2006

Evaluation of computer input devices for use in standard office applications while moving in an off-road vehicle

Thorsten Baldus
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Agriculture Commons, Bioresource and Agricultural Engineering Commons, and the Industrial Engineering Commons

Recommended Citation

Evaluation of computer input devices for use in standard office applications while moving in an off-road vehicle

by

Thorsten Baldus

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Industrial Engineering

Program of Study Committee:
Patrick E. Patterson, Major Professor
Samuel K. Adams
Roger E. Baer
John K. Jackman
Mack C. Shelley

Iowa State University

Ames, Iowa

2006

UMI Number: 3229052

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Thorsten Baldus

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

TABLE OF CONTENTS

GENERAL INTRODUCTION

1.1 INTRODUCTION

With computers taking over ever more control functions in modern vehicles, automated systems like traction control and Anti Lock Brakes increase safety while assisting the operator with the driving task in extreme conditions. It is easy to imagine that automated driving systems will be developed in which the operator will merely have to monitor the vehicle while an automated driving system controls it.

While these types of systems are still in the development phase for on-road vehicles, semi-autonomous controls are already available for use in tractors. One example of such a system is John Deere's Green-Star Auto-Trac System™, which is a hands free steering system for straight-line driving applications. Steering is automated, with the operator's task reduced to monitoring tractor operations and turning the tractor around when the end of the field is reached.

Future generations of this system are expected to include path planning software and implement control functionality. This means once a field work task is programmed and started all driving operations including turning when the end of the field is reached and all control functions will be performed by the system.

With such systems, the computer is taking over more and more of the operator's workload and the operators task will switch from steering the tractor and controlling the implement to a monitoring task making sure everything works as intended.

When using a tractor that is equipped with one of these systems, the work load of the operator would be lightened and enable the performance of other tasks while doing fieldwork. For this reason many people believe a tractor cab could become a mobile office in the near future in which the operator could uses a personal computer to take care of all types

of office tasks while working in the field. With the advances in mobile technology, it is easy to imagine that in the near future a farmer could browse the Internet to find weather data or market prices or sell a harvest while sitting in the tractor doing fieldwork.

1.2 OBJECTIVES

The goal of this study was to investigate the feasibility of using a computer, running standard office applications such as Microsoft Office, in a moving off-road vehicle with semi autonomous controls.

Most standard office applications require the user to interact with the software in two different ways. One is by manipulating a pointer used in a graphical user interface to select menu buttons, move the cursor, drag and drop items etc. and the second is via a keyboard to realize various functions and alphanumeric input. Several studies (MacKenzie 1991, Bohan et.al. 2003, Cerney et.al 2004, Pavlovych and Stuerzlinger 2004) established the dominance of the mouse and the standard QWERTY keyboard as the interface devices that provide the highest usability in an office environment. It is expected that due to factors like movement and vibration the performance achievable with a computer system in a mobile off-road environment will be lower than the performance achievable in an office environment and that therefore other devices might be better suited for this environment.

To determine if a system that provides a tractor operator with a computer workplace in a tractor cab is feasible a series of factors have to be considered.

What interface devices provide the best usability to the operator and are therefore best suited for users in this environment

What user performance can such a system support as compared to the performance in an office environment and how does a user rate its usability

1.3 THESIS ORGANIZATION

To investigate these issues this study was conducted in three parts. The first part focused on pointing devices with the goal of determining which commercially available pointing device is best suited for use in the given moving environment and what factors influence the usability. The second part focused on alphanumeric input devices with a goal to determine which of the commercially available alphanumeric input techniques provides the best usability in the tractor environment. The goal of the third part was to determine the performance that could be achieved with such a system compared to an office environment and how a user rates the usability of such a system.

Part 1: Evaluation of pointing devices

The performance of standard pointing devices in a moving off-road environment (a tractor) was examined to develop a general understanding of use within such an environment. The devices were evaluated to determine those that allow the user to perform office computer applications in a tractor best and to determine the performance parameters a user can achieve with those devices. Off-road environmental factors, such as vibration of work surfaces and the relative movement of the operator's body to the input device can have a significant influence on the performance. It was therefore expected that the performance the operator could achieve with the tested devices would decrease in a mobile environment as compared to a stationary one, leading to lower performance values and higher error ratios when compared to a stationary environment. The performance a user could achieve with the tested devices was evaluated using throughput, task time and error rate as performance measurements. Initially a pilot study was conducted to evaluate the usability of a wide variety of commercially available pointing devices for their use in the tested environment. In the pilot study multi directional discrete pointing tasks were used to evaluate the performance a user could achieve with the tested devices. The devices allowing the best performance in the pilot study were than included in a more thorough study. Test software was created that presented a multi-stage selection task which simulated the use of a pointer in a standard windows application. The objective performance and the subjective user evaluation of the tested devices determined those that provide the best usability in the tested environment and indicated the performance a user can achieve with those devices.

Part 2: Evaluation of alphanumeric input devices

The second part of this study was designed to determine which alphanumeric input device was best suited for used in the tested environment. It was expected that, as in the first part, factors like vibration and movement of the participant and the input device will have a negative effect on the usability in the moving as compared to the stationary environment. To evaluate these factors a wide variety of input devices was presented to a focus group to determine which of the devices and techniques are expected to provide good usability and allow a good performance and therefore should be evaluated in more detail. The devices that were selected by the focus group were than included in a study to determine which provides the best usability and allows the highest performance in a moving tractor. Eighteen test subjects were asked to perform a series of typing tasks with each of the test devices providing objective performance data and subjective usability ratings to determine which device is better suited for the moving environment.

Part 3: Evaluation of overall performance

The goal of the third part of the study was to determine overall performance and usability that can be expected when using a standard application in the tested environment compared to an office environment. The devices determined to be best suited for the moving tractor environment in the first two parts of the study were combined to provide a full user interface. The performance a user can achieve with the resulting interfaces in a moving tractor was than compared to the performance a user can achieve with a standard application in an office environment. A subjective usability rating and subjective impression of performance with the tested interfaces was also evaluated.

1.4 LITERATURE REVIEW

Previous work in this area was primarily to evaluate the performance of interface devices in either an office environment or for mobile devices like laptops or PDAs, and focused on either pointing devices or alphanumeric input.

Fitts' law for the information processing capacity of the human motor system (Fitts, 1954) was developed to predict performance of the human motor system. MacKenzie (1991) examined the theory, prediction power, and relevance of Fitts' law as a performance model in human computer interaction. He extended the Law's applicability to movement tasks that are common on interactive computer systems like dragging and compound dragging-plus-pointing tasks. He developed two methods to accommodate the approach angle in the calculation of task difficulty and demonstrated that movement time, error rate, and the index of performance are all affected by the choice of device or task.

McKenzie et.al. (1991) compared a mouse a tablet with stylus and a trackball in the performance of pointing and dragging tasks both modeled after Fitts' reciprocal tapping task. In this study the tablet with stylus and mouse performed similar during the pointing task but the tablet with mouse was slightly better for the dragging task. The trackball had a lower performance than the other two devices for both tasks. Another finding was that a dragging task requires more time and produces more errors than does a pointing task across all tested devices.

Bohan et.al. (2003) compared objective performance and subjective user preference of a RollerMouse to a standard mouse. The participants were asked to perform a series of tasks representative of the actions typically performed using input devices. The independent variables were the approach angle, the target size, the target distance, and the task (pointing or dragging). The standard mouse had better performance in pointing and dragging for all target sizes and distances. The participants also rated the standard mouse as significantly easier to use than the RollerMouse in their subjective evaluation.

While the mouse is the most popular pointing device for an office environment for mobile devices like laptops or PDAs, other devices are commonly used.

Douglas et.al. (1999) evaluated the scientific validity and practicality of the ISO 924, Part 9 draft for two pointing devices used with laptop computers. This draft represented an

international standard for testing computer pointing devices for both performance and comfort. Performance was evaluated by measuring the throughput for one-directional and multi-directional pointing and selecting tasks. Throughput is the ISO recommended primary measurement of performance. It is the Fitts' index of performance using the effective width instead of the actual measured size of the target to incorporate the variability observed of human performance by including both speed and accuracy (Douglas et. al., 1999). Additionally, the standard requires the subjective evaluation of the tested devices. A post test questionnaire is administered allowing participants to rate the tested devices for comfort including operation, fatigue, and usability. In the course of this study two pointing devices for laptop computers, the isometric joystick (TouchStick) and the touchpad were evaluated. The *Generalized Fitts' Law Model Builder* (Soukoreff and MacKenzie 1995) was used to present both one and multidirectional pointing tasks to the participants. The isometric joystick performed better in the objective performance measurement with no significant difference between the two devices in subjective comfort ratings. Based on the results of his study Douglas (Douglas et. al. 1999) recommended that while throughput allows a single measure of both speed and accuracy to compute both movement time and error rate as separate dependent variables. .

While no work was found on testing of pointing devices in moving vehicles, Rider et. al. (2003) analyzed the effect of land-vehicle ride motion on the movement time and accuracy of in-vehicle reaching tasks. In this study "soft" buttons on a touch screen and "hard" push-buttons were used to present reaching and pointing tasks in a moving vehicle under different driving conditions, for different reach distances, different reach directions, and different target sizes. To evaluate performance, the movement time and spatial error were recorded. The participants were also asked to complete a post test questionnaire. It was found that movement completion time for reaches, accuracy in pointing tasks, and perceived difficulty ratings were adversely affected by vehicle motion.

