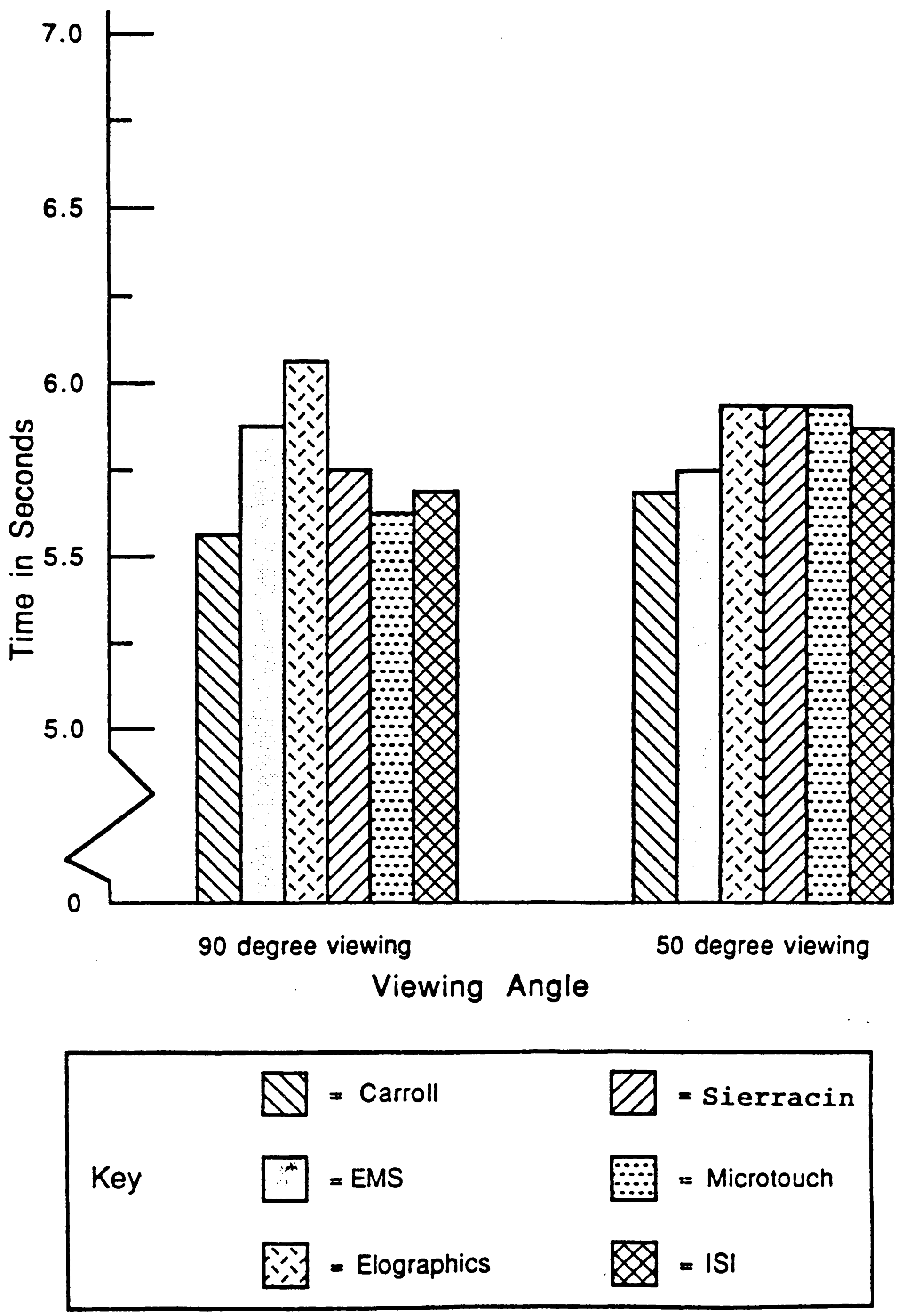

Figure 6: Tinker Task Reading Time Performance,
Device by Viewing Angle Interaction at Night Light Level.

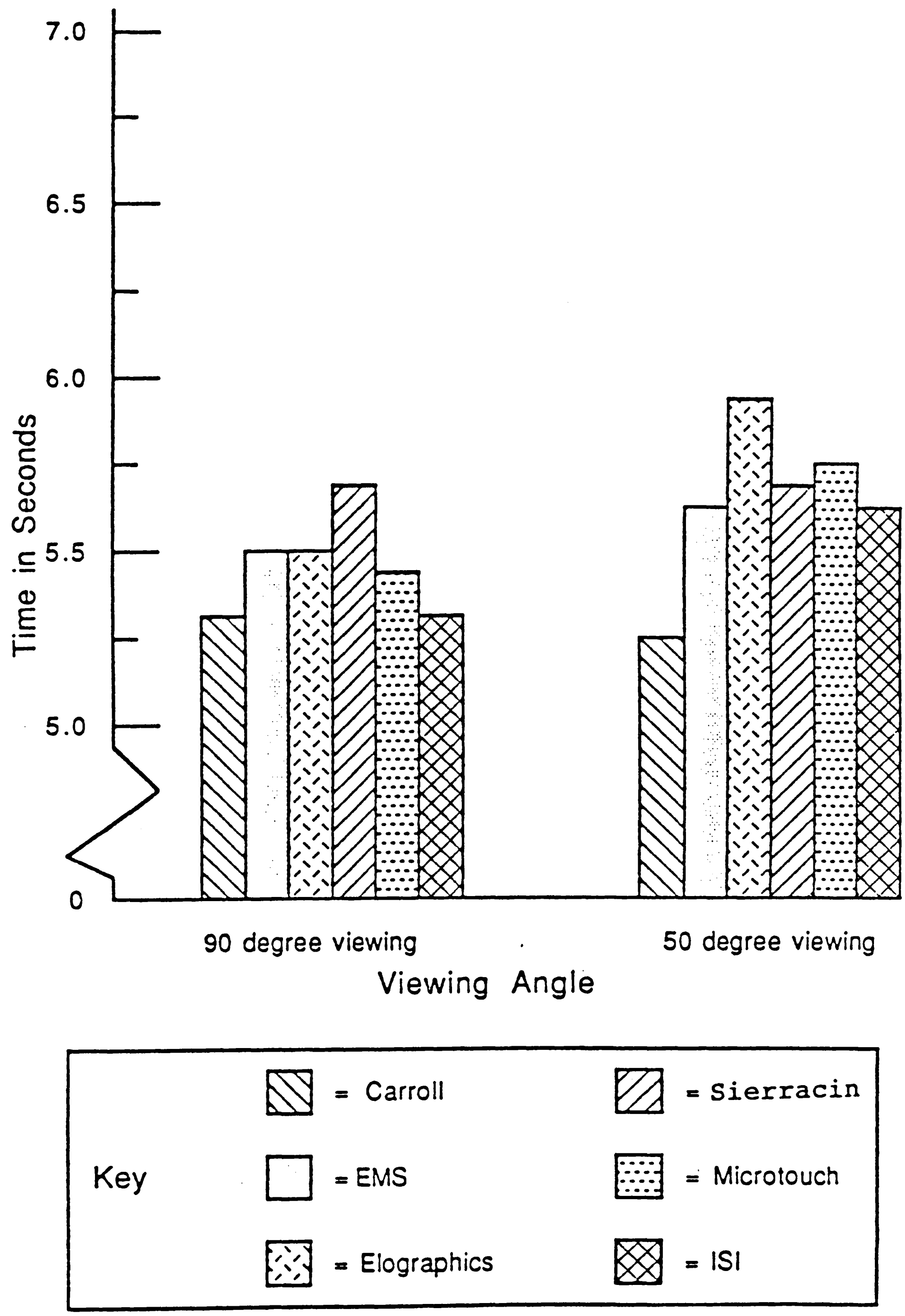

Figure 7: Tinker Task Reading Time Performance, Device by Viewing Angle Interaction at Office Light Level.

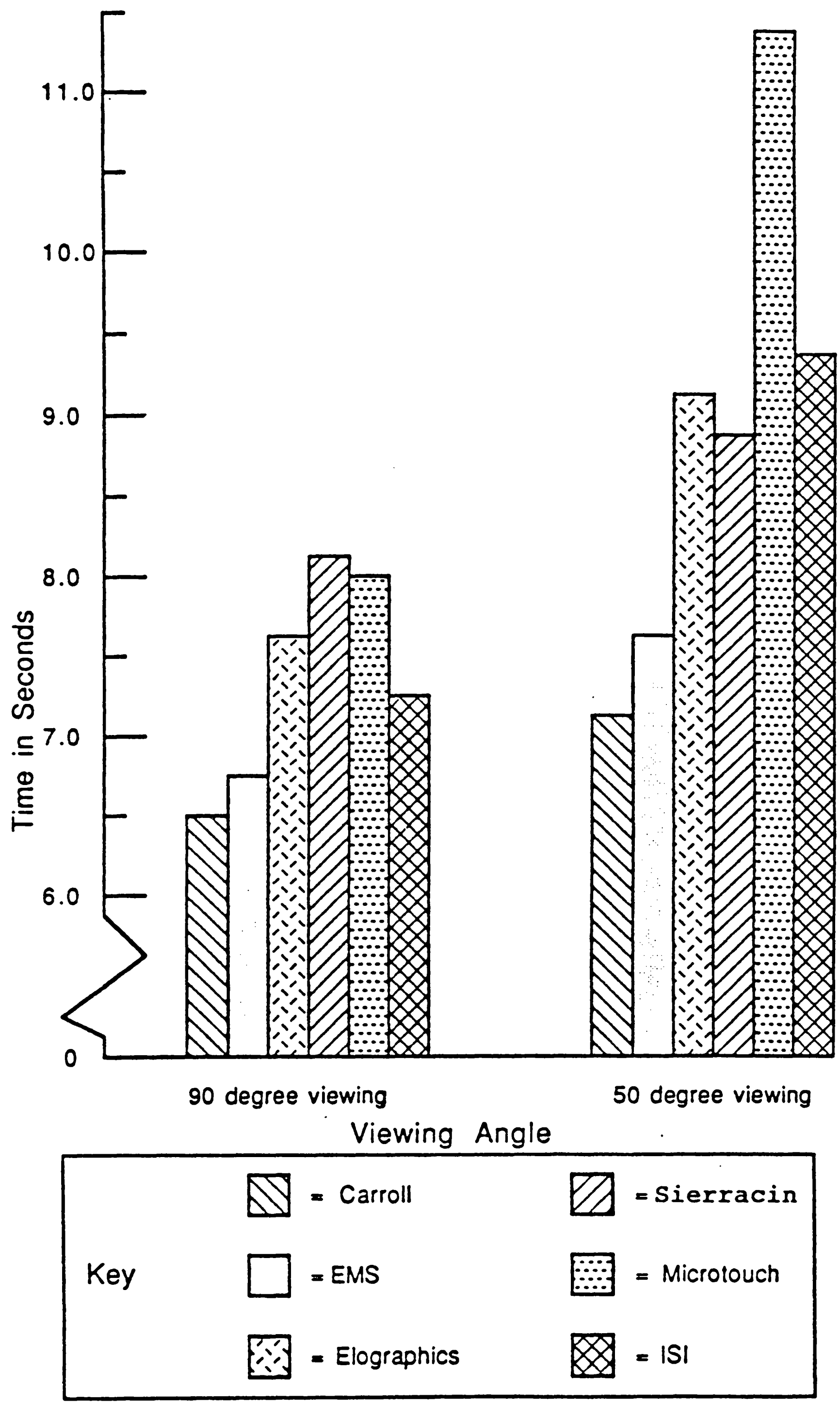

Figure 8: Tinker Task Reading Time Performance,
Device by Viewing Angle Interaction at Daytime Light Level.

angle treatment combination means at each level of lighting.

These figures and the significant findings reported above show that at nighttime and office conditions, it makes little difference, in terms of reading speed, which of the six devices is used.

Under daytime ambient conditions, visual differences between devices begin to significantly affect reading performance, and these differences are further exacerbated by adding off-angle viewing to daytime lighting. The additional impact of off-angle viewing is most pronounced in the Microtouch and ISI panels, while Carroll, EMS, and Sierracin TSDs are less affected. Across all levels of lighting and viewing angle, the Carroll and EMS devices consistently have the shortest mean reading times.

Dependent variable entry time. An analysis of variance of the entry time data, summarized in Table 5, uncovered significant main effects of device type ($F_{5,55}$ = 6.41, $p < .001$) and light level ($F_{2,22} = 4.93$, $p < .02$), with no significant interactions. Newman-Keuls comparisons were completed for group means under the two significant main effects.

The comparisons for the main effect of device are included in Table 6. These data indicate that Carroll Touch and EMS TSDs have significantly faster entry times

TABLE 5. ANOVA Summary Table for the Tinker Task Entry Times (Experiment 1)

Source	df	MS (s/100)	*F*	*p*
<u>Between</u> <u>Subjects</u>				
Subjects (Sub)	11	4950.06		
<u>Within Subjects</u>				
Device (Dev)	5	7553.22	6.41	.0001
Dev x Sub	55	1177.93		
Lighting (Lt)	2	5408.71	4.93	.0170
Lt x Sub	22	1097.18		
Angle (An)	1	168.21	.29	.6020
An x Sub	11	583.34		
Dev x Lt	10	789.15	1.17	.3209
Dev x Lt x Sub	110	676.34		
Dev x An	5	449.47	.87	.5100
Dev x An x Sub	55	519.13		
An x Lt	2	869.58	2.11	.1455
An x Lt x Sub	22	412.78		
Dev x An x Lt	10	187.11	.34	.9680
Dev x An x Lt x Sub	110	549.55		
Total	431			

TABLE 6. Newman-Keuls Comparisons for the Tinker Task

Entry Time Measure: Comparing Device Means

Device	Mean (s)	Comparisons	
Carroll	0.6529	A	
EMS	0.6719	A	
Elographics	0.7267	A	B
Microtouch	0.8185		B
Sierracin	0.8650		B
ISI	0.8682		B

averaged across subjects, light levels, and viewing angles than the Microtouch, Sierracin, and ISI devices. Elographics' mean entry time falls between these two groups and is not significantly different than any of the other devices.

Newman-Keuls comparisons for the main effect of light level, seen in Table 7, showed no significant differences between light levels at the $p < .05$ criterion level.

Dependent variable touch count. Summary information for the analysis of variance of touch count data is included in Table 8. One significant main effect, device type, resulted from this analysis ($F_{5,55} = 8.9$, $p < .001$). A breakdown of this effect can be seen in Newman-Keuls comparisons and condition means in Table 9. The comparisons found the Sierracin TSD to require significantly fewer touches per target acquisition than any other device. Next best in terms of touch count were the Elographics and EMS TSDs, which were not found to differ significantly from each other. The ISI device was significantly poorer than the Elographics panel but no different than the EMS TSD. Carroll Touch's device required significantly more touches on the average than any of the above mentioned devices but was significantly better than the poorest device, which was the Microtouch panel.

TABLE 7. Newman-Keuls Comparisons for the Tinker Task Entry Time Measure: Comparing Light Level Means

Light Level	Mean (s)	Comparisons
Office	0.7346	A
Daytime	0.7716	A
Nightime	0.7954	A

TABLE 8. ANOVA Summary Table for the Tinker Task Touch Count (Experiment 1)

Source	df	MS	*F*	*p*
<u>Between</u> <u>Subjects</u>				
Subjects (Sub)	11	.0997		
<u>Within</u> <u>Subjects</u>				
Device (Dev)	5	.4460	8.90	.0001
Dev x Sub	55	.0501		
Lighting (Lt)	2	.0423	1.59	.2273
Lt x Sub	22	.0267		
Angle (An)	1	.0394	4.04	.0697
An x Sub	11	.0098		
Dev x Lt	10	.0247	1.61	.1116
Dev x Lt x Sub	110	.0058		
Dev x An	5	.0313	1.58	.1821
Dev x An x Sub	55	.0199		
An x Lt	2	.0184	1.02	.3785
An x Lt x Sub	22	.0181		
Dev x An x Lt	10	.0075	.44	.9222
Dev x An x Lt x Sub	110	.0169		
Total	431			

TABLE 9. Newman-Keuls Comparisons for the Tinker Task Touch Count Measure: Comparing Device Means

Device	Mean	Comparisons				
Sierracin	1.2498	A				
Elographics	1.3766		B			
EMS	1.4538		B	C		
ISI	1.5524			C		
Carroll	1.6722				D	
Microtouch	1.9622					E

Alphanumeric Search Experiment

Hypothesis testing on the data from the Alphanumeric search task proceded in a way similar to that of the reading speed experiment. Analyses of variance for each dependent measure were followed by simple effects and Newman-Keuls tests of significant overall effects. The criterion level for rejection of the null hypothesis was again held at $p < .05$ for all F tests and comparisons. The same three dependent measures were analyzed for this task as for the Tinker task.

Dependent variable search time. Search time data for the search task were collected in the same fashion and represent a measure similar to reading time in the Tinker task experiment. An analysis of variance of these data revealed significant main effects of viewing angle ($F_{1,11}$ = 12.06, $p = .005$), target size ($F_{2,22} = 4.77$, $p = .019$) and device type ($F_{5,55} = 2.93$, $p = .020$). An ANOVA summary table for search time is provided in Table 10.

The mean for the 90-deg condition in the main effect of viewing angle is 2.04 while the 50-deg mean is 2.31. Thus, performance is reliably better at a 90-deg or straight on viewing angle than at the 50-deg off-angle condition.

TABLE 10. ANOVA Summary Table for the Alphanumeric Search Task Search Time (Experiment 2)

Source	df	MS (s/100)	*F*	*p*
<u>Between Subjects</u>				
Subjects (Sub)	11	428.86.50		
<u>Within Subjects</u>				
Device (Dev)	5	12298.58	2.93	.0203
Dev x Sub	55	4192.33		
Target Size (TT)	2	7581.18	4.77	.0190
TT x Sub	22	1587.78		
Angle (An)	1	85798.43	12.06	.0052
An x Sub	11	7113.30		
Dev x TT	10	1266.06	.51	.8771
Dev x TT x Sub	110	2462.79		
Dev x An	5	1829.65	.59	.7113
Dev x An x Sub	55	3127.14		
An x TT	2	2641.76	1.67	.2107
An x TT x Sub	22	1579.22		
Dev x An x TT	10	2578.54	.98	.4648
Dev x An x TT x Sub	110	2630.35		
Total	431			

Newman-Keuls comparisons were conducted for both device type and target size means (described in Tables 11 and 12). No significant pairwise differences were found for device type at an experiment-wise alpha criterion level of $p<.05$. Under the main effect of target size, the average search time for the smallest target was significantly shorter than for the largest target. This finding was unexpected because changing touch target size had no effect on the displayed targets. Therefore, it is probable that this significant effect is an artifact of the random generation of search patterns for each search task session. The random patterns of letters generated for the trials under the small target size might have been less difficult to find the target letter in on average than those generated for large target trials.

Dependent variable entry time. As can be seen in the ANOVA summary table for entry time (Table 13), this analysis led to findings of significant two-way interactions of device type by target size ($F_{10,110} = 3.89$, $p = .0002$) and target size by angle ($F_{2,22} = 6.31$, $p = .0068$). All three main effects were also significant.

Subsequent one-way simple effect F tests were conducted for the effects of target size for each device. A complementary set of one-way tests was also completed for the effects of device at each level of target size.

TABLE 11. Newman-Keuls Comparisons for the Search Task
Search Time Measure: Comparing Device Means

Device	Mean (s)	Comparisons
EMS	2.0708	A
Carroll	2.0744	A
Elographics	2.0924	A
ISI	2.1453	A
Microtouch	2.3215	A
Sierracin	2.3628	A

TABLE 12. Newman-Keuls Comparisons for the Search Task Search Time Measure: Comparing Target Size Means

Target Size	Means (s)	Comparisons	
Small	2.1021	A	
Medium	2.1848	A	B
Large	2.2467		B

TABLE 13. ANOVA Summary Table for the Alphanumeric Search Task Entry Time Measure (Experiment 2)

Source	df	MS (s/100)	F	p
Between Subjects				
Subjects (Sub)	11	16,651.2		
Within Subjects				
Device (Dev)	5	44,089.28	6.74	.0001
Dev x Sub	55	6,536.8		
Viewing Angle (An)	1	16,338.9	12.09	.0052
An x Sub	11	1.351.2		
Target Size (TT)	2	492,323.6	59.41	.0001
TT x Sub	22	8,286.5		
Dev x An	5	2,083.3	.72	.6083
Dev x An x Sub	55	2,877.2		
Dev x TT	10	11,924.8	3.89	.0002
Dev x TT x Sub	110	3,069.1		
TT x An	2	8,231.1	6.31	.0068
TT x An x Sub	22	1,303.6		
Dev x TT x An	10	1,723.74	1.01	.4364
Dev x TT x An x Sub	110	1,699.9		
TOTAL	431			

Results of these tests are included in Tables 14 and 15 as well as in Figure 9.

For each device, significant one-way effects of target size were present. Further investigation through Newman-Keuls comparisons for each device revealed that for the Sierracin touch panel 30 x 30 and 40 x 40 targets were not significantly different nor were 20 x 20 pixel and 30 x 30 pixel targets. However, the 40 x 40 pixel condition was significantly better in terms of entry time performance than was the 20 x 20 pixel size.

All five of the remaining devices showed a pattern of results from the Newman-Keuls tests that was different from the Sierracin device. For these five devices, the 40 x 40 and 30 x 30 pixel square target sizes were not significantly different than each other while the 20 x 20 pixel was significantly worse than either of the larger sizes. These findings are summarized in Table 16.

One way simple effect tests for the effect of devices at each level of target size led to a different pattern of results. One-way tests were significant at 20 x 20 pixel and 30 x 30 pixel target sizes, but nonsignificant at the 40 x 40 pixel target size.

Newman-Keuls comparisons at the 20 x 20 pixel target size (Table 17) revealed that the Sierracin, Elographics, and EMS panels did not differ significantly and were

TABLE 14. Simple Effect F Tests for the Search Task, Entry Time Measure: Tests for Effects of Target Size at Levels of Device

Device	MS	*F*	*p*
Carroll	161,149.149	52.507	< .01
EMS	61,701.674	20.104	< .01
Elographics	38,726.293	12.618	< .01
Sierracin	22,387.147	7.294	< .01
Microtouch	166,750.000	54.3316	< .01
ISI	101,232.444	32.98	< .01

TABLE 15. Simple Effect F Tests for the Search Task Entry Time Measure: Test for Effect of Device at Levels of Target Size

Target Size	MS	*F*	*p*
Small	54,014.996	17.600	< .01
Medium	10,600.208	3.45	< .01
Large	3,323.722	1.08	> .25

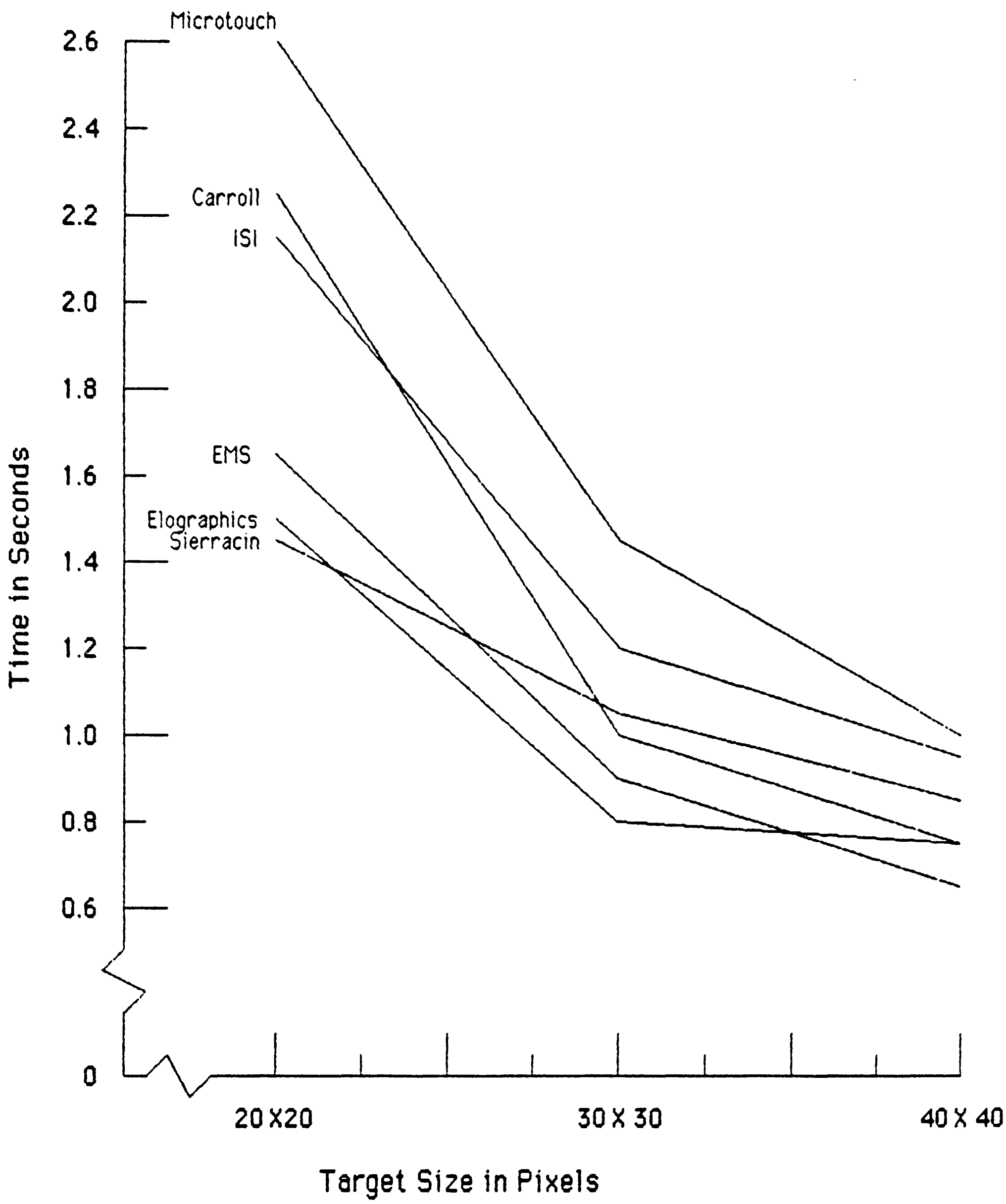

Figure 9: Search Task Entry Time Performance,

Device by Target Size Interaction.

TABLE 16. Newman-Keuls Comparisons for the Search Task, Entry Time Measure: Comparing Target Size Means at Levels of Device

Target Size	Carroll Device Mean (s)	Comparisons	
Small	2.29	A	
Medium	.99		B
Large	.77		B

Target Size	EMS Device Mean (s)	Comparisons	
Small	1.63	A	
Medium	.89		B
Large	.66		B

Target Size	Elographics Device Mean (s)	Comparisons	
Small	1.50	A	
Medium	.87		B
Large	.75		B

Target Size	Sierracin Device Mean (s)	Comparisons	
Small	1.45	A	
Medium	1.05	A	B
Large	.85		B

Table 16, continued

Target Size	Microtouch Device Mean (s)	Comparisons
Small	2.59	A
Medium	1.43	B
Large	.97	B

Target Size	ISI Device Mean (s)	Comparisons
Small	2.16	A
Medium	1.18	B
Large	.93	B

Note: Means followed by the same letter are not significantly different.

TABLE 17. Newman-Keuls Comparisons for the Search Task Entry Time Measure: Comparing Device Means at Small Target Size

Device	Mean (s)	Comparisons	
Sierracin	1.45	A	
Elographics	1.50	A	
EMS	1.63	A	
ISI	2.16		B
Carroll	2.29		B
Microtouch	2.59		B

Note: Means followed by the same letter are not significantly different.

reliably better than the remaining three devices with respect to entry time.

At the 30 x 30 pixel target size (Table 18), comparisons among means showed the Elographics device to have a significantly shorter mean entry time than the ISI and Microtouch TSDs. No other pairwise comparisons at this level of target size were significant.

Following the overall analysis of variance for entry time, the significant viewing angle by target size interaction was further described by Newman-Keuls comparisons (seen in Table 19 and Figure 10) among all possible combinations of interaction means. These tests resulted in significant differences among all pairs of means in the test except the combination of the two 30 x 30 target size means at 90-deg and 50-deg angles and the combinations of 40 x 40 pixel target with the two viewing angles. Across viewing angles, the 40 x 40 pixel target size was reliably better than the 30 x 30 pixel target which was, in turn, better than the 20 x 20 pixel size. Within each target size, the only condition under which viewing angle had a significant effect on entry time data was the 20 x 20 pixel size.

Dependent variable touch count. The analysis of variance for the mean number of touches required to complete each trial, summarized in Table 20, reflects the

TABLE 18. Newman-Keuls Comparisons for the Search Task, Entry Time Measure: Comparing Device Means at Medium Target Size

Device	Mean (s)	Comparisons	
Elographics	.87	A	
EMS	.89	A	B
Carroll	.99	A	B
Sierracin	1.05	A	B
ISI	1.18		B
Microtouch	1.43		B

Note: Means followed by the same letter are not significantly different.

TABLE 19. Newman-Keuls Comparisons for the Search Task Entry Time Measure: Comparing Target Size by Viewing Angle Interaction Means

Target Size	Viewing Angle	Mean (s)	Comparison
Large	90^o	.7997	A
Large	50^o	.8451	A
Medium	90^o	1.0580	B
Medium	50^o	1.0844	B
Small	90^o	1.7877	C
Small	50^o	2.0850	D

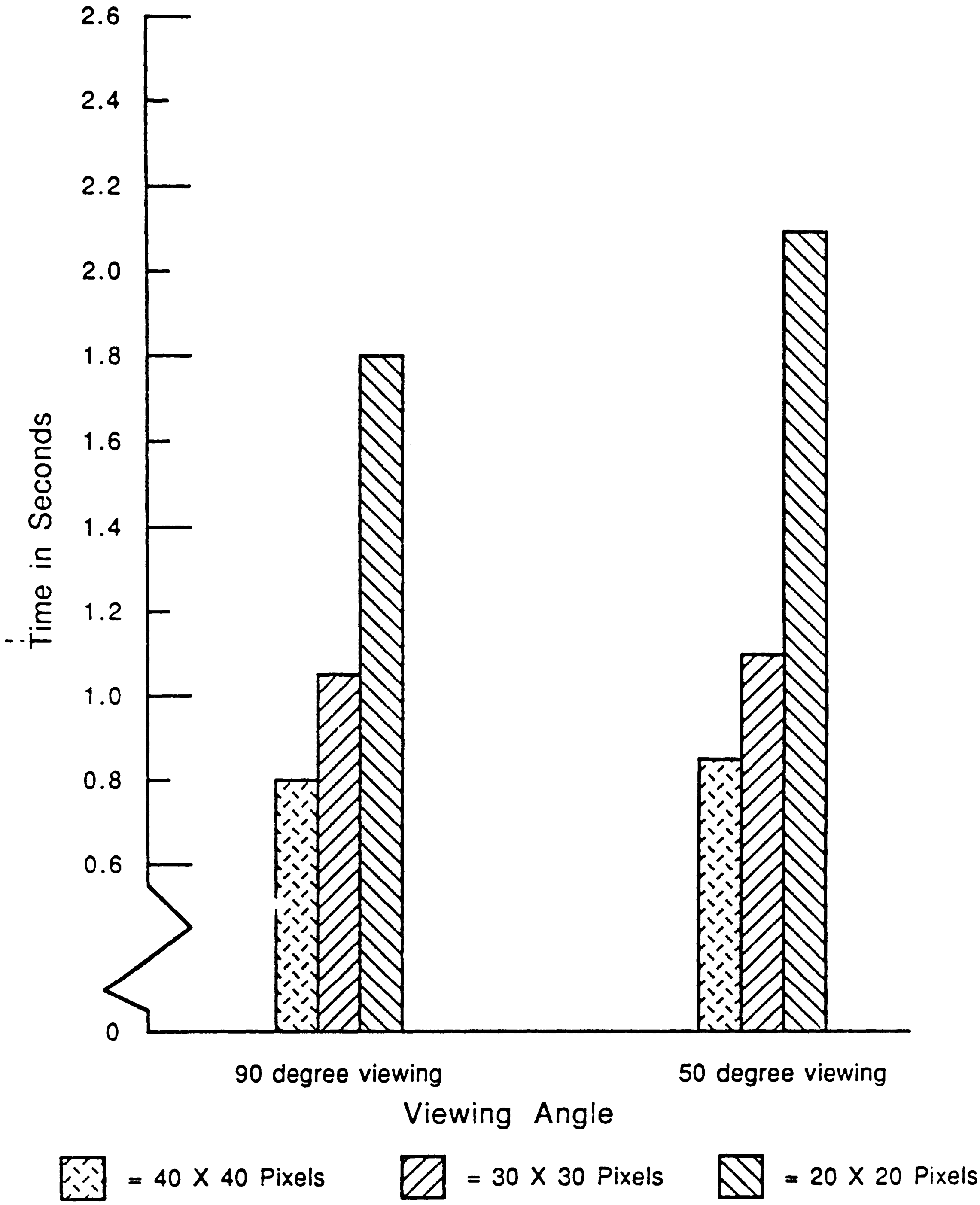

Figure 10: Search Task Entry Time Performance, Viewing Angle by Target Size Interaction.