Besides pointer operation alphanumeric input is the second common way to interact with a standard application. While the omnipresent QWERTY keyboard is the primary text entry

device on desktop systems, to date mobile and handheld systems lack an equivalent dominant technology or technique for the same task.

MacKenzie and Soukoreff (2002) studied text entry techniques for mobile computing. They did not evaluate the performance of any of the devices but instead discussed key factors for the evaluation of new technologies for mobile text entry including the methodology and the experimental design. They indicated that the most important issues for consideration when evaluating text entry techniques for mobile computing are focus of attention, text creation vs. text copy tasks, novice vs. expert performance, quantitative vs. qualitative measures, and the speed-accuracy trade off. They also discussed the optimization of different text input techniques by movement minimization and language prediction. They provided a survey of mobile text entry techniques both in research papers and in commercial products.

Cerney et.al (2004) conducted a study to compare the standard QWERY keyboard with different alternatives of mobile text entry methods such as an onscreen QWERTY soft keyboard, a letter recognition system on a pocket PC, and a T9 text-input system on a cellular phone. Users were asked to perform common input tasks and user performance, accuracy and overall user preference for the four input methods were examined. They also compared the measured performance with the performance for the different input devices as predicted by Fitts' law. In this study, the standard QWERTY keyboard had the best performance and the lowest cognitive load indicated by the users. This study also demonstrated that Fitts' law does not account for, nor predict, performance values including cognitive load or skill transfer and is only valid if the user's familiarity of the device placed interaction on a level of pure motor reflex.

MacKenzie (2002) evaluated different techniques for three key text entries for different mobile devices. Ten participants were asked to enter a series of phrases from the 'Phrase Sets for Evaluating Text Entry Techniques' [MacKenzie and Soukoreff 2003]. Interaction issues that were examined in this study included the use of linguistic knowledge to accelerate input

and the challenges in using auto repeat keying strategies to reduce the number of physical key presses.

Fleetwood et. al. (2002) evaluated two text entry methods common to Personal Digital Assistants (PDA), an on-screen virtual keyboard and a single-character handwriting recognition system called Graffiti. 48 expert and novice users and their character entry rates were evaluated with respect to a predicted rate of entry based on Fitts' and the Hick-Hyman laws for the virtual keyboard and pen and paper printing for Graffiti. Better performance using Graffiti was found for the expert users while the novice users had better performance using the virtual keyboard.

One of the most common methods for text entry in mobile computing devices is the use of a virtual soft keyboard on a touch screen. Several studies have evaluated the performance for different layouts. Pavlovych and Stuerzlinger (2004) compared the performance of a traditional QWERTY layout soft keyboard with several other layouts. A set of phrases of English text from MacKenzie and Soukoreff (2003) was used to present the subject with typing tasks with the text entry speed and error rate measured to evaluate the performance of the tested layouts. This study showed that for a population of regular keyboard users the QWERTY layout was the best alternative with the highest speed and better error rate.

MacKenzie and Zhang (1999) showed that for expert users other layouts of soft keyboards could achieve higher performance than the standard QWERTY layout. In their study a longitudinal evaluation over 20 sessions showed that the average entry rates using an OPTI layout increased rapidly and exceeded those for the QWERTY layout.

Zhai et. al. (2000) evaluated two different computerized quantitative design techniques to search for the optimal keyboard layout, comparing their estimated performance to four existing keyboard layouts. To predict the performance of the different layouts a performance model for a virtual keyboard was established, summing the Fitts' law movement times between all digraphs, weighted by the occurrence of the digraphs. This model only describes

the movement time for tapping on a key by an expert user and doesn't reflect the visual search required by a novice or intermediate user. Based on this model, better performance was found for the new layouts.

Sears et al. (1993) investigated the relation between keyboard size and data entry rates for soft keyboards. As a result of this study it was shown that users made, and corrected, fewer errors when entering text on larger keyboards, with data entry speeds higher on larger keyboards, and the larger keyboards had better results in the subjective user preference.

Soukoreff and MacKenzie (2003) identified and described shortcomings in two statistics used to measure accuracy in text entry evaluations, minimum string distance (MSD) error rate and keystrokes per character (KSPC). In their study they developed a new framework and new measurements for evaluating text entry methods but fail to provide means to measure the bandwidth between humans and machines during text entry that includes both speed and accuracy.

MacKenzie and Soukoreff (2003) describe and publish a phrase set for evaluating text entry techniques. The phrase set is a collection of 500 phrases of moderate length that are easy to remember and are representative for the English language. They also provide utility programs to calculate statistical properties like letter and word frequencies for this and other phrase sets.

ISO 9241(1998) is an international standard of ergonomic requirements for office work with visual display terminals. The different parts of this standard regulate different aspects of the design and evaluation of office computer workstations. Part 4 regulates the design factors and evaluation of keyboards, part 9 the design and evaluation of non-keyboard input devices, and part 11 provides a guide to measure the usability of a computer work place.

Much research has been undertaken to evaluate the performance and usability of different input devices for a wide variety of application and environments. According to these studies

the mouse and the QWERTY keyboard are the dominant input devices for standard applications in an office environment. For other applications and other environments a wide variety of other devices seem to have benefits for specific situations. Even though none of the prior research evaluated the use of input devices for standard applications in a moving vehicle, the research done for other environments and tasks provide a good source of established evaluation techniques and testing methods which were adopted to be used in the course of this study.

Evaluation of the usability of input devices in elemental pointing tasks in a moving off-road environment

A pilot study

Thorsten Baldus

ABSTRACT

Five standard input devices (mouse, touch stick, touch pad, touch screen, trackball) were evaluated to determine those best suited for performing office computer application tasks in a mobile off-road environment (a tractor). Fifteen subjects performed a series of multi-directional discrete pointing tasks in a standstill tractor and while moving in a tractor. The mouse, touch screen and touch pad seemed well-suited in some aspects and poorly in others and need further testing in this environment. With the trackball and the touch stick the performance of the participants was poor and these devices are not recommended for use in the tested environment.

1.0 INTRODUCTION

Elemental pointing tasks are a fundamental control mechanism in computer interfaces. Many studies have evaluated the performance parameters of pointing devices for these tasks.

McKenzie et.al. (1991) compared the usability of a mouse, a tablet with stylus, and a trackball when performing pointing and dragging tasks modeled after Fitts' reciprocal tapping task. In this study the participants had similar performance during the pointing task when using the tablet with stylus as with the mouse but they performed slightly better for the dragging task when using the tablet with stylus. When using the trackball the participants had lower performance than with the other two devices for both tasks. Another finding was that a dragging task requires more time and produces more errors than does a pointing task across all tested devices.

Douglas et.al. (1999) evaluated the scientific validity and practicality of the ISO 9241, Part 9 draft. This draft recommended an international standard for testing the usability of computer pointing devices. The performance a user can achieve with a device was evaluated by measuring the throughput for one-directional and multi-directional pointing and selecting tasks. Throughput is the ISO recommended primary measurement of performance. Throughput is the Fitts' index of performance using the effective width instead of the actual measured size of the target to incorporate the observed variability of human performance by including both speed and accuracy (Douglas et. al., 1999). In the course of this study the usability of two pointing devices for laptop computers, the isometric joystick (TouchStick) and the touchpad, were evaluated. The *Generalized Fitts' Law Model Builder* (Soukoreff and MacKenzie 1995) was used to present both one- and multi-directional pointing tasks to the participants. This study showed that the participants had better performance when using the isometric joystick compared to the touch pad.

While the performance of pointing devices in an office environment has been studied extensively, little is currently known about their performance in a mobile environment. Rider et. al. (2003) analyzed the effect of land-vehicle ride motion on the movement time and accuracy of in-vehicle reaching tasks. In this study "soft" buttons on a touch screen and "hard" push-buttons were used to present reaching and pointing tasks in a moving vehicle under different driving conditions, for different reach distances, different reach directions, and different target sizes. Movement time and spatial errors were recorded to evaluate performance. In addition, the participants were asked to complete a post-test questionnaire that found vehicle motion adversely affected movement completion time for reaches, accuracy in pointing tasks, and perceived difficulty ratings.

The goal of this study is to examine the usability of standard, commercially available pointing devices in a moving off-road environment (a tractor) to develop a general understanding of pointing devices in such an environment. The devices were evaluated to determine those with the best capabilities to perform office computer applications in a tractor

and to determine the performance a user can expect with them. Off-road environmental factors, such as vibration of work surfaces and the relative movement of the operator's body to the input device, can have a significant influence on the performance of those devices. It is therefore expected that the performance a user can achieve with the tested input devices will decrease in a mobile environment as compared to a stationary one.

2.0 METHOD

An experiment was designed to evaluate the performance a user can achieve with five different pointing devices in a moving tractor. Performance was tested using a multi-directional discrete pointing task (2D Fitts discrete task Douglas et al 1999). The pointing devices considered were off-the-shelf standard pointing devices that are commonly used with standard application interfaces. The devices included in this study were:

- Microsoft Optical IntelliMouse
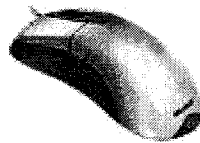
- Kensington Orbit Optical Track Ball

- Cirque Smart Cat USB Touchpad
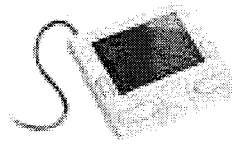
- DELL Latitude build in Touch Stick

- Earth Computer Tech MTR-EVUE-10T Touch Screen

Figure 1. Tested pointing devices

These devices were selected because they are either standard input devices used in an office environment (mouse, track ball, touch pad, touch stick) or because they are widely used in mobile applications such as Personal Digital Assistants (touch screen).

The goal of this preliminary study was to determine the pointing devices with potential to perform well in the given environment for standard office applications.

2.1 Participants

After the study was approved by the Iowa State Human Subject Review Committee fourteen male employees at the John Deere Product Engineering Center in Waterloo, Iowa, volunteered to participate in this study. The group included 13 right-handed subjects and one left-handed subject. All participants had experience in operating tractors and felt comfortable using a Microsoft Windows or similar computer environment. A pre-test questionnaire was administered to determine the extent of a subject's exposure to computer pointing devices. Twelve of the participants indicated the mouse as their preferred pointing device.

2.2 Apparatus

A Dell Latitude laptop computer with a 14" LCD screen was fitted in a John Deere 7920 Series tractor. See Figure 2 for the test setup.
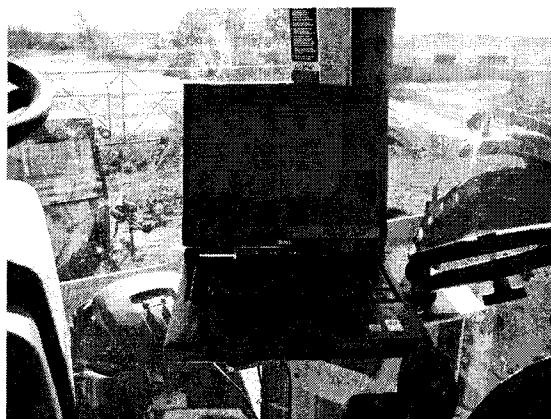
Figure 2. Laptop computer fitted in 7920 tractor cab

Additionally, a touch screen was installed in the cab. The touch screen was mounted on an adjustable arm that allowed all participants to adjust it to their preference (Figure 3).
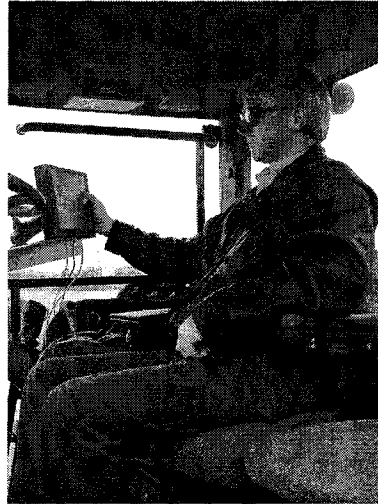
Figure 3. Touch screen in 7920 tractor cab

A custom armrest (40 cm x 20 cm) replacing the standard armrest was installed in the tractor to provide a surface for the use of the mouse, trackball and touch pad (Figures 4 and 5).
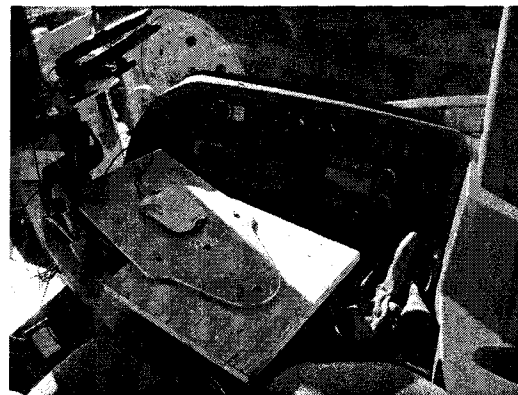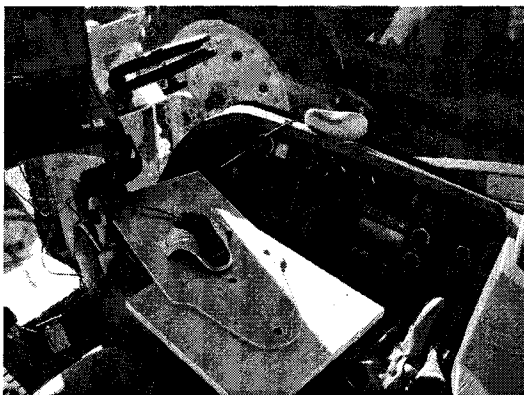
Figures 4 and 5. Custom armrest in 7920 tractor cab with trackball and mouse

As recommended in prior studies (Douglas et al 1999) gain was set to the middle value in the standard driver software for pointing device sensitivity for all devices.

The participants were asked to perform a series of discrete pointing tasks presented to them by two different test programs. For the mouse, touch pad, touch stick and trackball the *Generalized Fitts's Law Model Builder* program (Soukoreff, MacKenzie, 1995) was used. This program, written in C, runs under Microsoft Windows in MSDOS mode.

Due to compatibility issues between the *Fitts's Law Model Builder* and the touch screen it was necessary to use different software for this part of the test. Software was developed using the MetaCard platform, running under Microsoft Windows. Figure 6 shows a typical screen of this program.
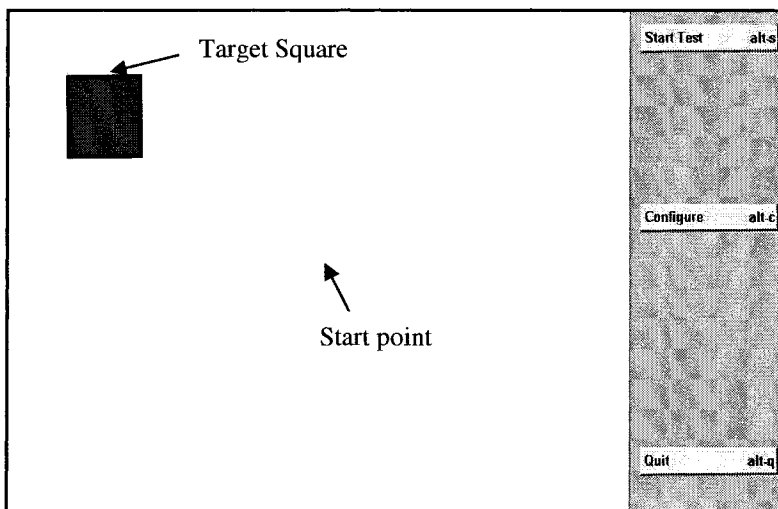
Figure 6. Screen shot of test software

Both programs are very similar and presented subjects with identical pointing tasks and recorded the same parameters: movement time (MT) and an error indicator (E). The dependent variable movement time (MT) was defined as the time from the start of the task until the target square was hit and error indicator (E) was defined as an attempted object selection that landed outside the target area.

2.3 Moving Vehicle Environment

The tractor used for this study was a John Deere 7920 Series tractor. To conduct the study the tractor was controlled by the investigator from the "Instructor Seat" while the participant performed the pointing tasks sitting in the driver seat. The tractor followed a predetermined path at the John Deere Product Engineering Center in Waterloo, Iowa. To ensure the test conditions were representative for a fieldwork environment the tractor speed was set to 7.5 km/h with acceleration data of the cab, driver seat and frame of the tractor taken while traveling on the path. It was determined by experts at John Deere that the test conditions were representative for fieldwork conditions.

2.4 Experimental Design

According to MacKenzie (2005) a within-subject design is generally preferred because effects due to the disposition of participants are minimized. Such effects are likely to be consistent for each condition under which a participant is tested. This is beneficial because the variability in measurements is more likely due to differences among conditions than to behavioral differences between participants if a between subjects design had been used. For this study a "within-subject" experimental design was used with the following independent variables:

- Vehicle state (stopped, moving)
- Target Size (20mm, 10mm)
- Target Distance (80mm, 160 mm)
- Angle (0°, 60°, 90°, 140°, 230°, 300°)
- Pointing device (mouse, trackball, touch stick, touch pad, touch screen)

2.5 Procedure

Participants in the study completed a pre-test questionnaire to determine their computer experience, experience with the tested devices and their preferred device. This was done to see if the preference of the participants in an office environment would influence performance in a moving tractor. The actual testing was then performed in two sessions. In the first session the mouse, trackball, touch stick and touch pad were tested. In the second session the touch screen was tested utilizing different test software.

For each session the participants had an opportunity to familiarize themselves with the tested devices and the test software. The participants were placed in the driver seat of the tractor for this test and were instructed to be as fast and accurate as possible. The subjects were then asked to perform a series of 48 tasks (2 target sizes x 6 angles x 2 distances x 2 repetitions per condition) for each device with the vehicle stopped and the engine off. Each task began with the mouse pointer at the start point. The subject then moved the pointer to the target square and selected it. The target square then disappeared and the time between these instances was recorded as MT. The mouse pointer was then automatically returned to the start point and after a delay of 500 milliseconds (ms) the next task started. If the subject missed the target square a beep signal was given and the software recorded an error for that trial. However, the time for this task kept running and the subject still had to complete the task. The total time for the subject to correctly select the target was recorded as MT. Combinations of distance, angle and target size were presented following a predetermined configuration file. Two different target sizes, two different distances and six different angles varied the tasks. Each set was presented twice to each subject, for a total of 48 tasks per block.

After a short break the participants were asked to perform the same series of tasks with each device while the tractor was traveling along the predetermined test track at a constant speed of 7.5 km/h. The participants completed the pointing task as an investigator sat in the instructor seat of the tractor and handled all of the tractor controls.

The time spent by each participant ranged from 30-45 minutes for the first session and from 10-20 minutes for the second session.