TABLE 20. ANOVA Summary Table for the Alphanumeric Search Task Touch Count Measure (Experiment 2)

Source	df	MS	*F*	*p*
Between Subjects				
Subjects (sub)	11	1.263		
Within Subjects				
Device (Dev)	5	4.523	5.78	.0002
Dev x Sub	55	.783		
Viewing Angle (AN)	1	2.433	11.73	.0057
AN x Sub	10	.207		
Target Size (TT)	2	48.713	80.34	.0001
TT x Sub	22	.606		
Dev x AN	5	.142	.55	.7398
Dev x AN x Sub	55	.260		
Dev x TT	10	1.328	4.80	.0001
Dev x TT x Sub	110	.277		
TT x AN	2	1.631	8.50	.0018
TT x AN x Sub	22	.192		
Dev x TT x AN	10	.173	1.04	.4146
Dev x TT x AN x Sub	110	.166		
Total	431			

same pattern of significant results that was observed in the entry time analysis. Significant interactions of device by target size ($F_{10,110}$ = 4.8, p = .0001) and target size by viewing angle ($F_{2,22}$ = 8.5, p = .0018) were found. These interactions can be seen in Figures 11 and 12. In addition, all three main effects are significant.

Within the device type by target size interaction, one-way simple effect tests at each level of device, as seen in Table 21, also matched the results obtained for the entry time variable with all six tests significant. As with the entry time data, Newman-Keuls comparisons were conducted among touch count means for each device. These results, presented in Table 22, also followed the pattern of the previous measure except in the case of the Microtouch TSD. The Microtouch panel was significantly better at the 40 x 40 pixel condition than at the 30 x 30 pixel group which in turn was significantly better than the 20 x 20 pixel target size.

Looking at the device by target size interaction from the opposing viewpoint, one-way simple effects were tested for each target size. These tests too matched the findings from the entry time data (see Table 23) with significant tests at the 20 x 20 and 30 x 30, but not at the 40 x 40 pixel target sizes.

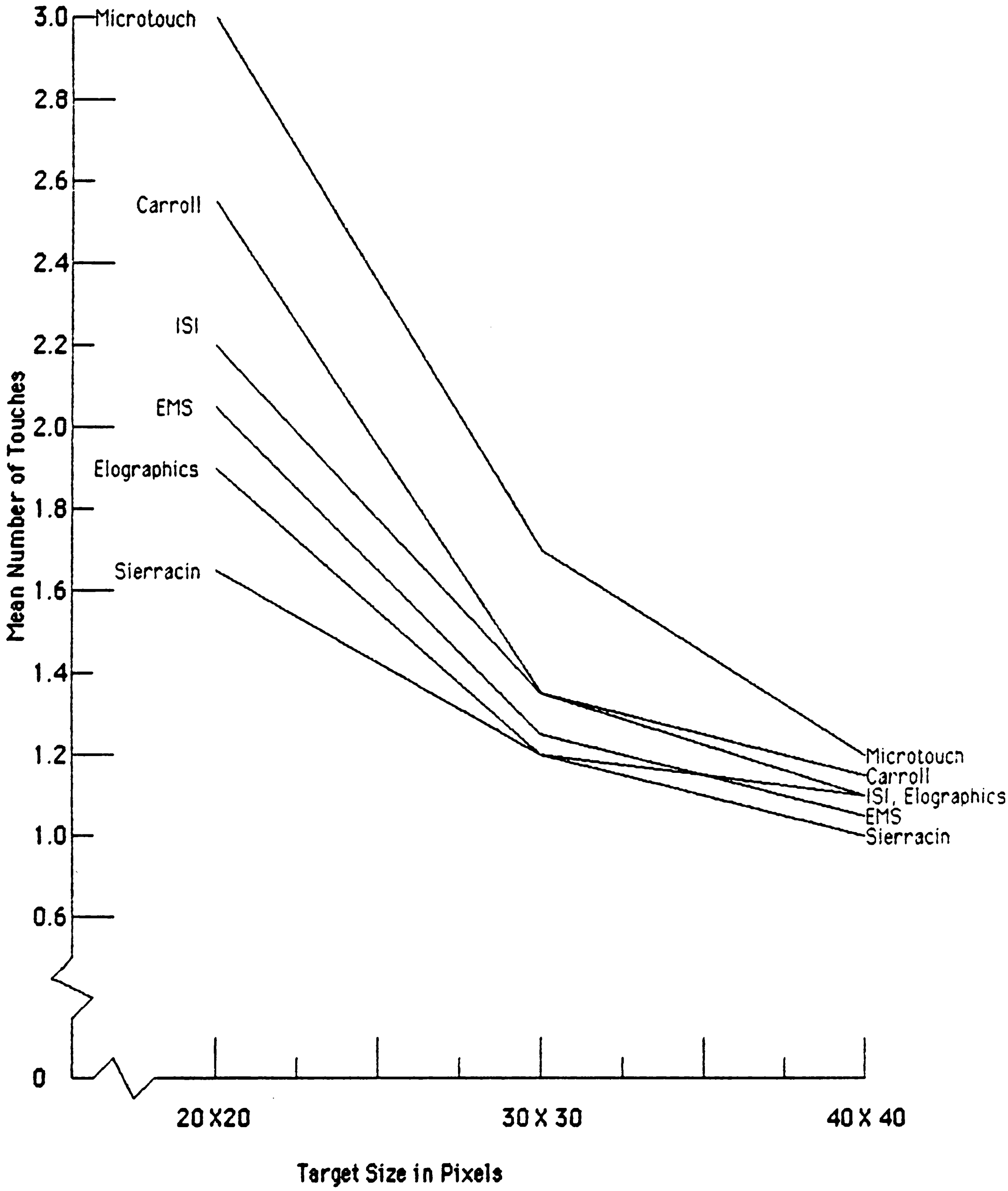

Figure 11: Search Task Touch Count Performance, Device by Target Size Interaction.

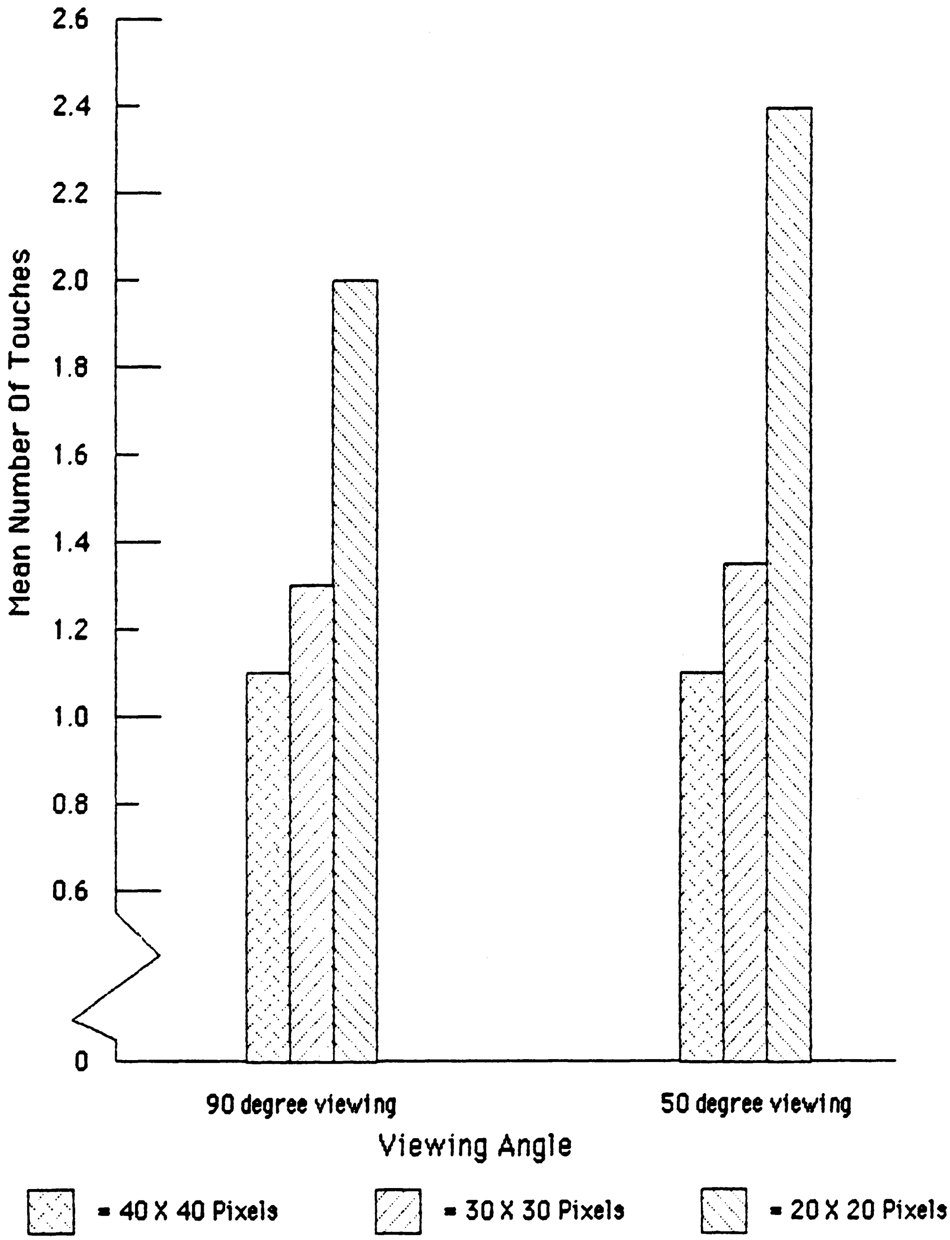

Figure 12: Search Task Touch Count Performance, Viewing Angle by Target Size Interaction.

TABLE 21. Simple Effect F Tests for the Search Task Touch Count Measure: Tests for Effects of Target Size at Levels of Device

Device	MS	*F*	*p*
Carroll	13.5092	48.84	< .01
EMS	6.5820	23.80	< .01
Elographics	4.6120	16.67	< .01
Sierracin	1.6521	5.97	< .01
Microtouch	20.2614	73.25	< .01
ISI	8.7339	31.58	< .01

TABLE 22. Newman-Keuls Comparisons for Search Task Touch Count Measure: Comparing Target Size Means for Each Device

Target Size	Carroll Device Mean (s)	Comparisons
Small	2.53	A
Medium	1.34	B
Large	1.15	B

Target Size	EMS Device Mean (s)	Comparisons
Small	2.05	A
Medium	1.27	B
Large	1.05	B

Target Size	Elographics Device Mean (s)	Comparisons
Small	1.88	A
Medium	1.19	B
Large	1.06	B

Target Size	Sierracin Device Mean (s)	Comparisons
Small	1.54	A
Medium	1.19	A B
Large	1.03	B

Table 22, continued

Target Size	Microtouch Device Mean (s)	Comparisons
Small	2.98	A
Medium	1.70	B
Large	1.20	C

Target Size	ISI Device Mean (s)	Comparisons
Small	2.23	A
Medium	1.34	B
Large	1.09	B

Note: Means followed by the same letter are not significantly different.

TABLE 23. Simple Effect F Tests for the Search Task Touch Count Measure: Tests for Effect of Device at Levels of Target Size

Target Size	MS	F	p
Small	6.1891	22.38	< .01
Medium	.8841	3.20	< .01
Large	.1052	.381	> .25

Further analysis of the two significant one-way tests revealed that the pattern of significant differences among device group means at 20 x 20 and 30 x 30 pixel targets was different from the entry data results of similar tests. The Microtouch device required significantly more touches at the smallest target size than any other device at that size (see Table 24). The Sierracin panel required significantly fewer touches than all but the Elographics TSD at the 20 x 20 pixel target condition. Carroll Touch showed the second highest average number of touches, which made it significantly worse than Sierracin, Elographics, and EMS but not reliably different from the ISI device. Touch count continued to decrease for ISI, EMS, and Elographics but these devices did not differ from each other significantly at the small target size.

The comparison of device means at the 30 x 30 pixel target size condition (Table 25) resulted in no significant differences among devices at an alpha criterion level of $p < .05$.

The final significant result from the analysis of variance of touch count data was a target size by viewing angle interaction. Newman-Keuls multiple comparisons were employed to assess the meaning of this finding. Results of these comparisons are tabulated in Table 26, which shows a significant difference between 20 x 20 pixel group means at

TABLE 24. Newman-Keuls Comparisons for the Search Task Touch Count Measure: Comparing Device Means at Small Target Size

Device	Mean (s)	Comparisons			
Sierracin	1.54	A			
Elographics	1.88	A	B		
EMS	2.05		B		
ISI	2.23		B	C	
Carroll	2.53			C	
Microtouch	2.98				D

Note: Means followed by the same letter are not significantly different.

TABLE 25. Newman-Keuls Comparisons for the Search Task Touch Count Measure: Comparing Device Means at Medium Target Size

Device	Mean (s)	Comparison
Sierracin	1.186	A
Elographics	1.188	A
EMS	1.27	A
Carroll	1.337	A
ISI	1.338	A
Microtouch	1.70	A

Note: Means followed by the same letter are not significantly different.

TABLE 26. Newman-Keuls Comparisons for the Search Task Touch Count Measure: Comparing Viewing Angle by Target Size Interaction Means

Target Size	Viewing Angle	Mean	Comparisons			
40 x 40	90^{o}	1.09	A			
40 x 40	50^{o}	1.10	A			
30 x 30	90^{o}	1.32		B		
30 x 30	50^{o}	1.36		B		
20 x 20	90^{o}	2.00			C	
20 x 20	50^{o}	2.40				D

Note: Means followed by the same letter are not significantly different.

90-deg versus 50-deg viewing angles. No similar effect of viewing angle is observed for conditions with 30 x 30 or 40 x 40 pixel target sizes. These data agree with the results of similar tests conducted on entry time data.

HARDWARE PARAMETERS RESULTS

Overview

Several measurements were made of characteristics of each touch device as reported in the method section. Many of these measurements are system based, meaning that the reported findings are not independent of the computer and display system used as the test bed in this evaluation. However, since the same computer-display system was used with each TSD, the relative differences among TSDs on these measures are independent of the system. System dependent measures include parallax, all image quality, RMS, and Wiener spectrum. System independent measures include RTF, resolution measures, and touch force.