3.0 ANALYSIS

The test software directly collected the data for movement time MT and errors for single tasks. The data was then prepared for further statistical analysis by computing values for throughput and error ratio E.

3.1 Adjustment to Data

All movement time values were truncated to a maximum of 10 seconds, that is, all MT values over 10 seconds were scored as 10 seconds. Since the means for the movement time was 2.2 seconds and the standard deviation was 1.8 seconds, this only affected values that were more then four times the standard deviation from the mean. This didn't change the overall results of the data but eliminated some outliers from the statistical analysis. The movement time, index of performance and throughput values were normalized using a natural logarithmic transformation.

3.2 Computed Formulas

To compare the different devices the Index of Performance (IP) was calculated. The IP values were then calculated using the formula:

$$IP = ID / MT \qquad\qquad (1)$$

where MT is the movement time in seconds and ID is the index of difficulty in bits. The Shannon formulation was used to calculate ID because it always gives a positive rating for the index of difficulty and yields a better fit with empirical data (MacKenzie et al., 1991).

$$ID = \log_2 (D / W'+1) \tag{2}$$

where D is the distance to the target and W is the width of the target. Since this study presented two-dimensional tasks, the W model was used; this model substitutes the span of the target as viewed from an approach vector for the target width (MacKenzie 1991).

The primary performance measurement for pointing devices recommended by ISO is the Throughput (TP) (Douglas, 1999).

$$TP = ID_e / MT \tag{3}$$

where MT is the movement time in seconds and ID_e is the effective index of difficulty.

$$ID_e = \log_2 (D / W_e+1) \tag{4}$$

ID_e is the effective index of difficulty, in bits, and is calculated from D, the distance to the target and W_e, the effective width of the target. W_e was computed from the observed distribution of selection coordinates in participants' trials:

$$W_e = 4.133 \ SD \tag{5}$$

where SD is the standard deviation of the selection coordinates (Douglas, 1999).

$$SD = \sqrt{\left(\frac{\sum_{i=1}^{n}[(x_i - \bar{x})^2 + (y_i - \bar{y})^2}{n-1} \right)} \tag{6}$$

For this study both performance measures, TP and IP, were calculated and compared. A Pearson Product-Moment Correlation study between the TP and IP values indicted a strong correlation (.963). In general, it appears the TP values are slightly offset to the lower side as compared to the IP values but have the same order across devices. Therefore, only the ISO recommended measurement TP was used in the further progress of this study.

4.0 RESULTS

4.1 Mean movement time MT

A general linear model analysis found a significant difference in MT between devices (F=405.1,p<0.0001), between tractor states (moving/stationary) (F=618.3 p<0.0001), between different target widths (F=322.41,p<0.0001), between different target distances (F=113.1,p<0.0001), the interaction between tractor state and pointing device (F=27.57,p<0.0001), the interaction of target distance and the device (F=8.42,p<0.0001) and the interaction between target width and tractor state (F=56.28, p<0.0001).

Across all devices the participants had the best MT performance when using the touch screen with a least square MT of 995 ms followed by the mouse with 1474 ms, the track ball with 2342 ms, the touch pad with 2476 ms and the touch stick with 2612 ms. The detailed results are shown in Figure 7.

Current effect: F(4, 6660)=373.00, p=0.0000
Vertical bars denote 0.95 confidence intervals
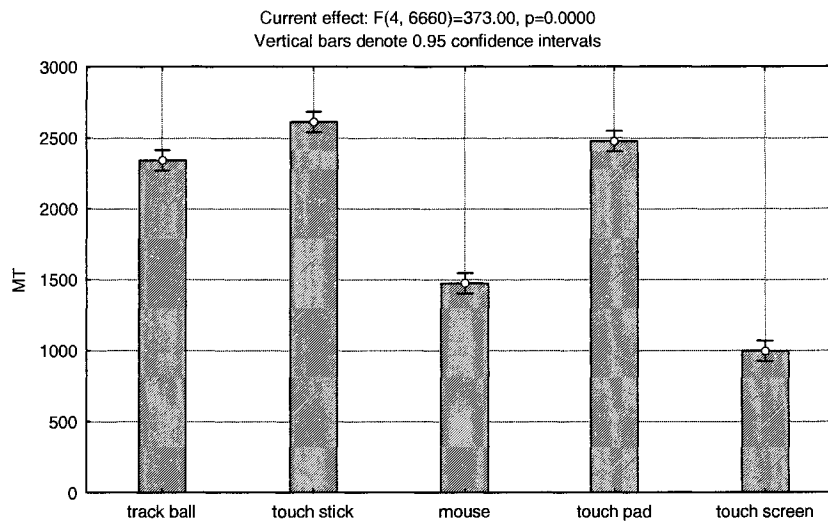
Figure 7. MT by device

A significantly higher MT with a least square mean value of 2371 ms was found in the moving tractor compared to the stationary tractor with a least square mean MT of 1589 ms. There was also a significantly higher least square mean MT found for the small or 10mm wide targets (2262 ms) compared to the big or 20 mm wide targets (1698 ms) and for the large target distance (2147 ms) compared to the small target distance (1813 ms).

For the interaction between typing device and tractor state it is shown that the participants had a significantly higher MT in the moving tractor compared to the stationary with all devices. The best performance was found when using the touch screen with a least square mean MT of 808 ms stationary and 1182 ms for the moving tractor followed by the mouse with 1258 ms stationary and 1690 ms moving. With the track ball a least mean square MT of 1724 ms stationary and 2860 ms moving, with the touch pad 1989 ms stationary and 2962 ms moving and with the touch stick 2166 ms stationary and 3059 ms in the moving tractor was found. For the track ball, the touch stick and the touch pad there was no significant difference found between devices in the moving tractor. With those devices the participants also had the highest MT in both the moving and the stationary tractor and also the biggest difference in MT between the different tractor states. The detailed results are shown in Figure 8.

Current effect: F(4, 6660)=25.382, p=0.0000
Vertical bars denote 0.95 confidence intervals
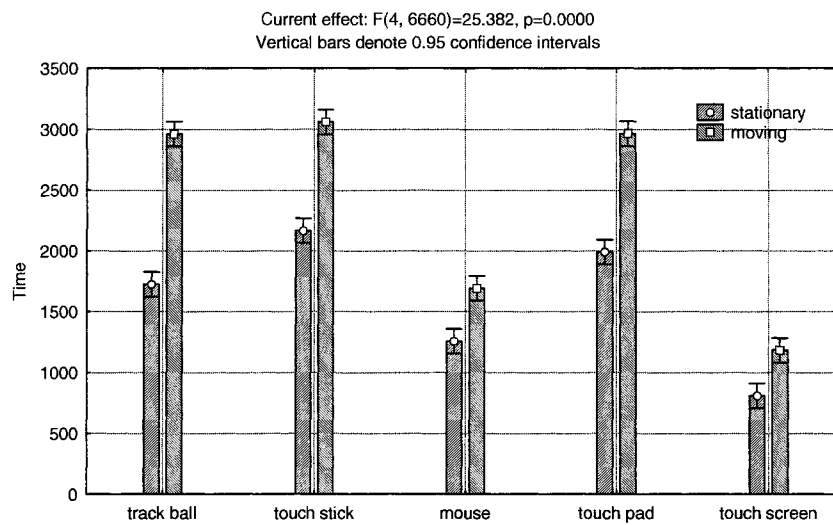
Figure 8. MT by tractor state and device

For the interaction between target distance and pointing device for all devices except the touch screen a higher least square mean MT for the large distance compared to the small distance was found. For the touch screen there was no significant difference in MT found between target distances (Figure 9).

Current effect: F(4, 6697)=8.4151, p=.00000
Vertical bars denote 0.95 confidence intervals
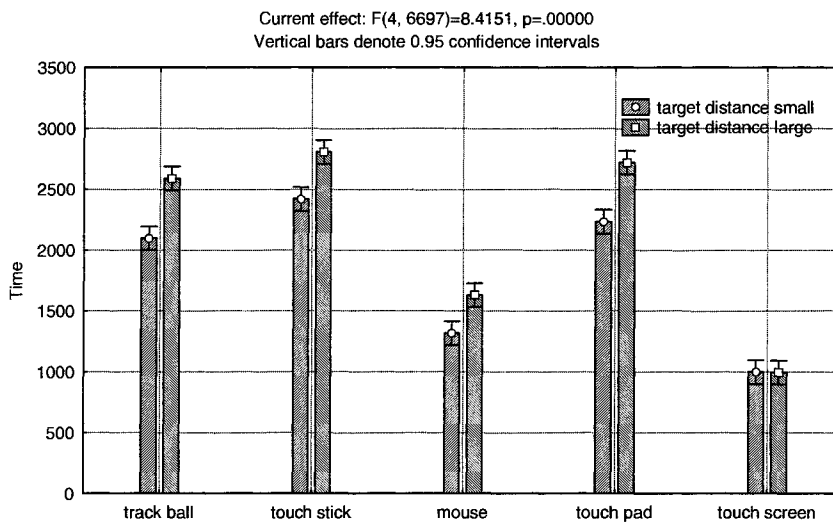
Figure 9. MT by target distance and device

For the interaction of target width and tractor state there was a significantly higher difference in MT between target sizes found in the moving tractor compared to the stationary tractor (Figure 10).