The results of hardware parameter measurements are too numerous to itemize in text; therefore, they are presented in tables. Tables 27 and 28 contain the Ambient MTF and all other image quality signal metrics respectively. Table 29 has RTF while Table 30 includes Wiener spectrum and RMS noise measures. Signal/noise ratios are provided in Table 31. Touch characteristics make up Tables 32 and 33. Table 32 contains parallax measures, and Table 33 resolution and touch force.

TABLE 27. Ambient MTF Measures

Device	Light Level	Viewing Angle	MTF
Carroll	Nightime	90°	2.1731
Carroll	Nightime	50°	2.5230
Carroll	Office	90°	.9713
Carroll	Office	50°	1.3623
Carroll	Daytime	90°	.1094
Carroll	Daytime	50°	.1358
EMS	Nightime	90°	2.1731
EMS	Nightime	50°	2.523
EMS	Office	90°	.9713
EMS	Office	50°	1.3623
EMS	Daytime	90°	.1094
EMS	Daytime	50°	.1358
Elographics	Nightime	90°	1.3191
Elographics	Nightime	50°	1.1682
Elographics	Office	90°	.6205
Elographics	Office	50°	.4915
Elographics	Daytime	90°	.0602
Elographics	Daytime	50°	.0431

Table 27, continued

Device	Light Level	Viewing Angle	MTF
Sierracin	Nightime	90°	1.2806
Sierracin	Nightime	50°	1.0330
Sierracin	Office	90°	.8199
Sierracin	Office	50°	.5323
Sierracin	Daytime	90°	.0802
Sierracin	Daytime	50°	.0395
Microtouch	Nightime	90°	.9656
Microtouch	Nightime	50°	.6595
Microtouch	Office	90°	.3571
Microtouch	Office	50°	.2083
Microtouch	Daytime	90°	.0321
Microtouch	Daytime	50°	.0189
ISI	Nightime	90°	1.2210
ISI	Nightime	50°	.8747
ISI	Office	90°	.5038
ISI	Office	50°	.3319
ISI	Daytime	90°	.0458
ISI	Daytime	50°	.0279

TABLE 28. Image Quality Signal Strength Measures

Object	MTF	MTFA	EP	SSF
Carroll				
Horizontal line	2.0499	13.549	1.2115	12.4042
Vertical line	2.1044	14.305	1.2733	12.2692
Mean	2.0722	13.989	1.2424	12.3367
EMS				
Horizontal line	2.0499	13.549	1.2115	12.4042
Vertical line	2.1044	14.305	1.2733	12.2692
Mean	2.0772	13.989	1.2424	12.3367
Elographics				
Horizontal line	1.3638	9.896	.7559	4.1717
Vertical line	1.4051	10.251	.7815	4.2931
Mean	1.3845	9.9610	.7687	4.2324
Sierracin				
Horizontal line	1.7600	11.392	.9166	11.5567
Vertical line	1.7007	11.365	.8780	10.0387
Mean	1.7304	11.378	.8973	10.7977
Microtouch				
Horizontal line	1.0488	8.053	.6398	1.3563
Vertical line	1.0304	7.917	.6233	1.2936
Mean	1.0414	7.985	.6315	1.3250
ISI				
Horizontal line	1.6945	8.886	.7094	1.9972
Vertical line	1.2099	9.208	.7315	1.9375
Mean	1.4522	9.047	.7222	1.9674

TABLE 29. Reflectance Transfer Function Measure

Device	Angle of Incidence	RTF
Carroll	20°	.0857
Carroll	40°	.0150
EMS	20°	.0857
EMS	40°	.0150
Elographics	20°	.0148
Elographics	40°	.0645
Sierracin	20°	.0278
Sierracin	40°	.0147
Microtouch	20°	.0013
Microtouch	40°	.0006
ISI	20°	.0023
ISI	40°	.0009

TABLE 30. Image Quality Noise Measures

Device	RMS Luminance (cd/m^2)	WS (Luminance Power/Cycles/mm)
Carroll	8.1564	6.574
EMS	8.1564	6.574
Elographics	14.1666	22.390
Sierracin	27.2830	14.280
Microtouch	7.8840	7.685
ISI	10.3813	11.130

TABLE 31. Signal/Noise Measures

Device	Viewing Angle	MTF/RTF	MTFA/RTF	EP/RTF	SSF/RTF
Carroll	90^{o}	24.18	163.24	14.50	143.95
Carroll	50^{o}	138.15	932.64	82.83	822.45
EMS	90^{o}	24.18	163.24	14.50	143.95
EMS	50^{o}	138.15	932.64	82.83	822.45
Elographics	90^{o}	93.55	637.04	51.94	285.97
Elographics	50^{o}	21.47	154.43	11.92	65.62
Sierracin	90^{o}	62.25	400.64	32.28	388.42
Sierracin	50^{o}	117.71	757.67	61.04	734.56
Microtouch	90^{o}	801.08	6142.15	485.77	1,019.23
Microtouch	50^{o}	1,735.67	13308.00	1,052.50	2,208.33
ISI	90^{o}	631.39	3933.35	314.00	855.39
ISI	50^{o}	1,613.56	10051.88	802.44	2,186.00

Device	MTF/RMS	MTFA/RMS	EP/RMS	SSF/RMS
Carroll	0.2541	1.7150	0.1523	1.5125
EMS	0.2541	1.7150	0.1523	1.5125
Elographics	0.0977	0.7030	0.0543	0.2988
Sierracin	0.0634	0.1470	0.0329	0.3958
Microtouch	0.1321	1.0130	0.0801	0.1681
ISI	0.1399	0.8710	0.0696	0.1895

Table 31, continued.

Device	MTF/WS	MTFA/WS	EP/WS	SSF/WS
Carroll	0.3152	2.1280	0.1890	1.8770
EMS	0.3152	2.1280	0.1890	1.8770
Elographics	0.0618	0.4450	0.0343	0.1890
Sierracin	0.1212	0.7790	0.0628	0.7561
Microtouch	0.1355	1.0390	0.0822	0.1724
ISI	0.1305	0.8130	0.0649	0.1768

TABLE 32. Touch Characteristics: Parallax Measures

Measurement Condition	Parallax A (mm)	Parallax B (mm)
Elographics		
90° Viewing Angle	2.2776	8.7998
50° Viewing Angle	7.4871	11.0900
Sierracin		
90° Viewing Angle	1.3397	5.1764
50° Viewing Angle	4.4042	6.6631
ISI		
90° Viewing Angle	2.6795	10.3500
50° Viewing Angle	8.8084	13.3262
Carroll		
90° Viewing Angle	2.6795	10.3500
50° Viewing Angle	8.8084	13.3362
Microtouch		
90° Viewing Angle	2.4111	6.7292
50° Viewing Angle	7.9275	8.6620

TABLE 33. Touch Characteristics: Resolution and Touch Force Measures

Device	Limiting Resolution (pts/cm)	Effective Resolution ER1	ER2	Touch Force (N)
Carroll	3.15	3.15	3.15	0.0000
EMS	3.15	3.15	3.15	0.0000
Elographics	78.74	5.98	5.70	1.0101
Sierracin	1.08	1.08	1.08	2.7145
Microtouch	78.75	7.02	5.21	0.0226
ISI	7.88	3.78	2.53	0.0147

In addition to a tabularized presentation of results a brief description of each tabled category is included as an aid to their interpretation.

MTF, MTFA, EP, and SSF

These four measures of system image quality are grouped together in Table 28 because each describes image quality as an output property of a system independent of ambient viewing conditions. Each metric has six observations, one for each TSD-display system. Measurement of these image quality metrics is done under identical optimal conditions.

MTF, MTFA, EP, and SSF are all similar in that each is an area under a frequency domain function. For each, a larger area generally indicates a higher image quality.

Ambient MTF

Ambient MTF (Table 27) is also an area metric of image quality. Unlike the previous measures, however, it does not represent the ideal condition output of the system. Ambient MTF as the name implies takes into account the conditions or setting in which the system is operating. This metric included 36 observations, one for each unique TSD-viewing angle-light level combination under which operator performance was evaluated.

RMS, WS

RMS luminance and WS are ideal system noise metrics which do not take ambient conditions into account. Six levels of each measure were recorded and are reported in Table 30. RMS is a distance domain metric, being the mean variation in luminance output (in candelas per square meter) of a system over a given distance. Therefore, higher numbers in this measure indicate greater luminance variability or noise.

WS is a measure of the same type of luminance variability transformed into the frequency domain. WS is reported as an area under a function of luminance power across spatial frequencies. Larger area under the WS function means more total noise power in the system output.

RTF

The RTF (Table 29) is a TSD noise measure which is independent of the display system and which accounts for varying viewing angle. Twelve RTF values were recorded, with one observation for each TSD-viewing angle condition in the performance experiment. RTF is also an area under a function with higher areas representing a higher resolution of directly reflected images from the operational setting.

Sharper reflected images possess a higher potential for transmitting meaningful visual information that is not related to the information being transmitted by the display. Therefore, under some viewing conditions, an increase in RTF will produce an increase in interference or noise in system output.

Under other viewing conditions, increased RTF can indicate a decrease in visual noise interference from the surrounding environment. A high RTF indicates that directly reflected images will be sharp while relatively little scatter of light from sources at other than direct angles will occur. If the ambient conditions include high information light sources at locations that will be directly reflected into the viewer's eyes, RTF is a measure of the TSD's transmission of that potential interference. If the situation is such that all light sources are at off angles to the viewer, then RTF measures tha ability of the first surface of the TSD display to prevent transmission of these images.

LR, ER1, ER2, and Touch Force

Touch resolution and touch force (Table 33) are TSD specific measures that do not vary with different conditions. Touch resolution, however, is expected to interact with touch target size in its impact on user performance.

Limiting resolution (LR) is measured as the number of possible touch coordinates per centimeter as defined by the physical and software limitations of the TSD. ER1 and ER2 are TSD performance-based measures also reported in touch points per centimeter. For all three measures of touch resolution higher values indicate increased touch recognition resolving power.

Touch force (TF) is reported in Newtons required to activate a TSD touch recognition.

Parallax A and B

Both measures of TSD-display parallax (Table 32) vary with viewing angle. The measures are based both on distance in millimeters between a target seen on the display and its corresponding touch point. High parallax values are indicative of poor TSD-display system integration.

Signal/Noise Ratios (MTF/RTF, etc.)

Table 31 contains data for the 12 possible signal-to-noise ratios. Units of measure for each ratio are not necessarily meaningful. For example, MTF/RMS is a frequency domain metric being divided by a distance domain measure. Though not in easily interpretable units, each ratio represents a measure of system signal strength weighted by system noise.

The different ratios are not scaled and are therefore not directly comparable across measures. Within each ratio, observations with relatively high values represent better overall system information output capabilities.

PERFORMANCE MODELING RESULTS

Overview

Modeling of operator performance from hardware parameters was done with two primary goals. First, hardware variables that had an impact on performance needed to be identified and the strength of their individual relationships with user performance had to be assessed. Second, from these variables, models were to be constructed which accounted for as much as possible of the variation in operator performance across the various experimental conditions.

These goals were reached by evaluating simple correlations between dependent performance variables and hardware measures and by using multiple linear regressions. Modeling followed a progressive exploratory strategy starting with simplest regression models predicting overall performance and ending with best fitting models predicting performance under specific subsets of conditions. An overall criterion of maximizing the predicted amount of performance data variation determined the course of model selection. The primary tool in construction of models was a stepwise model selection procedure.

The analyses of potential models were broken into four steps. First, each hardware measure was evaluated as a single predictor model for both performance tasks and all dependent measures. For this and all subsequent steps, a fourth dependent variable was added to the three used in the performance evaluation. This variable was total task time, created by adding together reading/search time and entry time for each observation.

The second step was a stepwise selection to arrive at the best regression models for each task and dependent variable. These models were constructed using the full data sets from Tinker and search tasks. Results from this procedure accounted for significant but not large proportions of performance variation.

Step 3 consisted of the inclusion of some experimental condition variables in the stepwise procedure in an attempt to increase the variance accounted for by the models. This strategy was successful, indicating that a substantial proportion of performance variability was due to the manipulation of the conditions under which the TSDs were compared.

Having found a large contribution by including the experimental condition variables, Step 4 attempted to maximize prediction by hardware measures. This was accomplished by using the stepwise procedure with hardware parameters for specific subsets of data from different experimental conditions.

Upon review of the ANOVA mean squares for subjects from the performance evaluation, it was thought that some accounting for interindividual variability would add to the goodness of fit of models. Step 5 assessed this possibility by adding the class variable subjects to the stepwise procedures for the Tinker and alphanumeric search tasks model selection.

Again, adding a class variable increased the fit of models in most cases. Therefore, Step 6 was undertaken as an attempt to gain fit from models made up of hardware measures. In step 6, data from the 12 subjects in the study were averaged, reducing the number of unique observations in the Tinker and search task data sets from 432 each to 36 each. Models were then constructed with stepwise selection to fit performance of these "average operator" data.

Finally, "average operator" data sets were split into specific levels of lighting and target size and models were constructed for each of these six data subsets.

The results of these seven stages of model building are presented in tabular form. The accompanying text describes the general contents of each table and points out some general trends and counterintuitive findings in the results. No attempt is made to exhaustively describe these results in the text on the assumption that tables are less confusing and lend themselves better to interpretation.

Step 1: Single Predictor Models

Tables 34, 35, 36, and 37 contain Pearson product-moment correlations between hardware and performance measures for Tinker and alphanumeric search tasks. Tables 38 and 39 provide R^2 statistics for image quality and touch characteristic measures used as single predictor models of performance.

In general, some performance measures are more meaningful for certain hardware measures than for others. For example, reading speed in the Tinker task is expected to be related most closely with image quality and not related to touch weight or touch resolution. Also, touch characteristic measures should have a strong relationship with search task entry time and touch count and a similar but somewhat weaker tie to Tinker task entry and touch measures. These trends are borne out in the the data with a few exceptions.

Coincidentally, WS and RMS each have a strong negative correlation with touch force. Because of these high correlations, WS and RMS show unexpected significant relationships with touch count while touch force is significantly related to reading time.

TABLE 34. Pearson Product-Moment Correlations Between Image Quality Measures and Performance Measures for the Tinker Task Data

Image Quality	Reading Time	Entry Time	Number Touches	Total Time
MTF	-.1605	-.1166	.0685	-.1690
MTFA	-.1586	-.1360	.1091	-.1698
Ambient MTF	-.3662	-.0014	.1650	-.3502
EP	-.1547	-.1179	.1129	-.1636
SSF	-.1322	-.1272	.0470	-.1434
RTF	-.1022	-.0773	.1621	-.1080
RMS	.0538	-.0543	-.3096	.0442
WS	.0740	-.1370	-.3219	.0525
MTF/RTF	.1571	.1987	.0419	.1767
MTF/RMS	-.1449	-.0233	.3019	-.1417
MTF/WS	-.1491	-.0219	.2742	-.1454
MTFA/RTF	.1633	.1957	.0574	.1822
MTFA/RMS	-.1348	-.0216	.3197	-.1318
MTFA/WS	-.1423	-.0183	.2916	-.1384
EP/RTF	.1635	.1963	.0561	.1825
EP/RMS	-.1412	-.0298	.3090	-.1390
EP/WS	-.1463	-.0274	.2789	-.1425
SSF/RTF	.1680	.1981	-.0117	.1870
SSF/RMS	-.1589	-.0986	.2092	-.1651
SSF/WS	-.1547	-.0880	.1884	-.1596

TABLE 35. Pearson Correlation Coefficients Between Touch Characteristic Measures and Performance Measures for Tinker Task Data

Touch Characteristics	Reading Time	Entry Time	Number Touches	Total Time
Touch Force	.1078	-.1361	-.1978	-.03910
LR	.1210	-.0083	-.0115	.11425
ER1	-.0117	.1320	.1851	.09540
ER2	.0589	-.0432	.0911	.05050
Parallax A	.1196	.0143	-.0115	.11623
Parallax B	.0004	-.0867	-.0804	.01110

TABLE 36. Pearson Product-Moment Correlations Between Image Quality and Performance Measures for the Alphanumeric Search Task Data

Image Quality Metric	Search Time	Entry Time	Number Touches	Total Time
MTF	-.0918	-.1508	-.1204	-.1741
MTFA	-.0992	-.1477	-.0923	-.1763
AmbMTF	-.0988	-.1358	-.0760	-.1668
EP	-.0822	-.1363	-.0885	-.1569
SSF	-.0453	-.1577	-.1242	-.1507
RTF	-.0157	-.0575	-.0176	-.0544
RMS	.1092	-.1273	-.2022	-.0314
WS	.0039	-.1411	-.1627	-.1074
MTF/RTF	.1180	.2465	.2048	.2648
MTFA/RTF	.1232	.2527	.2185	.2728
EP/RTF	.1240	.2525	.2177	.2732
SSF/RTF	.1462	.2286	.1707	.2683
MTF/RMS	-.1263	.0086	.0861	-.0715
MTFA/RMS	-.1203	-.0211	.1072	-.0668
EP/RMS	-.1178	.0066	.0902	-.0678
SSF/RMS	-.1106	-.0876	-.0136	-.1362
MTF/WS	-.0892	-.0133	.0484	.0656
MTFA/WS	-.0853	-.0015	-.0668	-.0540
EP/WS	-.0854	-.0140	-.0530	-.0638
SSF/WS	-.0858	-.0910	-.0304	-.1239

TABLE 37. Pearson Product-Moment Correlations Between Touch Characteristics and Performance Measures for the Alphanumeric Search Task Data.

Image Quality Metric	Search Time	Entry Time	Number Touches	Total Time
Touch Force	.1077	-.1360	-.1977	-.0387
ER1	-.0118	.1319	.1852	.0954
ER2	-.0469	.0574	.1364	.0141
LR	.0282	.0728	.1140	.0742
Parallax A	.1575	.1327	.1568	.2007
Parallax B	.0111	.0989	.1044	.0762

TABLE 38. R^2 For Image Quality Measure Single Predictor Models

Measure	Tinker Reading Time	Tinker Enter Time	Tinker Number Touches	Tinker Total Time	Search Search Time	Search Enter Time	Search Number Touches	Search Total Time
MTF	.0257*	.0136*	.0047	.0285*	.0084	.0227*	.0145*	.0303*
MTFA	.0252*	.0184*	.012*	.0288*	.0098	.0218*	.0085	.0309*
EP	.0239*	.0139*	.0127*	.0268*	.0068	.0186*	.0078	.0246*
SSF	.0175*	.0162*	.0022	.0206*	.0021	.0249*	.0154*	.0227*
AmbMTF	.1341*	.0000	.0272*	.1227*	.0098*	.0185*	.0058	.0278*
RMS	.0029	.0029	.0959*	.0020	.0119*	.0162*	.0409*	.0010
RTF	.0105*	.006	.0263*	.0117*	.0002	.00033	.0003	.0003
WS	.0740*	.0188*	.1036*	.0028	.0001	.0199*	.0265*	.0115*
MTF/RTF	.0247*	.0395*	.0018	.0312*	.0139	.0608*	.0419*	.0701*
MTF/RMS	.0210*	.0005	.0911*	.0201*	.016*	.0001	.0074	.0051
MTF/WS	.1034*	.0005	.0222*	.0211	.0080	.0002	.0023	.0043
MTFA/RTF	.0265*	.0382*	.0032	.0332*	.0152*	.0638*	.0477*	.0744*
MTFA/RMS	.0182*	.0004	.1022*	.0174*	.0151*	.0004	.0114*	.0033
MTFA/WS	.0202*	.0003	.0851*	.0191*	.0072	.00001	.0044	.0029
EP/RTF	.0267*	.0385*	.0032	.0333*	.0154*	.0638*	.0474*	.0747*
EP/RMS	.0199*	.0009	.0955*	.0193*	.0139*	.0001	.0081	.0046
EP/WS	.0211*	.0007	.0777*	.0203*	.0073	.0002	.0028	.0041
SSF/RMS	.0252	.0097*	.0438*	.0272*	.0121*	.0077	.0002	.0186*
SSF/RTF	.0282*	.0393*	.0001	.0350*	.0214*	.0522*	.0291*	.0720*
SSF/WS	.0239*	.0077	.0355*	.0255*	.0074	.0083	.0009	.0153*

* indicates significance at $p < .05$ level

TABLE 39. R^2 For Touch Characteristic Measures Single Predictor Models

	Tinker				Search			
Measure	Reading Time	Enter Time	Number Touches	Total Time	Search Time	Enter Time	Number Touches	Total Time
TF	.0032	.0055	.0915*	.002	.0116*	.0185*	.0391*	.0015
PXA	.0143*	.0002	.0001	.0157*	.0248*	.0176*	.0246*	.0403*
PXB	.0000	.0075	.0065	.0001	.0000	.0098*	.0109*	.0058
LR	.0146*	.0001	.0001	.0131*	.0008	.0053	.0130*	.0055
ER1	.008	.0016	.0118*	.0082	.0001	.0174*	.0343*	.0091*
ER2	.0035	.0019	.0083	.0026	.0022	.0033	.0186*	.0002

* indicates significance at $p \leq .05$ level

A trend in the data that fits the intuitive models upon which the hardware metrics were based is the existence of generally negative correlations between image quality metrics and time based performance measures. Also intuitive are the positive correlations found between noise and time measures. The notable (but totally logical) exception to this trend is consistently negative correlations between RTF and reading, entry, and total times.

It was expected that signal-to-noise ratios would be negatively correlated with reading/search and total times. This is the case with all ratios containing WS or RMS; however, ratios containing RTF in the denominator show the opposite relationship, again for a logical reason to be discussed later.

Overall, the best single predictor model for predicting reading and total task times is ambient MTF. This was to be expected because the ambient MTF takes viewing conditions into account while the other image quality metrics measure hardware parameters under ideal conditions only. Signal-to-noise ratios with RTF in the denominator account for the highest average percentages of variance across all dependent measures.

Step 2: Stepwise Model Selection from Hardware Measures

The stepwise procedure used here produces best regression models that contain only those variables which produce a significant increment in the variance accounted for in the presence of all other variables in the model. The criterion for significance was set at the 0.15 probability that the increment would be exceeded by chance. This level was selected as the inclusion criterion so that all predictors with a reasonable chance to significantly improve model fit would be tested while still preventing overly large models which include many minimal contributors.

Tables 40 and 41 contain best models for the Tinker and alphanumeric search tasks, respectively. Four models are included for each task, with a separate model for each dependent variable. In viewing the tables, special attention should be given to the total time dependent variable which reflects both visual and visual/motor portions of the tasks.

Each table contains standardized (beta) and nonstandardized parameter estimators. Standardized estimators reflect the relative importance of each variable to the model. Nonstandardized estimators are useful for predicting performance from measured values of parameters. R^2 statistics are also included for each model.

TABLE 40. Best Regression Models for the Tinker Test Data Using Hardware Measures as Predictors

Dependent Variable Model (Nonstandardized Weights)	R^2
Read Time = 688.6 - 108.3 Amb MTF + 8.98 PXA	.148
Entry Time = 133.4 - 2.08 WS - 160.6 EP/WS	.063
Touch Count = 1.26 + 0.029 Amb MTF - 0.0035 RMS - 0.009 WS - 0.005 PXA + 0.0006 LR	.157
Total Time = 780.68 - 108.4 Amb MTF + 9.13 PXA	.136

Dependent Variable Model (Standardized or Beta Weights)
Read Time = - 0.366 Amb MTF + 0.119 PXA
Entry Time = - 0.392 WS - 0.330 EP/WS
Touch Count = 0.124 Amb MTF - 0.144 RMS -0.288 WS -0.087 PXA + 0.129 LR
Total Time = - 0.350 Amb MTF + 0.115 PXA

Note: Amb MTF = Ambient MTF, PXA = Parallax A

TABLE 41. Best Regression Models for the Alphanumeric Search Data Using Hardware Parameters as Predictors

Dependent Variable	Model (Nonstandardized Weights)	R^2
Search Time	= 246.79 - 24.47 MTF - 2.04 WS + 0.171 TF + 4.87 PXA	0.075
Entry Time	= 129.172 - 1.266 WS + 0.004 MTFA/RTF	0.071
Touch Count	= 1.358 + 0.112 MTFA/WS + 0.0001 MTFA/RTF - 0.0005 SSF/RTF	0.079
Total Time	= 473.521 - 72.340 MTF - 5.233 WS - 0.186 TF +7.875 PXA	0.111

Dependent Variable	Model (Standardized or Beta Weights)
Search Time	= - 0.143 MTF - 0.178 WS + 0.271 TF + 0.213 PXA
Entry Time	= - 0.233 MTFA/RTF - 0.088 WS
Touch Count	= 0.092 MTFA/WS + 0.694 MTFA/RTF - 0.464 SSF/RTF
Total Time	= - 0.261 MTF - 0.282 WS + 0.182 TF + 0.213 PXA

Note: PXA = Parallax A, TF = Touch Force

Comparing models, it can be seen that the largest R^2 for any Tinker task model is 0.16. This means that at best, 84% of all variation in performance is due to variables other than the one measured. In the search task data, only 10% of the variance can be accounted for with a model made up of hardware parameter predictors.

Step 3: Adding Lighting and Target Size to the Stepwise Procedure

Two experimental condition variables were added to the stepwise procedure to determine if a larger percentage of performance variability could be accounted for. Light level was included as a categorical variable with values 1, 2, and 3 corresponding to nightime, office, and daytime conditions in the Tinker task experiment. Target area was added to the search task predictors as a continuous variable in units of mm^2.

Table 42 provides the best models for the Tinker task data with light levels as an added predictor. Reading and total time dependent measures both show large gains in R^2 over the best models without light level. From the standardized estimator models for these variables it can be seen that light level plays an important role in the performance prediction.

TABLE 42. Best Regression Models for the Tinker Test Data Using Hardware Parameters and Lighting as Predictors

Dependent Variables	Model (Nonstandardized Weights)	R^2
Read Time =	119.74 + 175.36 Amb MTF	
	- 3795.56 MTF/WS + 5183.40 EP/WS	
	+ 0.08 SSF/RTF + 240.46 Glare	0.2860
Entry Time =	133.36 - 2.08 WS - 160.63 EP/WS	0.0630
Touch Count =	1.166 +0.059 Amb MTF - 0.003 RMS	
	- 0.009 WS - 0.005 PXA + 0.0008 LR	
	+ 0.029 Glare	0.1627
Total Time =	186.25 + 184.77 Amb MTF	
	- 3740.68 MTF/WS + 5090.43 EP.WS	
	+ 0.09 SSF/RTF + 246.29 Glare	0.2710

Dependent Variable	Model (Standardized or Beta Weights)
Read Time =	0.919 Glare + 0.295 SSF/RTF + 0.592 Amb MTF
	- 1.760 MTF/WS + 1.486 EP/WS
Entry Time =	- 0.392 WS - 0.330 EP/WS
Touch Count =	0.249 Amb MTF - 0.112 RMS - 0.299 WS
	- 0.082 PXA + 0.173 LR + 0.139 Glare
Total Time =	0.900 Glare + 0.318 SSF/RTF + 0.597 Amb MTF
	- 1.658 MTF/WS + 1.395EP/WS

Note: Amb MTF = Ambient MTF, PXA = Parallax A

Table 43 includes the best models for search task performance with target size included in the predictor pool. Target size adds significantly to R^2 for all the best models in this task. Large gains are evident in entry time, touch count, and total time prediction models.

Step 4: Best Regression Models for Differing Conditions of Lighting and Target Size

Gains in variance accounted for by models which included experimental condition variables led to the question of whether or not hardware measures would predict performance better under more specific sets of conditions. Therefore, the stepwise selection procedure was repeated again with all hardware measures as variables and not including the experimental condition variables. The performance data upon which the procedure was carried out were split so that each subset of Tinker task data contained observations from only one light level and each subset of search task data contained one level of target size. This procedure resulted in 12 possible best models for each Task. Tables 44, 45, and 46 are the best models for nighttime, office, and daytime lighting conditions. Small, medium, and large touch target models are presented in Tables 47, 48, and 49, respectively.

TABLE 43. Best Regression Models for the Alphanumeric Search Data Using Hardware Parameters and Target Size as Predictors

Dependent Variables	Model (Nonstandardized Weights)	R^2
Search Time =	235.34 - 24.47 MTF - 2.04 WS + 0.17 TF + 4.87 PXA + 0.08 TS	0.083
Entry Time =	215.350 - 1.266 WS + 0.004 MTFA/RTF - 0.614 TS	0.364
Touch Count =	2.214 +0.112 MTFA/WS + 0.0001 MTFA/RTF - 0.0005 SSF/RTF - 0.006 TS	0.370
Total Time =	548.247 - 72.340 MTF - 5.233 WS + 0.186 TF + 7.875 PXA - 0.532 TS	0.245

Dependent Variable	Model (Standardized Beta Weights)
Search Time =	0.090 TS - 0.143 MTF + 0.123 PXA + 0.271 TF - 0.178 WS
Entry Time =	- 0.088 WS + 0.233 MTFA/RTF - 0.542 TS
Touch Count =	0.092 MTFA/WS + 0.694 MTFA/RTF - 0.464 SSF/RTF - 0.540 TS
Total Time =	- 0.261 MTF - 0.282 WS + 0.182 TF + 0.213 PXA - 0.365 TS

Note: PXA = Parallax A, TS= Target Size, TF = Touch Force

TABLE 44. Best Regression Models for Tinker Task Performance at Nighttime Lighting Level

Dependent Variable	Model (Nonstandardized Weights)	R^2
Entry Time	= 112.210 - 1.318 WS	0.057
Touch Count	= 0.988 - 0.074 Amb MTF + 0.254 MTFA/RMS	0.222

Note: For Read and Total Time dependent variable no predictor variables met 0.15 entry criterion

Dependent Variable	Model (Standarized or Beta Weights)
Entry Time	= - 0.137 WS
Touch Count	= - 0.246 Amb MTF +0.652 MTFA/RMS

Note: Amb MTF = Ambient MTF

TABLE 45. Best Regression Models for Tinker Task Performance at Office Light Level

Dependent Variable	Model (Nonstandardized Weights)	R^2
Entry Time = 160.771 - 22.353 MTFA+ 183.706 EP		0.171
Touch Count = 1.195 - 0.008 WS - 0.008 PXA		
+ 0.012 ER1		0.093
Note: For Read and Total Time dependent variable no predictor variables met 0.15 entry criterion		

Dependent Variable	Model (Standarized or Beta Weights)
Entry Time = 0.220 SSF/RTF - 0.063 PXA	
Touch Count = - 2.419 MTFA + 2.101 EP	

TABLE 46. Best Regression Models for Tinker Task Performance at Daytime Light Level

Dependent Variable Model (Nonstandardized Weights)	R^2
Read Time = 829.49 +21.070 PXA - 6138.31 MTF/RMS + 8619.78 EP/RMS + 0.094SSF/RTF	0.247
Entry Time = 90.075 + 0.12 MTF/RTF	0.049
Touch Count = 1.171 - 0.010 WS + 0.016 ER1	0.147
Total Time = 913.663 + 21.088 PXA - 5981.788 MTF/RMS + 8396.76 EP/RMS + 8396.76 EP/RMS + 0.104 SSF/RTF	0.242

Dependent Variable Model (Standardized or Beta Weights)
Read Time = 0.232 PXA - 1.760 MTF/RMS + 1.535 EP/RMS + 0.277 SSF/RTF
Entry Time = 0.221 MTF/RTF
Touch Count = - 0.357 WS + 0.207 ER1
Total Time = 0.291 SSF/RTF - 1.631 MTF/RMS + 0.221 PXA + 1.422 EP/RMS

Note: PXA = Parallax A

TABLE 47. Best Regression Models for Search Task Performance at Small Target Size