Current effect: F(1, 6697)=56.284, p=.00000
Vertical bars denote 0.95 confidence intervals
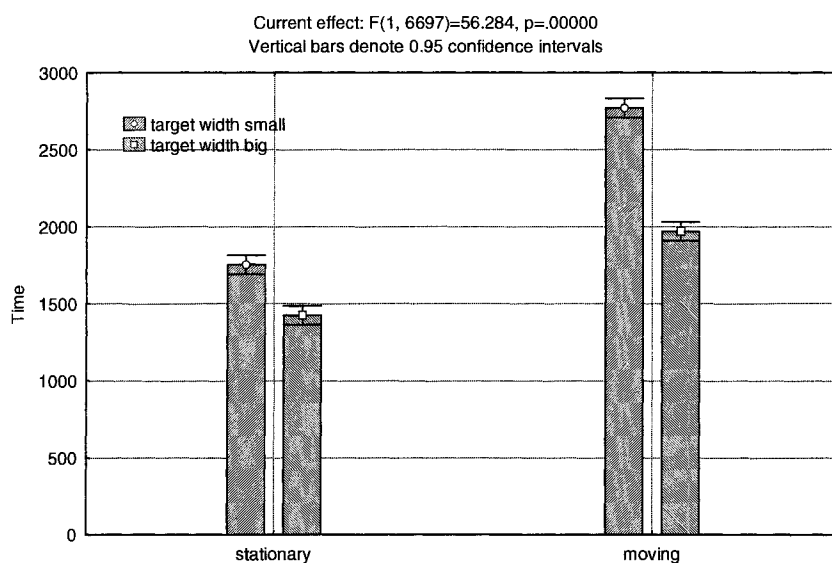
Figure 10. MT by target width and tractor state

4.2 Throughput TP

For the TP values there was a significant difference found for the performance with different devices (F=1966, p<0.0001), between tractor states (F=754, p<0.0001) and for the interaction between device and tractor state (F=22.69, p<0.0001).

For the different tractor states a least square means TP of 2.28 bits/sec was found in the stationary tractor compared to 1.79 bits/sec for the moving tractor. Between the different pointing devices when using the touch screen the participants had the highest TP values with a least square mean of 3.51 bits/sec followed by the mouse with 2.27 bits/sec, the track ball with 1.56 bits/sec, the touch pad with 1.46 bits/sec and the touch stick with 1.38 bits/sec (Figure 11).
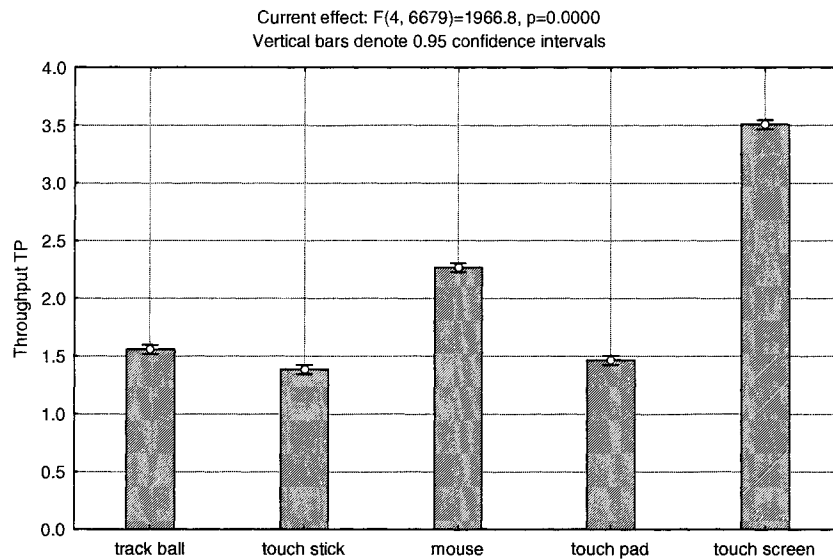
Figure 11. TP by pointing device

For all devices a higher TP value was found in the stationary tractor compared to the moving tractor. The highest TP values were found when using the touch screen with a least square mean TP of 3.19 bit/sec in the stationary and 3.10 bit/sec in the moving tractor followed by the mouse with 2.50 bit/sec stationary and 2.03 bit/sec moving, the track ball 1.81 bit/sec stationary and 1.30 bit/s moving, the touch pad with 1.63 bit/sec. stationary and 1.30 bit/sec moving and the touch stick with 1.56 bit/sec stationary and 1.20 bit/sec moving. For both the stationary and the moving environment the order of performance between devices used was identical but there was no significant difference in TP found in the moving tractor between the track ball, the touch stick and the touch pad (Figure 12).
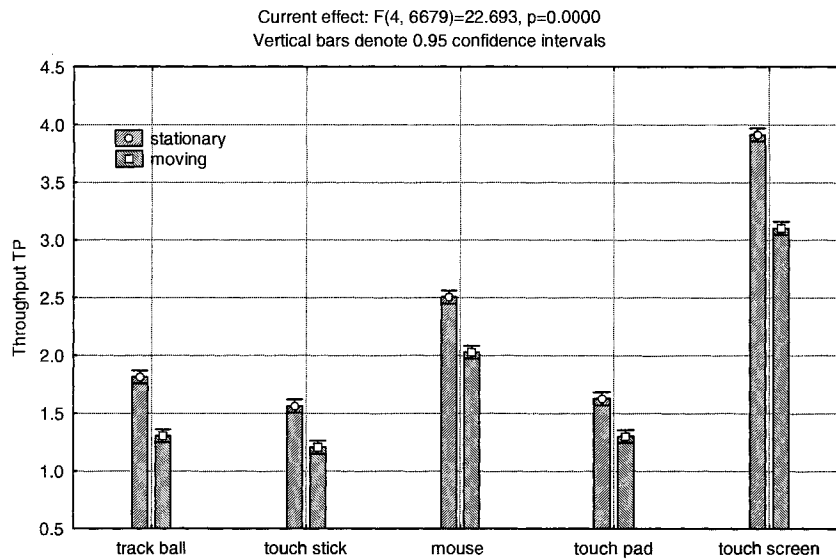
Figure 12. TP by tractor state and device

4.3 Error ratio E

For the error ratio there was a significant difference in the performance of the participants found between devices used (F= 42.43, p<0.00010), between tractor states (F=222.6,p<0.0001), between target width (F= 97.9,p<0.0001), the interaction of target width and device used (F=25.2,p<0.0001), target width and tractor state (F=36.9, p<0.0001) and device used and tractor state (F=14.9, p<0.0001).

Between the different devices the lowest error ratio was found when using the mouse with a least square mean error ratio of 3.8% followed by the touch pad with 3.9%, the touch stick with 5.9%, the track ball with 6.8%, and the touch screen with 14.4 % (Figure 13).
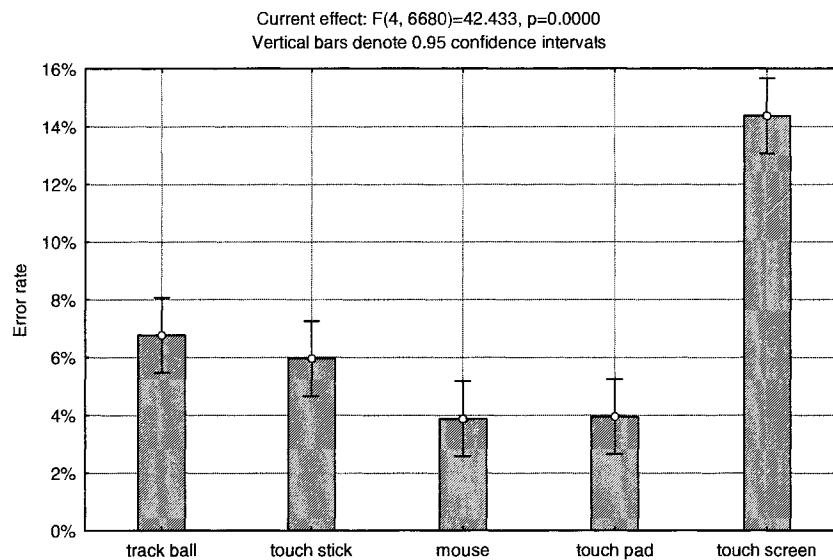
Figure 13. Error rate E by pointing device

There was a significantly higher error rate found in the moving tractor compared to the stationary tractor. The least square mean error rate for the stationary tractor was 2.5% and 11.4% for the moving tractor. Between the different target widths the error ratio for the 10mm targets was significantly higher with a least square mean of 9.9% compared to 4.0% for the 20mm targets.

For the interaction of target width and pointing device the participants had a higher error ratio with the small (10mm) targets compared to the large (20mm) targets for all devices. The biggest difference between the target sizes was found when using the touch screen which had a least square mean error ratio of 5.5% for the large and 23.2% for the small targets (Figure 14).
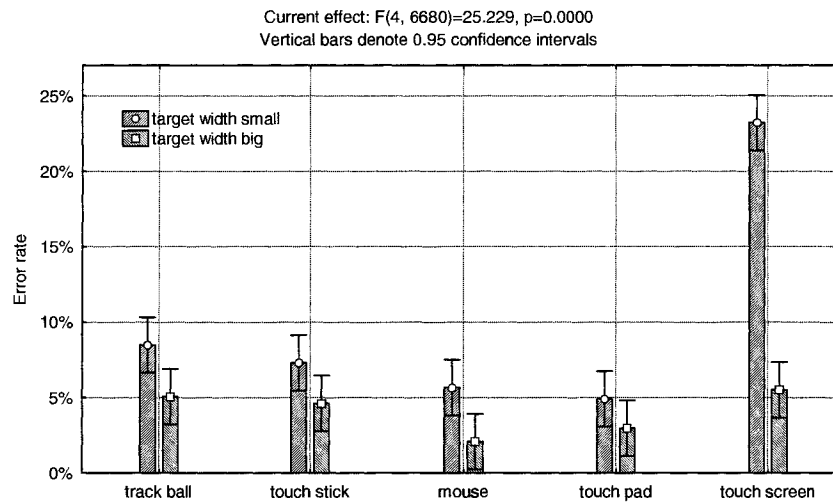
Figure 14. Error rate E by target width and device

For the interaction of tractor state and pointing device with all devices the participants had a significantly higher error ration in the moving tractor compared to the stationary tractor. Here again the biggest difference in error ratio between tractor states was found when using the touch screen which had a least square mean error ratio of 5.9% for the stationary and 20.9% for the moving tractor (Figure 15).
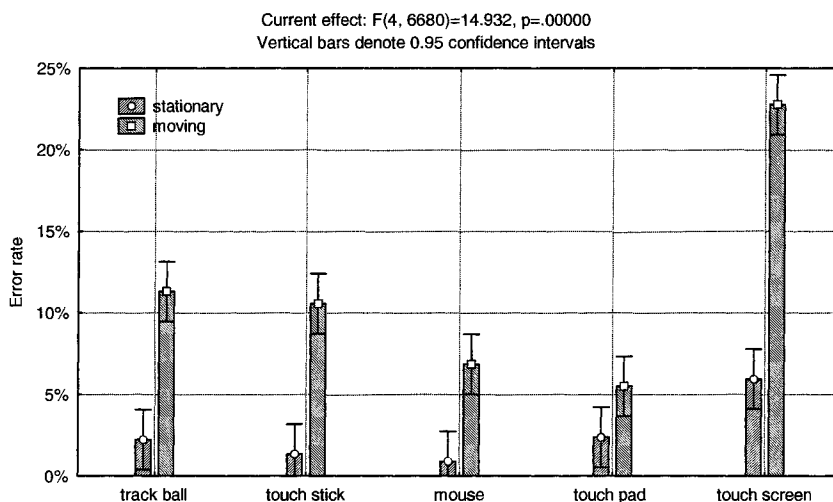
Figure 15. Error rate E by device and tractor state

5.0 DISCUSION

Consistent with the Rider et al. (2003) study, this study showed that participants had significantly worse performance in MT, TP and error ratio in the moving tractor compared to the stationary tractor for all tested devices. Target size was also a significant influence on the performance found for all devices. This was more substantial in the moving tractor than in the stationary.

As MacKenzie (1991) showed, different values are found for performance parameters when using the same device in different studies. MacKenzie suggested looking at across-study agreements on within-study ranking rather than the absolute values to cope with the disparity between studies. This means that rather than comparing the absolute values of performance measurement of different studies one should look at the ranking of the devices in the study.

For the tested devices in this study the best performance for MT and TP was found when using the touch screen followed by the mouse. This is consistent with the results of MacKenzie et. al. (1991) where the tablet with stylus, comparable to a touch screen, allowed slightly better performance than the mouse.

When using the track ball, the touch stick and the touch pad participants had lower performance values with only small differences between the three devices. For the track ball, the touch stick and the touch pad the biggest influence on the performance parameters MT and TP was caused by motion and target size.

For the performance parameter error ratio the best performance was found when using the mouse and the touch pad with almost no difference between these two devices. The worst performance was found when using the touch screen, which caused more than double the error ratio than the next best device.

Consistent with the results found by Rider et. al. (2003) the target size and the tractor state had a major influence on the participants' error ratio for all devices. The influence of these two factors was most significant when using the touch screen.

For the tested conditions, the touch screen allowed the best MT and TP performance. The drawback of the touch screen is the high error ratio, especially for small targets.

Conversely, with the touch pad the participants had the best error ratio but a bad performance in mean movement time and TP. With the mouse the participants had the second best performance for both MT and TP and the second best error rate. With the trackball and the touch stick the participants performed more poorly for MT and TP and had the second highest error ratios.

6.0 CONCLUSION

The goal of this study was to determine which of the tested input devices have the potential to enable the user to utilize office computer applications in a mobile off-road environment and provide the best usability to the users of such a system. For the tested devices the mouse, touch screen, and touch pad showed good usability in some aspects of the study but were not satisfying in others. The track ball and the touch stick showed low usability under the tested conditions for all aspects of this study. Without significant changes in their structure and function, these two devices are not well-suited for the tested environment.

To find the device that is best suited for this environment, further investigations are necessary to determine which aspects are most important for the use of office computer applications in a mobile environment. It is necessary to consider not only the speed and error ratio but also the effect an error could have to determine which devices are best suited for this kind of application in a mobile environment.

Future studies need to take into consideration that there is a big difference in the usability of the tested devices between different target sizes. This could cause problems for the use of standard office computer applications since these applications frequently use pull-down menus and small tool icons. Future studies that further investigate the usability of the tested devices should therefore investigate the performance a user can achieve with a real application rather than an abstract pointing task to give more insight in the real world usability of the tested devices.

REFERENCES

Afifi, A., Clark, V.A., & May, S., (2004) Computer-aided multivariate analysis, Chapman & Hall/CRC, Boca Raton.

Douglas, S.A., Kirkpatrick, A.E. and MacKenzie I.S., (1999) Testing pointing device performance and user assessment with the ISO 9241, Part 9 Standard. Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI'99 pp.215-222 ACM: New York.

Fitts, P. M., (1954) The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 47 (1954), 381-391

Hochberg, Y., Tamhane, A.C., (1987) Multiple comparison procedures. Wiley, New York

ISO 9241-9, (1998) Ergonomic requirements for office work with visual display terminals, Non-keyboard input device requirements

ISO 9241-11, (1998) Ergonomic requirements for office work with visual display terminals, Guidance on usability

Jeffries, R., Miller, J.R., Wharton, C., Uyeda, K.M., (1991) User interface evaluation in the real world: a comparison of four techniques. Association for Computing Machinery, 1991

MacKenzie, I. S., (1991) Fitts' law as a performance model in human-computer interaction. Doctoral dissertation. University of Toronto: Toronto, Ontario, Canada.

MacKenzie, I. S., Sellen, A., & Buxton, W., (1991) A comparison of input devices in elemental pointing and dragging tasks. Proceedings of the CHI `91 Conference on Human Factors in Computing Systems, pp. 161-166. New York: ACM.

MacKenzie, I.S., & Buxton, W., (1993) A Tool for rapid evaluation of input devices using Fitts' law models. SIGCHI Bulletin, 25 (3), 58-63.

MacKenzie, I.S., (2005) Within-subject vs. between-subjects design :which to use?. York University: Toronto, Ontario, Canada.

Rider, K.A., Park, W., & Prohaska, M.P., (2003) Analyzing the effect of ride motion on the performance in-vehicle reaching tasks. Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society, Denver, Colorado, 2003.

Soukoreff, W., & MacKenzie, I. S., (1995). Generalized Fitts' law model builder. Companion Proceedings of the ACM Conference on Human Factors in Computing Systems – CHI '95, pp. 113-114. New York: ACM

Usability of pointing devices for office applications in a moving off-road environment

Thorsten Baldus

ABSTRACT

Three pointing devices (mouse, touch pad, touch screen) were evaluated to determine the one that provides the best usability with a standard windows application in a moving off-road environment. Eighteen subjects performed a series of complex pointing tasks that simulated the use of a standard application in a moving tractor and rated the tested devices for their subjective usability. The mouse and the touch screen both allowed the best objective performance with the mouse being the device that received the best subjective usability ratings. With the touch pad the participants in this study had the lowest performance and it was also the least favorite device amongst the participants.

1.0 INTRODUCTION

Advances in modern farming technology provide a lot of automation features that lighten the workload for operators of farming equipment. One of the many technologies incorporated in modern agricultural equipment is automated steering systems like the John Deere's Green-Star Auto-Trac System™, which is a hands-free steering system for straight-line driving applications.

Future generations of these systems are expected to include path planning software and implement control functionality. This means once a field work task is programmed and started all driving operations, including turning when the end of the field is reached, and all control functions will be performed by the system.

With such systems, the computer is taking over more and more of the operator's workload and the operator's task will switch from steering the tractor and controlling the implement functions to monitoring tasks to ensure everything works as intended.

When using a machine that is equipped with one of these systems, the work load of the operator would be lightened and enable the operator to perform other tasks while doing field work. For this reason, many people believe a tractor cab could become a mobile office in the near future with the operator using a personal computer to take care of various office tasks while working in the field. With the advances in mobile technology, it is easy to imagine that in the near future a farmer could browse the Internet to find weather data or market prices or sell a harvest while sitting in the tractor doing field work. To enable the operator to do this efficiently it is necessary to provide a usable computer interface for the environment that includes a pointing device and an alphanumeric input device.

Previous work in this area primarily evaluated the performance and usability of interface devices in either an office environment or for mobile devices like laptops or PDAs, and focused on either pointing devices or alphanumeric input.

Fitts' law for the information processing capacity of the human motor system (Fitts, 1954) was developed to predict performance of the human motor system. MacKenzie (1991) examined the theory, prediction power, and relevance of Fitts' law as a performance model in human computer interaction. He extended the law's applicability to movement tasks that are common on interactive computer systems like dragging and compound dragging-plus-pointing tasks. He developed two methods to accommodate the approach angle in the calculation of task difficulty and demonstrated that movement time, error rate, and the index of performance are all affected by the choice of device or task.