Dependent Variable	Model (Nonstandardized Weights)	R^2
Search Time	= 224.604 - 584.543 MTF/RMS + 49.011 MTFA/WS +3.668 PXA	0.084
Entry Time	= 118.822 + 44.629 MTFA/RMS + 0.009 MTFA/RTF	0.164
Touch Count	= 2.198 - 0.034 RMS - 0.002 MTF/RTF + 0.0003 MTFA/RTF + 0.063 PXA	0.213
Total Time	= 354.620 + 164.772 RTF + 0.012 MTFA/RTF	0.156

Dependent Variable	Model (Standardized or Beta Weights)
Search Time	= - 0.751 MTFR/RMS + 0.570 MTFA/WS + 0.181 PXA
Entry	= 0.215 MTFA/RMS + 0.385 MTFA/RTF
Touch Count	= - 0.224 RMS + 1.19 MTFA/RTF + 0.172 PXA - 1.001 MTF/RTF
Total Time	= 0.163 RTF + 0.441 MTFA/RTF

Note: Amb MTF = Ambient MTF, PXA = Parallax A

TABLE 48. Best Regression Models for Search Task Performance at Medium Target Size

Dependent Variable Model (Nonstandarized Weights)	R^2
Search Time = 173.955 + 4.804 PXA + 1.703 RMS	0.056
Entry Time = 95.824 + 0.014 MTFA/RTF - 0.073 MTF/RTF	0.167
Touch Count = 1.326 + 0.0001 MTFA/RTF - 0.001 SSF/RTF	0.174
Total Time = 305.077 + 0.030 SSF/RTF	0.074

Dependent Variable Model (Standarized or Beta Weights)
Search Time = 0.206 PXA + 0.169 RMS
Entry Time = 1.373 MTFA/RTF - 0.992 MTF/RTF
Touch Count = 1.216 MTFA/RTF - 0.902 SSF/RTF
Total Time = 0.268 SSF/RTF

TABLE 49. Best Regression Models for Search Task Performance at Large Target Size

Dependent Variable Model (Nonstandardized Weights)	R^2
Search Time = 274.606 - 2.870 WS + 5.756 PXA	0.084
Entry Time = 156.706 -21.847 MTFA + 172.381 EP + 1.159 PXA	0.253
Touch Count = 1.076 - 0.006 WS + 0.023 ER1	0.099
Total Time = 259.247 + 0.111 TF + 5.126 PXA + 0.024 SSF/RTF	0.109

Dependent Variables Model (Standardized or Beta Weights)
Search Time = 0.185 PXA - 0.286 MTF/RMS - 0.185 WS
Entry Time = - 2.442 MTFA - 2.035 EP + 0.158 PXA
Touch Count = 0.216 ER1 - 0.197 WS
Total Time = 0.229 SSF/RTF + 0.132 PXA + 0.040 TF

Note: PXA = Parallax A, TF = Touch Force

Splitting the search task data by target size was expected to increase R^2 for entry time and touch count performance models. Additionally, it was hoped that these increases would improve total task time prediction. Improvements in all three of these touch related performance measures were exhibited in the best models for the smallest target size. Improvements under medium and large target conditions were limited to touch count and entry time models.

Subsetting of the Tinker task data by light level was expected to have its main impact on models of reading time performance and to a lesser extent on all other performance measures because all contain visual performance components. Reading time prediction was markedly improved by the best model at daytime light level. A large improvement was also made in modeling total time by including only the daytime light data. However, no gains were made at the daytime condition in prediction of entry time or touch count.

At nightime and office light conditions, no model could be found that would account for a significant proportion of variance in either reading or total time performance measures. At the nightime light level, the model of touch count performance was substantially better than the corresponding full data set model. Under normal office lighting conditions, entry time was better predicted than across all conditions. In general, only the daytime subset of data represented an overall improvement in modeling over models built from the full Tinker task data set.

Step 5: Stepwise Model Selection With Subjects as a Regression Prediction

Both performance evaluation experiments employed within subjects designs. Therefore, each subject had his/her performance evaluated under all conditions in each experiment. This design allows subjects to be coded as a regressor in stepwise model selection. The utility in including subjects as a regressor is that it allows models to account for variability among individual operators.

Step 5 regression models were constructed by stepwise selection with hardware measures and subjects as regressors. As can be seen in Table 50 for Tinker task data and Table 51 for search task data, significant improvement was made in models of most measures of performance when the subjects variable was added to the stepwise selection pool. In Tinker task results, subjects contributed significantly to all models except touch count. For alphanumeric search performance, entry time and touch

TABLE 50. Best Regression Models for Tinker Task Data with Subjects as a Predictor

Dependent Variable Model (Nonstandardized Weights)	R^2
Read Time = 748.268 - 108.322 Amb MTF + 8.98 PXA - 9.180 Subject	0.170
Entry Time = 125.685 - 2.078 WS - 160.628 EP/WS + 1.181 Subject	0.081
Touch Count = 1.234 + 0.029 Amb MTF - 0.004 RMS - 0.009 WS - 0.005 PXA + 0.001 LR + 0.004 Subject	0.164
Total Time = 832.677 - 108.377 Amb MTF + 9.133 PXA - 7.999 Subject	0.151

Note: AMB MTF = Ambient MTF, PXA = Parallax A

TABLE 51. Best Regression Models for Search Task with Subjects as a Predictor

Dependent Variable	Model (Nonstandardized Weights)	R^2
Search Time	= 246.789 - 24.467 MTF - 2.042 WS + 0.171 TF + 4.870 PXA	0.075
Entry Time	= 108.571 - 1.267 WS + 0.0043 MTFA/RTF + 3.169 Subject	0.089
Touch Count	= 1.135 + 1.585 MTFA/WS + 0.0001 MTFA/RTF - 0.001 SSF/RTF + 0.036 Subject	0.102
Total Time	= 309.007 + 5.009 PXA - 5.691 ER1 - 0.169 MTF/RTF + 0.521 MTFA/RTF + 2.385 Subject	0.106

Note: PXA = Parallax A, TF = Touch Force

count models were significantly improved by adding the subject variable.

Step 6: Modeling Performance of the "Average Operator"

Because of the influence of inter-individual variability on the fit of hardware parameter based models, performance data were pooled across subjects. This was accomplished by computing a mean of the 12 subject's performance at each experimental condition for each experimental task. Modeling of this "average operator" performance was undertaken with stepwise selection. Models were constructed for each performance measure under each task, resulting in eight best fitting models of average performance. Table 52 contains models for the Tinker task data and Table 53 holds the search task models.

There are two statistical reasons why an increase in variance accounted for would be expected when modeling mean performance across subjects. First, differences among means are generally less variable than among individual observations because of the central limit theorem. Therefore, computing means across subjects produces better estimates of the population average performance than considering each individual's performance separately would. This reduction in unaccounted for variance should generally improve the fit of the hardware based models.

TABLE 52. Best Regression Models for Tinker Task "Average Operator" Means

Dependent Variable Model (Nonstandardized Weights)	R^2
Read Time = 731.554 - 108.418 AmbMTF	0.316
Entry Time = 133.364 - 2.078 WS - 160.638 EP/WS	0.385
Touch Count = 1.260 + 0.029 Amb MTF - 0.004 RMS	
- 0.009 WS - 0.00 PXA + 0.001 LR	0.683
Total Time = 824.362 - 108.475 Amb MTF	0.298

Dependent Variable Model (Standardized or Beta Weights)
Read Time = - 0.562 Amb MTF
Entry Time = - 0.972 WS - 0.818 EP/WS
Touch Time = 0.259 Amb MTF - 0.300 RMS - 0.600 WS
- 0.182 PXA + 0.268 LR
Total Time = - 0.546 Amb MTF

Note: Amb MTF = Ambient MTF, PXA = Parallax A

TABLE 53. Best Regression Models for Alphanumeric Search Task "Average Operator" Means

Dependent Variable Model (Nonstandardized Weights)	R^2
Search Time = 246.790 - 24.467 MTF - 2.042 WS + + 0.171 TF + 4.870 PXA	0.627
Entry Time = 113.564 + 0.005 MTFA/RTF	0.131
Touch Count = 1.423 + 0.00004 MTFA/RTF	0.095
Total Time = 325.713 + 0.083 EP/RTF	0.242

Dependent Variable Model (Standardized or Beta Weights)
Search Time = - 0.412 MTF - 0.514 WS + 0.782 TF + 0.615 PXA
Entry Time = 0.361 MTFA/RTF
Touch Count = 0.308 MTFA/RTF
Total Time = 0.492 EP/RTF

Note: TF = Touch Force

Second, as the ratio between number of regressor variables and observations being predicted approaches unity, variance accounted for also approaches one regardless of the relationship between each predictor and the dependent variable. In the extreme, a model with as many predictors as there are observations will always account for all the variation in the data.

Reduction of the data sets being modelled from 432 to 36 observations should cause an increase in R^2 for the two statistical reasons noted above. However, the increases in R^2 by a factor of two seen in many measures is much larger than could be expected for these reasons alone.

Models for reading time in the alphanumeric search task and touch count in the Tinker task each account for over 60% of the variance in performance. These R^2 values are approaching what is generally thought of as a good fit when considering human performance data. Slightly higher R^2s might be expected from a good model of average performance, however.

Step 7: Modeling Average Performance Under Specific Conditions

The logical final step to construction of best fitting models was the combination of the "average operator" approach with subsetting of data under conditions of target size and light level. Stepwise selection was carried out from all hardware measures at the three levels of target size in the alphanumeric search task. Best fitting models were similarly constructed for Tinker task data at nightime, office, and daytime lighting conditions. Results of these procedures are found in Tables 54, 55, and 56 for the search task models and 57, 58, and 59 for the Tinker task.

Clearly, the fitting of hardware measures to average operator data under specific experimental conditions was very successful. As would be expected from earlier results, average performance was modelled best under daytime conditions for Tinker task performance and at the small target size for the alphanumeric search task. The highest R^2 value however was .958 for touch count in the alphanumeric search task at the medium target size. An R^2 of .920 was achieved by a model of Tinker task average performance on the reading time variable. This result was found for the model of the daytime condition.

A majority of the 23 best-fitting models produced at this stage of analysis fell in the .7 to .9 range of R^2. Only at nightime level in the Tinker task did the hardware measures fail to account for a significant proportion of total time performance. Although single predictors

TABLE 54. Best Regression Models for the Alphanumeric Search Task "Average Operator" at Small Target Size

Dependent Variable	Model (Nonstandardized Weights)	R^2
Search Time	= 152.164 + 3.258 WS - 725.136 MTF/RMS + 93.465 MTFA/WS + 0.004 MTFA/RTF + 1.854 PXA	0.950
Entry Time	= 118.820 + 44.631 MTFA/RMS + 0.009 MTFA/RTF	0.768
Touch Count	= 2.198 - .034 RMS -0.002 MTF/RTF + 0.0003 MTFA/RTF + 0.063 PXA	0.858
Total Time	= 354.619 + 164.779 RTF + 0.012 MTFA/RTF	0.778

Dependent Variable	Model (Standardized or Beta Weights)
Search Time	= 0.989 WS - 2.872 MTF/RMS + 3.351 MTFA/WS + 0.909 MTFA/RTF + 0.282 PXA
Entry Time	= 0.834 MTFA/RTF + 0.465 MTFA/RMS
Touch Count	= -0.451 RMS - 2.011 MTF/RTF + 2.409 MTFA/RTF + 0.345 PXA
Total Time	= 0.984 MTFA/RTF + 0.365 RTF

TABLE 55. Best Regression Models for the Alphanumeric Search Task "Average Operator" at Medium Target Size

Dependent Variable Model (Nonstandardized Weights)	R^2
Search Time = 173.955 + 1.703 RMS + 4.804 PCA	0.449
Entry Time = 95.824 - 0.073 MTF/RTF+ 0.012 MTFA/RTF	0.762
Touch Count = 1.467 - 0.037 ER1 - 0.0007 MTF/RTF + 0.002 MTFA/RTF - 0.0006 SSF/RTF	0.956
Total Time = 305.078 + 0.030SSF/RTF	0.506

Dependent Variable Model (Standardized or Beta Weights)
Search Time = 0.5998 PXA + 0.5196 RMS
Entry Time = 2.943 MTFA/RTF - 2.126 MTF/RTF
Touch Count = -2.237 MTF/RTF 5.353 MTFA/RTF - 2.246 SSF/RTF - 0.381 ER1
Total Time = 0.7115 SSF/RTF

TABLE 56. Best Regression Models for the Alphanumeric Search Task "Average Operator" at Large Target Size

Dependent Variable	Model (Nonstandardized Weights)	R^2
Search Time	= 276.828 - 2.622 WS - 1153.742 MTF/RMS + 1427.319 EP/RMS + 6.171PXA	0.871
Entry Time	= 156.706 -21.847 MTFA + 172.380 EP + 1.159 PXA	0.870
Touch Count	= 1.076 - 0.006 WS + 0.023 ER1	0.684
Total Time	= 259.247 + 0.111 TF + 5.126 PXA + 0.024 SSF/RTF	0.718

Dependent Variable	(Standardized or Beta Weights)
Search Time	= -0.635 WS - 3.648 MTF/RMS + 2.803 EP/RMS +0.750 PXA
Entry Time	= - 4.531 MTFA + 3.778 EP + 0.294 PXA
Touch Count	= 0.703 ER1 - 0.560 WS
Total Time	= 0.585 SSF/RTF + 0.472 PXA - 0.371 TF

Note: PXA = Parallax A, TF = Touch Force

TABLE 57. Best Regression Models for Tinker Task "Average Operator" at Nightime Condition

Dependent Variable Model (Nonstandardized Weights)	R^2
Read Time = 482.837 + 4.650 WS + 295.419 EP/WS	0.747
Entry Time = 112.212 - 1.318 WS	0.494
Touch Count = 0.988 - 0.073 Amb MTF + 0.254 MTFA/RMS	0.813
Note: For total time, no regressor variable met 0.15 criterion for inclusion	

Dependent Variable Model (Standardized or Beta Weights)
Read Time = 1.770 WS + 1.036 SSF/RTF + 1.224 EP/WS
Entry Time = - 0.703 WS
Touch Count = - 0.467 Amb MTF + 1.246 MTFA/RMS
Note: For total time, no regressor variable met 0.15 criterion for inclusion

TABLE 58. Best Regression Models for Tinker Task "Average Operator" at Daytime Condition

Dependent Variable Model (Nonstandardized Weights)	R^2
Read Time = 829.488 - 21.070 PXA - 6138.37 MTF/RMS + 8619.897 EP/RMS + 0.094 SSF/RTF	0.923
Entry Time = 90.076 + 0.012 MTF/RTF	0.408
Touch Count = 1.230 - 0.012 WS + 0.001 LR	0.664
Total Time = 913.662 - 5981.843 MTF/RMS + 8396.869 EP/RMS + 0.104 SSF/RTF + 21.0882 PXA	0.917

Dependent Variable Model (Standardized or Beta Weights)
Read Time = 0.649 PXA - 2.548 MTF/RMS + 1.8996 EP/RMS
Entry Time = 0.639 MTF/RTF
Touch Count = - 0.908 WS + 0.485 LR
Total Time = - 3.174 MTF/RMS + 2.767 EP/RMS + 0.566 SSF/RTF + 0.430 PXA

Note: PXA = Parallax A

TABLE 59. Best Regression Models for Tinker Task "Average Operator" at Office Condition

Dependent Variable Model (Nonstandardized Weights)	R^2
Read Time = 567.113 + 2.611 PXA - 154.959MTF/RMS	0.476
Entry Time = 160.776 - 22.354 MTFA + 183.720 EP	0.729
Touch Count = 1.197 - 0.007 WS	0.383
Total Time = 673.426 - 206.444 MTF/RMS	0.486

Dependent Variable Model (Standardized or Beta Weights)
Read Time = 0.404 PXA - 0.624 MTF/RMS
Enter Time = -4.994 MTFA + 4.337 EP
Touch Count = - 0.619 WS
Total Time = 0.697 MTF/RMS

Note: PXA = Parallax A

accounted for a greater proportion of the variance in these averaged data subsets, a much higher R^2 is required to pass the p = .15 significance criterion for inclusion in a model.