McKenzie et.al. (1991) compared the usability of a mouse, a tablet with stylus, and a trackball when performing pointing and dragging tasks both modeled after Fitts' reciprocal tapping task. In this study the participants had similar performance during the pointing task

when using the tablet with stylus and the mouse, but they performed slightly better for the dragging task when using the tablet with stylus. When using the trackball the participants had a lower performance than with the other two devices for both tasks. Another finding was that a dragging task requires more time and produces more errors than a pointing task across all tested devices.

Bohan et.al. (2003) compared objective performance and subjective user preference of a RollerMouse to a standard mouse. The participants in this study were asked to perform a series of tasks representative of the actions typically performed when using pointing devices. In this study the standard mouse had better performance in pointing and dragging for all target sizes and distances. The participants also rated the standard mouse as significantly easier to use than the RollerMouse in their subjective evaluation.

While the mouse is the most popular pointing device for an office environment for mobile devices like laptops or PDAs, other devices are commonly used.

Douglas et.al. (1999) evaluated the scientific validity and practicality of the ISO 9241, Part 9 draft. This draft recommended an international standard for testing the usability of computer pointing devices. The performance a user can achieve with the tested devices was evaluated by measuring the throughput for one-directional and multi-directional pointing and selecting tasks. Throughput is the ISO recommended primary measurement of performance. It is the Fitts' index of performance using the effective width instead of the actual measured size of the target to incorporate the variability observed of human performance by including both speed and accuracy (Douglas et. al., 1999). In the course of this study Douglas evaluated the usability of two pointing devices for laptop computers, the isometric joystick (TouchStick) and the touchpad. The *Generalized Fitts' Law Model Builder* (Soukoreff and MacKenzie 1995) was used to present both one- and multi-directional pointing tasks to participants. This study showed the participants had a better performance when using the isometric joystick compared to the touch pad. Based on the results of his study, Douglas (Douglas et. al. 1999)

recommended that while throughput allows a single measure of both speed and accuracy it is best to compute both movement time and error rate as separate dependent variables.

ISO 9241, part 11 (1998) provides a guide for the evaluation of the usability of an office computer terminal. It states that in order to determine the level of usability achieved, it is necessary to measure the performance and the satisfaction of users working with a product. Measurement of usability is particularly important in view of the complexity of the interactions between the user, the task characteristics, and the other elements of the context of use. A product can have significantly different levels of usability when used in different contexts. The standard also states that measures of the performance and satisfaction of the user can provide a basis for the comparison of the relative usability of products with different technical characteristics which are used in the same context (ISO 9241, part 11, 1998).

While no prior work was found on usability testing of pointing devices in moving vehicles, Rider et. al. (2003) analyzed the effect of land-vehicle ride motion on the movement time and accuracy of in-vehicle reaching tasks. In this study "soft" buttons on a touch screen and "hard" push-buttons were used to present reaching and pointing tasks in a moving vehicle under different driving conditions, for different reach distances, different reach directions, and different target sizes. To evaluate performance, the movement time and spatial error were recorded. The participants were also asked to complete a post-test questionnaire. It was found that movement completion time for reaches, accuracy in pointing tasks, and perceived difficulty ratings were adversely affected by vehicle motion.

The goal of this study was to identify the commercially available pointing device that provides the best usability when using a standard application in a moving off-road environment and identify the factors that influence the usability of the tested devices. To do this the performance a user can achieve with the tested devices in the tested environment as well as the subjective evaluation of the devices will be considered. In an off-road application it is expected that factors like the vibration of the work surface and the relative movement of the work environment have a significant influence on the usability of the tested devices.

While for an office environment the mouse is the most widely accepted pointing device to be used with standard applications it might not be the one offering the best usability for these applications in the tested environment.

In a pilot study five standard off the shelf pointing devices were tested to evaluate their usability in a mobile off-road environment. In the pilot study the test participants had to complete a series of discrete pointing tasks modeled after Fitts' reciprocal tapping task, where task time, throughput and error ratio were used to evaluate the performance the participants had with the tested devices. Three of the tested devices, the mouse, the touch screen and the touch pad, allowed good performance in some aspects of the pilot study and seem to be suited to be used in the tested environment. These three devices were chosen to be evaluated more thoroughly in the course of this study. Rather than using an abstract pointing task this study is trying to simulate the use of the tested devices with a standard application as close as possible. This will not only provide a better representation of the performance a user can expect with the tested devices but it will also provide a better idea of the real world usability for the subjective user evaluation.

2.0 METHOD

An experiment was designed to evaluate user performance with three devices used in a series of pointing tasks. The devices chosen for this study were:

- The Touch Pad (Cirque Smart Cat)
- The touch screen (Intelliworx Voice table)
- The mouse (Microsoft Optical IntelliMouse)

These devices were chosen as they demonstrated the most promising results in a pilot study and performed well in some aspects of it. In the pilot study, with the touch pad the users had

the lowest error ratio, with the touch screen the users had the highest throughput and with the mouse all participants had acceptable performance in all aspects of the study.

2.1 Participants

Eighteen employees at John Deere PEC in Waterloo, Iowa between the ages of 21 and 58 (mean 37.8) volunteered to participate in this study. The participants were categorized by age. Out of the 18 participants eight were age category one (<30 years), three category two (30-45 years) and seven category three (>45 years). The group included 17 right-handed subjects and 1 left-handed subject; 16 were male, and two were female. A questionnaire was administered to determine the extent of a subject's exposure too Microsoft Windows like computer environment, and tractor operation. Seventeen of the participants stated they were very experienced in using a Microsoft Windows like environment, with one participant stating medium experience. Concerning tractor operations eleven participants were very experienced, with two having medium experience, and five stating to only have little experience in operating a tractor. Concerning the experience with the different pointing devices 17 participants stated to be very experienced using a mouse with one participant having medium experience. For the touch screen three participants stated to be very experienced with 9 having medium experience and 6 having little prior experience using a touch screen. For the touch pad 10 participants were very experienced, five had medium experience and three had little experience using a touch pad (Table1).

Age Category	Mouse Exp.			Touch Pad Exp.			Touch screen exp.		
	Low	Med	High	Low	Med	High	Low	Med	High
1	-	-	8	1	5	2	1	1	6
2	-	1	2	2	1	-	-	2	1
3	-	-	7	3	3	1	2	2	3

Table 1. Participants summary

2.2 Apparatus

An Intelliworxx Voice tablet mobile computer with integrated touch screen was installed in a John Deere 7810 series tractor. The touch screen was mounted on a movable arm and all participants were able to adjust its position to their preference. Additionally there was a custom "work surface" (20 cm by 15 cm) installed on the right arm rest to enable the use of mouse and touch pad (Figure 1).
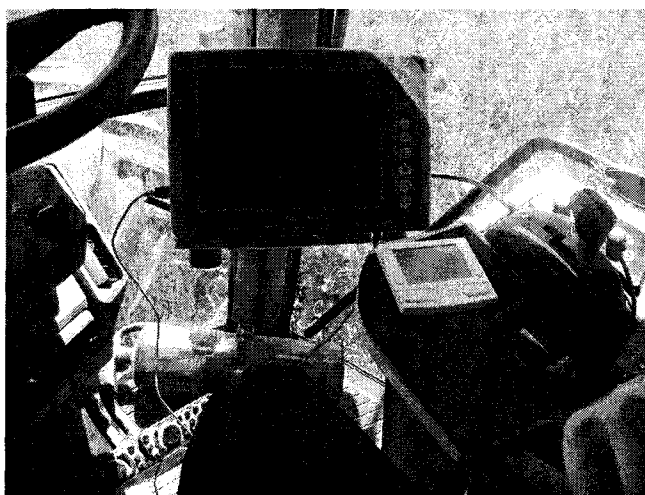
Figure 1. Computer fitted and work surface in 7910 tractor cab

As in the pilot study and in other studies (Douglas et al 1999) the gain was set to the middle value in the standard driver software for pointing device sensitivity for all devices. The participants were asked to perform a series of complex pointing tasks mimicking pointing tasks in a standard application presented to them by test software. The test software was a self developed META-card application that ran in a Microsoft Windows environment. The software presented the pointing tasks to the participants and recorded the task parameters, the task time (TT, time to perform the task correctly), the X, Y position on the screen of different selections, and the number of corrections and mouse clicks the participants performed.

2.3 Moving Vehicle Environment

The tractor used for this study was a John Deere 7910 Series tractor. The test was conducted over the same track and under similar conditions as the pilot study. The tractor was traveling on a predetermined path at a constant speed of 7.5 km/h with the investigator handling all tractor controls from the instructor seat while the participants performed the pointing tasks. During the pilot study acceleration data of the tractor frame and cab was recorded and it was decided by experts at John Deere that the conditions are representative for field work conditions. It was determined that it was not necessary to record any acceleration data again since the test conditions were very similar to the pilot study.

2.4 Experimental Design

For this study a "within-subject" experimental design was used with the following independent variables:

- Vehicle state (stopped, moving)
- Pointing device (mouse, touch pad, touch screen)
- Target size (small, large)
- Shape (circle, triangle, square, hex)
- Color (blue, red, black, green)

According to MacKenzie (2005) a within subject design is generally preferred because effects due to the disposition of participants are minimized. Such effects are likely to be consistent for each condition under which a participant is tested. This is beneficial because the variability in measurements is more likely due to differences among conditions than to behavioral differences between participants if a between subjects design was used.