Caution should be exercised in evaluating the results from the final stage of model construction. One reason for concern is that best fitting models have gone from two, three, or four regressors fitting 432 observations in step 2 to the same number of regressors predicting only 12 observations per model in step 7. For reasons cited earlier, this phenomenon above will tend to inflate R^2s for step 7 models.

Another cause for concern is that conditions under which the 23 best models in step 7 apply are quite narrowly defined. Therefore, it could not be expected that these models would necessarily generalize well to other data and thus they may not have good performance prediction capabilities.

In spite of these caveats, the fit of step 7 models to average performance data is very good. This at the least indicates that the hardware measures and the experimental manipulations are sufficient to account for most of the variability across two task types and a variety of performance measures. Table 60 provides a summary of R2 values for regression models from Steps 2 through 7. This

TABLE 60. Best Regression Models R^2 Summary

	Tinker Task			
Step	Reading Time	Entry Time	Number Touches	Total Time
2. Hardware Measures as Regressors	0.148	0.063	0.160	0.136
3. Add Lighting as Regressor	0.286	0.063	0.163	0.271
4. Split Data by Lighting:				
Nightime	*	0.057	0.223	*
Office	*	0.145	0.093	*
Daylight	0.247	0.049	0.147	0.242
5. Add Subjects as Regressor	0.170	0.081	0.164	0.151
6. "Average Operator"	0.316	0.385	0.683	0.298
7. "Average Operator" Split by Lighting:				
Nightime	0.759	0.494	0.815	*
Office	0.923	0.408	0.664	0.917
Daytime	0.476	0.617	0.383	0.486

* No significant model

Table 60, continued.

	Search Task			
Step	Reading Time	Entry Time	Number Touches	Total Time
2. Hardware Measures as Regressors	0.075	0.071	0.079	0.100
3. Add Lighting as Regressor	0.083	0.364	0.370	0.233
4. Split Data by Target Size				
Small	0.093	0.119	0.214	0.156
Medium	0.056	0.167	0.118	0.074
Large	0.084	0.259	0.099	0.109
5. Add Subjects as Regressor	0.075	0.089	0.102	0.106
6. "Average Operator"	0.627	0.131	0.094	0.242
7. "Average Operator" Split by Target Size:				
Small	0.951	0.768	0.861	0.779
Medium	0.449	0.766	0.958	0.506
Large	0.874	0.892	0.684	0.718

* No significant model

summary provides an overall view of the impact of different steps on performance variance accounted for.

PERFORMANCE EVALUATION DISCUSSION

Objectives

As stated in the Method section of this dissertation, the goal of the evaluation was to discriminate among TSD design alternatives. Therefore, the results that are most important here are those that contribute to discrimination among devices. Consistency of differences among devices across Tinker and search tasks is also important for generalizability of findings.

Different dependent variables within each task address different aspects of the human-TSD interface. Reading/search time, for example, is affected primarily by image quality and noise while entry time and touch count have both visual and touch related components. Because of these differences, performance measures must be addressed individually. The weight given to each measure in drawing final conclusions about devices depends upon the anticipated use of the TSD-display system.

Discussion of the performance evaluation findings will center on assessing the consistency of differences among devices across the two tasks and within tasks across different conditions. These assessments will initially be made within each performance measure. Finally, an overall assessment of device performance will be made, taking into account all measures.

Reading/Search Time

In Tinker task results, the main finding was a three-way device by viewing angle by light level interaction. Further analyses showed that at low and room lit conditions there were no significant differences among the devices in terms of reading time. At the high glare level, the slight differences in performance errors across the TSDs were enlarged and some differences among devices became significant. These differences were further exacerbated at high glare off-angle viewing conditions with two devices more affected by a change in viewing angle than the others.

The Carroll and EMS devices were best for reading across both viewing angles. The two capacitive devices, ISI and Microtouch were affected more by off-angle high glare than were the other devices. This difference is probably due to the fact that both have an etched first surface which becomes more opaque as viewing angle is decreased. The Sierracin device was least affected by changes in viewing angle with only a relatively small increase in reading time from 90° to 50° viewing conditions. Due to the larger increases in reading time for Elographics, ISI, and Microtouch, the Sierracin device's relative position went from worst at the 90^{o} angle to third at 50^{o}.

Looking at the search time data, the important finding was a significant main effect of device. This effect was not robust, probably because all conditions in the search task experiment were conducted under room lit conditions which were optimal for search time performance.

Although the main effect of device was significant, no pairwise comparisons between device means showed significant differences. However, the ordering of device means with respect to search time in the alphanumeric search task is quite similar to that of the Tinker task reading time measure under high glare direct viewing conditions. EMS and Carroll devices, for example, have the shortest average search times while Microtouch and Sierracin devices have the longest search times.

Across viewing angles both within tasks and across tasks, EMS and Carroll touch TSDs are consistently better in terms of reading/search performance than are the other devices. The remaining devices have advantages and weaknesses under specific viewing conditions but, when looked at overall, are not clearly or consistently better or worse than each other with respect to reading/search performance.

Differences among devices, however, are only significant under very poor viewing conditions. For most lighting conditions, differences in reading/search time are negligible and any of the devices would be acceptable.

Entry Time

In Tinker task results from entry time data, two significant main effects of device and light level were present. In the alphanumeric search task, significant two-way interactions of Device by Target Size and Target Size by Viewing Angle were found. Of these significant effects, the main effect of light level and the Target Size by Viewing Angle interaction reflect the successful manipulation of the independent variables across a range broad enough to affect performance. The significant entry time performance effects of interest to a comparison of TSDs were the main effect of device in the Tinker task and the Device by Target Size interaction in the alphanumeric search task.

For the Tinker task main effect of device, Carroll and EMS TSDs performed similarly and were significantly better than ISI, Sierracin, and Microtouch devices. These results are similar to the reading time results for the Tinker task in that operators performed the best on Carroll and EMS TSDs.

The search task Device by Target Size interaction can best be described as large differences among devices at the smallest target size being reduced to small differences at the largest target size. At the largest target size, no significant differences were found among device means.

The ordering of devices and the pattern of significant differences exhibited among devices at the small target size departs from the results discussed to this point. This divergence of results is logical if the demands of visual and motor task elements are taken into account.

Both the Tinker task and the large target size condition of the alphanumeric search task had touch target areas large enough so that fine touch accuracy was not required for speedy task completion. Therefore, results from these conditions primarily reflected only the visual differences between TSDs. The small target size condition of the search task, however, demanded a much higher degree of task accuracy, and this extra demand was reflected in the entry time and touch count dependent variables. Entry time differences between devices at the small target size showed the EMS device to be no different than the Elographics and Sierracin TSDs but significantly better than the Carroll, ISI, and Microtouch TSDs.

Touch Count

Touch count, like entry time, is a visual/motor measure of performance. Touch count focuses more specifically on the portion of the task when the operator's hand is in close proximity with the TSD. This part of the task involves fine motor movement and close eye-to-hand coordination. Also, touch count is not a time measure, so it reflects only accuracy of touch and not both speed and accuracy as entry time does.

For Tinker task touch count data a main effect of device was found. The ordering of devices in terms of touch count matches that of entry time for Tinker task data. For touch count performance Sierracin is significantly better than all other devices, followed by Elographics and EMS. Carroll is significantly poorer than these devices and Microtouch is significantly worse.

For the alphanumeric search task, touch count performance results match those of entry time with a target size by device interaction. A slightly different pattern of significance is evident, with the Sierracin and Elographics devices being significantly better than all others at the small target size.

Summary

In general, the performance evaluation results showed good consistency, for each dependent measure, across the two experimental tasks. The reading/search measure showed one pattern of results favoring devices with the best visual characteristics. Entry time and touch count agreed closely with each other and reflected a different emphasis which favored devices with good touch and visual characteristics.

The two infrared beam technology devices, EMS and Carroll, were best of all devices with respect to reading/search time. The matrix conductive film TSD of Sierracin was best in terms of touch count and entry time. The Sierracin device was relatively poor, however, for reading under daytime off-angle conditions.

Overall, the EMS TSD performed well across both tasks under all conditions and across all dependent measures. The Microtouch device was significantly poorer than all other devices in some analyses and generally ranked fifth or sixth across all measures, tasks, and conditions. All other devices had weaknesses and strengths which caused them to perform relatively better or worse under different task conditions and measures.

Interestingly, the EMS and Carroll TSDs differed with respect to touch-accuracy-related performance. Carroll and

EMS TSDs are identical in terms of visual and resolution characteristics. In fact, the only hardware characteristic upon which they differ significantly is parallax. The Carroll device has the largest measure of parallax, which it shares with the ISI TSD. EMS is second in terms of parallax to the Sierracin TSD. From the match between rankings of parallax measures and search task entry time performance at the small target size, it can be inferred and appears logical that parallax is an important contributor to accurate touch performance.

MODELING DISCUSSION

Correlations and Single Predictor Models

It was pointed out in the Modeling Results section that most hardware measures, by their nature, were expected to have a certain relationship with performance measures. For example, image quality, S/N ratios, and resolution should be positively correlated with performance while noise metrics and parallax should correlate negatively with performance. The signs of correlations in this evaluation were expected to be reversed, because for each of the performance measures a higher score meant poorer performance.

The signs of the correlations reported in Tables 34 to 37 generally follow the initially expected pattern with a couple of notable exceptions. First, RTF does not act like a noise measure because it is consistently negatively correlated with read/search, entry, and total time measures. Second, the relationship between touch count and all of the image quality measures is opposite of what was expected.

RTF is a measure of the ability of the first surface of the TSD to reflect light at an angle equal to the angle of incidence of that light. In this experiment, the subject always had an evenly lighted background reflected into his/her eyes from the TSD surface. In nighttime and office lit conditions, a black drop cloth provided the background and in daytime conditions, an equal luminance field was used. Therefore, no interfering high information images were presented at an angle to be reflected directly into the subjects eyes.

Conversely, RTF is also a measure of the ability of the first surface of the TSD to block interference from light sources that strike the device at angles other than that one angle equal to the angle of incidence. So in this study, RTF could be expected to be positively related to performance because it represents a measure of the ability of a TSD to block off angle glare from hitting the subject's eyes.

RTF acting like a signal strength rather than noise measure has an impact on the signal/noise ratios that have RTF in the denominator. All of the S/N metrics with RTF as denominator are positively correlated with reading, entry, and total time performance measures. On closer examination, it can be seen that relative changes in RTF from device to device are large in comparison to changes in MTF or other signal strength metrics. This makes the impact of RTF in the ratio greater than the numerator.

Thus, each S/N ratio with RTF is in effect an inverse of RTF weighted by an image quality metric.

As an inverse RTF metric, each of these S/N ratios would be expected to have a relationship with performance that is opposite that of RTFs. Also, since an inverse is a nonlinear transform, it would be expected to change the strength of the relationship between RTF and performance. In this case the weighted transforms of RTF represented by MTF/RTF, MTFA/RTF, EP/RTF, and SSF/RTF each have a stronger relationship with performance than RTF alone or any of the numerators alone except MTF and MTFA.

The second seemingly contrary result in the correlation data is the relationship between image quality noise measures and touch count performance. The noise measures RMS and WS in particular have fairly strong negative correlations with touch count. Also strongly related to touch count are any S/N ratios with RMS or WS in the denominator.

These findings can be better understood when it is known that RMS has a .99 correlation with touch weight and WS has a .55 correlation with touch weight. Therefore, RMS and touch weight are effectively the same measure. RMS and WS are also related to image quality so touch weight and image quality are confounded by coincidence in the six TSDs evaluated.

In general, a few hardware measures have stronger relationships with performance than the others. Ambient MTF has the highest correlation with reading time. RMS (touch force) has the strongest relationship with touch count performance. The best of the signal/noise metrics are anything with RTF in the denominator, especially MTFA/RTF which combines RTF with the best of the Image Quality metrics. WS and RMS are generally poor predictors of performance except touch count performance where they may be passing for touch force.

Best Models Across Conditions

Modeling individual performance across conditions. The models presented in Tables 40 and 41 were generated using the full Tinker and alphanumeric search task data sets. These are the most generalizable of the models produced in this evaluation. They are fitted to performance data taken under a variety of conditions and which include interindividual differences. As a consequence of the wide range of conditions across which they predict performance, full data set models account for the smallest proportions of performance variation of any models in the study. The tradeoff between generalizability and descriptive power is a primary consideration for the application of all of the models presented in this report.

Prior to the experiments, it was expected that touch characteristics, especially the resolution metrics, would play an important role in modeling of the alphanumeric search task performance. The expectation was due to the random placement of targets in the task. Entry time, touch count, and total time were expected to reflect the influence of touch related factors and should therefore be of primary importance in models of search task performance.

The Tinker task, on the other hand, is a reading task of the key entry type and as such should be most heavily impacted by image quality factors. Reading time and total time were expected to be the important measures in this task and models of their performance most interesting.

In Table 40 it can be seen that ambient MTF and parallax form the basis for the models of reading time and total time. The fact that the total time model is so similar to that of reading time confirms the importance of the visual part of the Tinker task to overall task performance.

In the alphanumeric search task, S/N ratios with RTF as denominator dominate entry time, and touch count. These ratios would be expected to enter into the total time model due to the emphasis in the search task on the touch aspects of the task. Total time, however, was modelled by the same variables that make up the search time model. This indicates that for both tasks, across conditions, the visual interaction with the display prior to touch is an important factor in the time taken to complete the tasks.

The entry of Touch weight into total time performance points out the increased importance of touch characteristics to search type tasks.

Ambient MTF in the Tinker reading time model is replaced by MTF, WS, and RMS in the search time model. This change makes sense when one considers that ambient MTF takes ambient lighting into account and that the search task has only one level of ambient lighting. Therefore, the utility of ambient MTF is greatly reduced in search task performance prediction.

Modeling average performance across conditions. The main advantage of the models of "average operator" performance in Tables 52 and 53 over the models previously discussed is that they have eliminated a fairly large source of unaccounted for variance in performance. This reduction allows hardware parameter models to fit the data better. The remaining unaccounted for variance in the data is consequently a better estimate of variability in performance due to unknown factors.

Modeling of average rather than individual performance does not reduce the generalizability of models across conditions. It must be remembered though that these models fit estimated population mean responses and, as such, cannot be compared with models of individual performance from other sources.

The most notable change brought about by averaging performance data across subjects is a general increase in accounted for variance. For the Tinker task, R^2s are about double those from individual data models. Alphanumeric search task "average operator" models show large increases in the R^2 for search and total time, but not in entry time and touch count.

The measures making up "average operator" models generally show the same influences and trends as the individual performance based models discussed above. Ambient MTF is important in the Tinker task reading time and total time models, and S/N ratios featuring RTF in the denominator are again important in alphanumeric search task models.

Best Models for Specific Conditions.

Individual performance models. The models in Tables 44 to 49 fit individual performance data at specific levels of lighting and target size. Like general models of individual performance, they are applicable to the prediction of a future operator's performance. Unlike general performance models, these equations are restricted in their applicability to the specific conditions of lighting and target size for which each was constructed. If an application matches the conditions under which one of these models was generated, that model is likely to predict performance better than one constructed with the full data set.

R^2 values for these equations are generally a little higher than for full data set individual performance models but not as large as for full data set average performance. Values fall in the .1 to .22 range for alphanumeric search task and .1 to .3 for Tinker task models.

Tinker task models at nighttime and office light conditions do not account for a significant proportion of read time or total time performance. This result indicates that individual differences in reading performance and other uncontrolled factors were large in comparison with reading performance differences caused by hardware differences between devices when lighting conditions were not demanding.

The values for image quality metrics reported in the hardware measures results fall in the range generally considered to be good to excellent when evaluating CRT produced images. An MTF area of 1.0 mod/cycle/mm, for example is generally achieved only by high quality monochrome VDT displays and is seldom attained by color monitors. Therefore it is not suprising that no significant effect of device was found at low and room lit conditions and that subsequent image quality based models of Tinker task reading performance for these conditions did not predict performance well.

Under daytime conditions, ambient visual noise levels are greatly increased, and the quality of the perceived image (as seen in the ambient MTF metric) is reduced to a level where performance suffers measurably. In these extreme conditions, image quality differences between TSDs had a greater impact on performance.

With daytime conditions, reading and total time performance R^2 values were considerably higher than for nightime and office lit conditions. Both reading and total time were best fit by models containing the same parameters and each of these models accounted for approximately 25% of the variance in performance.

As would be expected from the models in Step 2, signal to noise ratios are important in daytime condition models. EP/RMS and MTF/RMS, in particular, added to the ability of these models to fit the data.

The alphanumeric search task with its random placement of target and distractor items on the screen was quite variable with respect to search time performance.

Therefore, it is not unusual that search time performance was generally not well modeled.

Entry time, touch count, and total time models across the two smaller target sizes contained S/N ratios with RTF as the denominator. The contributions of RTF to models of conditions requiring fairly accurate touch positioning indicate that the visibility of the first or touch surface may have an impact on accurate touch positioning and overall task performance when touch requirements are demanding.

In the alphanumeric search task experiment, the simulated office lighting condition was used for all target size and viewing angle conditions. Therefore, in all search task conditions, the only image reflected directly into the subject's eyes from the first surface of the TSD/display was that of a flat black background drape cloth. This represents ideal viewing conditions for a highly specular TSD/display surface with direct glare blacked by a dark background and diffuse glare blacked by the specularity of the surface. It is important to note that under these conditions measures which discriminate between devices in terms of their visual characteristics play the major role in predicting user performance.

For targets requiring less accuracy of touches (large target size) touch weight, parallax, and ER1 each contributed to touch-related performance. An emphasis on touch characteristics such as parallax and resolution in search task models was expected because of the importance of touch to task performance. However, it was suprising that these measures did not also enter into models for the more demanding smaller target sizes.

Average operator models. All models based on specific sets of conditions share the advantage of being focused and therefore better models if one's application matches the conditions. In addition, modeling of performance averaged across subjects, removes variation due to individual differences from modeled data. This allows a clear picture of the impact of hardware parameters on general performance. What is lost in these models is an estimate of their ability to predict the performance of one individual.

Variance accounted for by models of average performance under specific conditions is dramatically improved over all other steps in this analysis. R^2 values fall generally in the 0.5 to 0.95 range. This result is quite good and shows that what we know and have measured about the human/TSD interface accounts for most of the variation in performance that was observed.

For reading and total time performance of the Tinker task, noghtime and office conditions were still not very well modeled compared to the daytime condition. Under nighttime conditions, no variable accounting for á significant proportion of the variance was found for total task time performance.

Under daytime conditions, however, 92% of total task time variance was accounted for. The model for this condition was quite similar to the daytime model for individual performance data with MTF/RMS and EP/RMS being the most important contributors.

Alphanumeric search task models of average performance resembled their individual data counterparts with an emphasis on S/N ratios with RTF in the denominator. Also, when specific conditions are modelled, total time models are generally dominated by the predictor variables that appear in the touch-related performance models of entry time and touch count. This again indicates the importance of the visual/motor portion of the alphanumeric search task performance.

Summary of Modeling Analysis

The following list of general statements summarizes the discussion of the seven-step performance modeling effort.

1. R^2 increases as the conditions being modeled become more specific.

2. The impact of particular predictor variables changes from one specific condition to the next.

3. Models of individual performance usually include the same variables as their average performance counterparts.

4. Looking at mean rather than individual performance allows a focus on the effects of hardware differences on performance. It also shows how much of the variability in performance can be accounted for by what was recorded about the human/TSD interface.

5. The "best model" depends upon its intended application. A tradeoff between generalizability and goodness of fit must always be considered.

6. Models of menu selection type tasks are influenced by the purely visual portions of the task while search type tasks are best fit by models emphasizing the visual/motor aspects of the task.

7. Image quality based predictor variables are important to both task types while touch characteristics have an impact primarily on search type tasks.

8. Total task time is a good general measure of performance because it reflects the influence of both visual and visual/motor task elements.

CONCLUSIONS

Performance Evaluation Conclusions

A primary goal of this dissertation was to compare and evaluate TSD technologies in terms of operator performance across a variety of conditions. This goal was approached through the performance testing of six TSDs.

Taking into account both task types and several measures of performance, the EMS TSD came out as the best all around device. The Sierracin TSD performed as well as or better than the EMS TSD in touch-related task components but was not as good in the purely visual elements of the tasks.

The "analog" devices (Microtouch, ISI and Elographics) did not show exceptionally good performance in either touch-related or visual task elements. From their high limiting resolution capabilities, it would be expected that these devices would do quite well in the alphanumeric search task, especially under small target size conditions. The most likely reason that they did not perform well under these conditions was that each of the analog devices has some problems with linearity of touch coordinates and with variability of the coordinates reported from a single touch location.

The Carroll Touch TSD, which is the same technology as the EMS, was equal to EMS in reading time performance but fell behind in the touch related portions of each task. The fall-off in Carroll TSD performance during the touch portion of each task might be attributed to the one hardware parameter on which EMS and Carroll differ. This parameter is the amount of parallax in the TSD-display interface. The touch-sensing grid of the Carroll TSD was mounted 10 mm from the surface of the display versus 6.5 mm for the EMS device. This difference in mounting created a large difference between the devices in parallax. All other hardware aspects of these two TSDs were virtually identical; therefore, it is likely that parallax was responsible for the decline in Carroll TSD performance during the touch portion of the tasks.

There are several factors which might explain why a low-parallax IR-beam technology would perform better than other TSDs across a variety of tasks and conditions. The IR beam does nothing to change the high quality image produced by the electroluminescent display. It is the only technology which does not require a modification of some sort to the display surface. The limiting resolution of IR TSDs is not as high as the capacative and analog resistive film TSDs, but the excellent linearity and stability of the coordinates reported by beam devices makes their effective resolution comparable to or better than the "analog" devices.

Modeling Conclusions

Modeling of performance with TSD hardware parameters led to a few general conclusions. For example, image quality metrics, especially S/N ratios with RTF in the denominator, are important in predicting both visual and visual/motor performance. RTF appears to be a particularly good predictor under conditions where there is an opportunity to block direct specular reflections with a dark background. Image quality metrics are particularly important to the prediction of reading time performance. These measures could be quite helpful to designers of an interface used primarily for data acquisition.

Touch characteristics contribute significantly to tasks which require accurate touch inputting. Of the touch characteristics measured in this study, parallax had the most consistent impact on task performance. Increases in TSD/display parallax consistently led to poorer entry time and touch count performance.

Overall, it is recommended from the results of this comparative evaluation that the EMS infrared beam TSD is the best current solution to a small-area TSD/EL panel interface. However, because of the comparative nature of this research, these conclusions do not generalize beyond the technologies evaluated. Other design solutions that are quantitatively better in the image quality and/or touch characteristics identified as important in the modeling evaluation would be prime candidates for inclusion in future comparison studies.

RECOMMENDATIONS FOR TSD INTERFACE DESIGN

The EL panel with its flat surface and bright high-resolution image is a very good display for a well-designed TSD/display interface. It offers opportunities for performance improving integration which were not fully developed by any of the products tested in this study. The primary recommendation from this dissertation is that regardless of the TSD technology selected, performance can be greatly facilitated with a cooperative effort between the display and TSD manufacturers with special attention to critical factors pointed out in this research.

The first example of a factor requiring a cooperative solution is the problem presented by parallax. The example of the EMS and Carroll TSDs, similar in all aspects other than the distance they were mounted from the display and their performance as touch-input devices, illustrates the importance of this point. All of the devices tested could have greatly reduced parallax if they had been able to use the display surface and housing as the structure for their devices. The flat surface of the display would allow the IR beam or any of the other technologies to be positioned within a few millimeters of the display surface rather then 6 mm or more as were all the devices tested. Using the display surface as the base could have improved the interface of the capacitive and conductive film devices by reducing parallax and also by reducing the number of surfaces across which the light from the display is required to pass. Thus, any of the devices in the study could have been improved by a cooperative effort with the display manufacturer to reduce the distance between the displayed image and the touch sensor array.

Significant contributions of image quality and especially RTF to reading and accurate touch performance point to another aspect of the interface which could greatly benefit from good system integration. The best TSD interface is one that does the least to degrade the image presented on the EL display. A highly specular first surface best preserves this image quality while blocking diffuse glare. The potential problem with a highly specular surface is that it by definition reflects direct glare quite well. This problem might best be approached by a direct glare treatment applied directly to or incorporated in the display glass or TSD film layer. A solution developed specifically for the emissive qualities of the EL display would interfere least with image quality and provide the best glare reduction.

REFERENCES

Abramson, S.R., Mason, L.H., and Snyder, H.L. (1983). The effects of display errors and font styles upon operator performance with a plasma panel. In Proceedings of the Human Factors Society 27th Annual Meeting (pp. 28-32). Santa Monica, CA: Human Factors Society.

Abramson, S.R., Mason, L.H., and Snyder, H.L. (1984). Operator performance on flat-panel displays with line and cell failures. (Technical Report HFL 83-3). Blacksburg, VA: Virginia Polytechnic Institute & State University, IEOR Dept.

Beaton, R.J., Schulze, L.J.H., and Snyder, H.L. (1983). Touch entry devices and human performance. Part I: Quantifying the display quality. Digest, Society for Information Display, 162-163.

Beaton, R.J., and Snyder, H.L. (1984). The display quality of glare filters for CRT terminals. Digest, Society for Information Display, 298-301.

Beaton, R.J., and Weiman, N. (1984). Effects of touch key size and separation on menu selection accuracy. (Technical Report No. HFL-500-02). Beaverton, OR: Tektronix, Inc.

Beringer, D.B., and Peterson, J.G. (1983). Touch input devices: training the device or the operator? In Proceedings of the Human Factors Society 27th Annual Meeting (pp. 206-210). Santa Monica, CA: Human Factors Society.

Beringer, D.B., and Peterson, J.G. (1985). Underlying behavioral parameters of the operation of touch-input devices: biases, models, and feedback. Human Factors, 27(4), 445-458.

Cornsweet, T.N. (1970). Visual Perception. New York: Academic Press.

Davis, G.I., and Badger, W.S. (1982). User-computer interface design of a complex tactical display terminal. In Proceedings of the Human Factors Society 26th Annual Meeting (pp. 768-771). Santa Monica, CA: Human Factors Society.

de Bruyne, P. (1980). Acoustic radar graphic input device. Computer Graphics, 14(3), 25-31.

Gaertner, K.P. and Holzhausen, K.P. (1980). Controlling air traffic with a touch sensitive screen. Applied Ergonomics, 11, 17-11.

Greenstein, J.S., and Arnaut, L.Y. (1987). Human Factors aspects of manual computer input devices. In G. Salvendy (Ed.) Handbook of Human Factors. New York: Wiley.

Hopkin, V.D. (1971). The evaluation of touch displays for air traffic control tasks. IEE Conference on Displays (Conf. Publ. #80), 83-90.

Hunter, M.W., Reger, J.J., Farley, W.W., and Snyder, H.L. (1985). Photometric evaluations of the glare suppression and image quality characteristics of various CRT surface/filter combinations. (Technical Report). Blacksburg, VA: Virginia Polytechnic Institute & State University, IEOR Dept.

Johnson, E.A. (1967). Touch displays: A programmed man-machine interface. Ergonomics, 10, 271-277.

Logan, J.D. (1985). Touch screens diversify. Electronic Products, Nov.1, 61-67.

Monty, R.W., Snyder, H.L., Farley, W.W., Donohoo, D.T., and Baggen, E.A. (1985). Eye movements and driver performance with electronic automotive displays. (Technical Report). Blacksburg, VA: Virginia Polytechnic Institute & State University, IEOR Dept.

Pfauth, M. and Priest, J. (1981). Person-computer interface using touch-screen devices. In Proceedings of the Human Factors Society 25th Annual Meeting (pp. 500-504). Santa Monica, CA: Human Factors Society.

Ritchie, G.J., and Turner, J.A. (1975). Input devices for interactive graphics. International Journal of Man-Machine Studies, 7, 639-660.

SAS User's Guide: Statistics (1982). Cary, NC: SAS Institute Inc.

Schulze, L.J.H., Beaton, R.J., and Snyder, H.L. (1983). Touch entry devices and human performance. Part II: Assessing operator performance. Digest, Society for Information Display, 164-165.

Schulze, L.J.H., and Snyder, H.L. (1983, October). A comparative evaluation of five touch entry devices. (Technical Report No, HFL-83-6). Blacksburg, VA: Virginia Polytechnic Institute & State Iniversity, IEOR Dept.

Slabey, C.A., Hartley, C.S., and Pulliam, R. (1985). Speech recognition, speech systhesis, and touch-panel technology applied to the control of remotely piloted spacecraft. In Proceedings of the Human Factors Society 29th Annual Meeting (pp. 1140-1143). Santa Monica, CA: Human Factors Society.

Snyder, H.L. (1980). Human performance and flat panel display image quality. (Technical Report No. HFL-80-1). Blacksburg, VA: Virginia Polytechnic Institute & State University, IEOR Dept.

Snyder, H.L. (1973). Image quality and operator performance. In L.M. Biberman (Ed.) Perception of displayed information. NY: Plenum.

Snyder, H.L. and Taylor, G.B. (1979). The sensitivity of response measures of alphanumeric legibility to variations in dot matrix display parameters. Human Factors, 21(4), 457-471.

Stammers, R.C. and Bird, J.M. (1980). Controller evaluation of a touch input air traffic data system: An "indelicate" experiment. Human Factors, 22, 581-589.

Tinker, M.A. (1955). Examiner's Manual for Tinker Speed Reading Test. Minneapolis, MN: University of Minnesota Press.

Waller, L. (1984). Military workstation is easy to reprogram. Electronics, Mar.22, 50.

Weiman, N., Beaton, R.J., Knox, S.T., and Glasser, P.C. (1985). Effects of key layout, visual feedback, and encoding algorithm on menu selection with LED-based touch panels. (Technical Report No. HFL-604-02). Beaverton, OR: Tektronix, Inc.

APPENDIX A

INSTRUCTIONS

The display that you see in front of you is equipped with a touch-entry device. This device allows you to communicate with a computer by touching the display screen. Your task in this experiment will be to ind different target words or letters on the screen, then signal to the computer that you have found the target by touching it with a finger. There will be two tasks, a reading task and a random letter search task.

TASK 1 INSTRUCTIONS

In this task you will be shown a two-sentence paragraph. Somewhere in the second sentence will be a word that does not fit with the meaning of the rest of the paragraph. To complete the task you must read the paragraph, pick out the word that doesn't fit, and point to that word.

When you touch the screen, you will hear a beep indicating that the computer received the touch. If you hit the correct target location, the screen will go blank and a new trial will be started. If the screen does not go blank, you must remove your finger and try again.

TASK 2 INSTRUCTIONS

Task 2 is a random pattern search task. Before each trial you will be shown a letter which will be your target for the trial. When the trial starts, a random pattern of letters will appear on the screen. You must find the target letter and point to it. The target will remain on the screen until you have successfully touched it.

Both tasks are self-paced. You will start each trial by pressing the black button on the grey box in front of you with your non-preferred hand. Your dominant hand must be resting on the brown board at the start of each trial. You must have this hand in the start position until you have found the target word or letter. When the target has been identified you must point to it using your dominant hand (the one on the board).

It is important that you leave your "pointing" hand at the start position until you have selected a target to point to. Therefore, a number of false trials have been included as a check that you do not move prematurely. A false trial will start like any other trial but when the pointing hand is moved, the screen will go blank and you will be required to touch the location on the screen where the target was last seen. Accuracy of this touch will be a measure of whether or not you had properly identified the target before moving.

The vita has been removed from the scanned document