2.5 Procedure

Each subject had a chance to familiarize themselves with the tested devices and the test software by performing a short series of the presented tasks with every device. The subjects were placed in the driver seat for the test and were instructed to complete the tasks presented as fast and as accurate as possible.

When the test was stared a target shape of a certain color was presented to the subject by the test software. The subject than had to select the right shape from a menu bar on top of the screen by moving the pointer to it and click on it. This opened a pull down menu with a selection of different colors and the subject had to select the correct color from this menu utilizing the pointer again. When this part of the task was completed a confirmation window opened and showed the subject the shape and color selected. The subject than had to either confirm their selection by moving the pointer and selecting the 'OK' button or selecting the 'CANCEL' button and correct their selection. A typical screen shot of the test software can be seen in Figure 2.
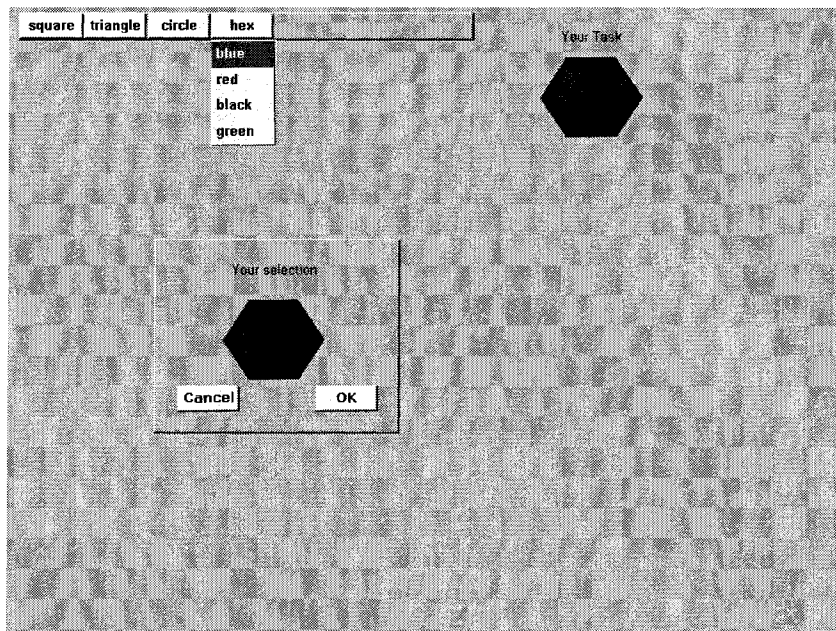
Figure 2. Screen shot of test software

A task was considered complete when the shape and color were selected correctly and the selection was confirmed. After a successfully completed task, the next task was presented to the participant after an 800 ms delay.

In case a participant hit the 'OK' button without having selected the right shape and color an error message appeared and asked the participant to cancel his selection and try again.

For each task the test software recorded task time (total time it took to correctly complete the task), number of selections (number of times the participant had to correct their selections), X and Y coordinates of every selection, time for every selection and for confirming the selection, total number of pointer clicks, and number of wrong confirmations of selections.

All subjects performed a set of 32 tasks at a time. Each set of 32 included 16 different tasks with small targets and 16 tasks with large targets in a random order. The different target sizes were 40 by 20 pixels (13mm x 6.5mm) for the small targets and 60 by 25 pixels (19.5mm x 8.125 mm) for the large targets. The target sizes were chosen according to the size of standard buttons in Microsoft Windows applications in different configurations. After each set, the subject had the chance to take a short brake before the next set was started. Overall, each subject performed a total of 12 sets in which every input device (3 devices total) and environment (moving/stationary) combination was tested twice. To guard against any effect of order a balanced design was used to present every subject with the devices in a different order. The test order is shown in Appendix B.

After all tests were completed, the participants were asked to complete a post test questionnaire where in addition to their experience with the different devices they were asked to give their subjective impression on how well the devices were suited for the tested environment. The time spent by every participant ranged from 40 to 70 minutes for the whole test.

3.0 ANALYSIS

The main goal of this study was to determine which of the tested devices is best suited to be used in a tractor to operate a standard windows application. To evaluate the usability of the tested devices objective performance measurements and the subjective user impressions had to be evaluated. For the objective performance measurements the test software directly collected the data for Task Time TT, number of mouse clicks and number of corrections for the task. The primary performance measurement for pointing devices recommended by the ISO 9241, Part 9 standard is Throughput (TP) (Douglas, 1999) which was calculated according to formula (1).

$$TP = ID_e / MT \qquad\qquad (1)$$

where MT is the movement time in seconds and ID_e is the effective index of difficulty.

$$ID_e = \log_2 (D / W_e + 1) \qquad\qquad (2)$$

ID_e is the effective index of difficulty, in bits, and is calculated from D, the distance to the target and W_e, the effective width of the target. W_e was computed from the observed distribution of selection coordinates in participants' trials:

$$W_e = 4.133 \, SD \qquad\qquad (3)$$

where SD is the standard deviation of the selection coordinates (MacKenzie 1991).

$$SD = \sqrt{\left(\frac{\sum_{i=1}^{n} [(x_i - \bar{x})^2 + (y_i - \bar{y})^2}{n-1} \right)} \qquad\qquad (4)$$

Since the task presented to the participants in this study is a multiple step complex selection task the TP values were calculated during one part of the tasks that closely resembles a multi directional discrete pointing task (2D Fitts discrete task Douglas et al 1999). For this the test software recorded the X,Y position of every selection and the time between selection and confirmation.

The second performance measurement the task time represents the performance for the whole task. It was necessary to normalize the data for the statistical analysis. This was done by inverting the task time and transforming it into tasks/second which was used in the further analysis of the data.

4.0 RESULTS

4.1 Throughput TP

A multivariate GLM found a significant difference in TP between pointing devices (F=2074, p<0.0001), between tractor states (F=1332, p<0.0001), between age categories (F=418, p<0.0001), for the interaction of pointing device and target size (F=4.24, p=0.014), the interaction of pointing device and tractor state (F=46.23, p<0.0001) and the interaction of pointing device and age category (F=15.96, p<0.0001). There was no significant difference in TP values found between the different shapes, colors or menu number or for any other interaction between factors.

Between the different pointing devices the highest TP values were found when using the touch screen with a least square mean of 3.42 bits/sec followed by the mouse with 3.14 bit/sec and the touch pad with 1.83 bits/sec.

As MacKenzie (1991) found, different values are found for performance parameters for the same device in different studies. MacKenzie suggested looking at across study agreements within study ranking rather than the absolute values to cope with the disparity between studies. This means that rather than comparing the absolute values of performance

measurement of different studies one should look at the ranking of the devices in the study. In this study the order of performance between the tested devices was identical to the pilot study with the touch screen having the highest TP values followed by the mouse and the touch pad.

Between the different tractor states there was a higher least square mean TP value found in the stationary tractor with 3.19 bits/sec compared to the moving tractor with 2.40 bits/sec.

For the different age categories there was no significant difference found between age category 1 with a least square mean TP value of 2.99 bits/sec and age category 2 with 3.01 bits/sec. Age category 3 had a significantly lower TP than the other two with a least square mean of 2.38 bits/sec.

With the mouse and the touch screen higher TP values were found for the big targets compared to the small targets. There was no significant difference among target sizes found when using the touch pad. The differences in TP between the device were used was more dominant than were the difference between target sizes though (Figure 3).

Current effect: F(2, 6840)=4.2406, p=.01444
Vertical bars denote 0.95 confidence intervals
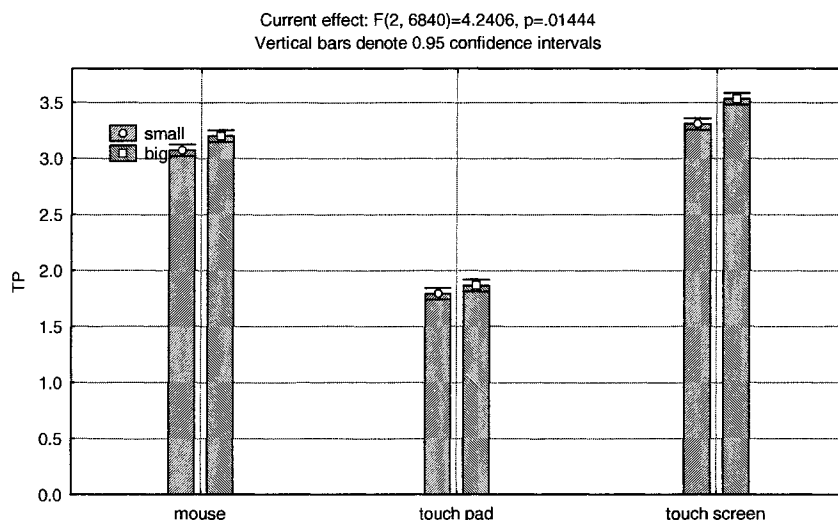
Figure 3. TP by pointing device and target size

For all devices the TP values in the moving tractor were significantly lower than in the stationary. For both the moving and the stationary environment the participants had the highest least square mean TP values with 3.77 bits/sec stationary and 3.07 bits/sec moving when using the touch screen followed by the mouse with 3.67 bits/sec stationary, 2.6 bits/sec. moving and the touch pad with 2.12 bits/sec. stationary and 1.54 bits/sec moving (Figure 4).
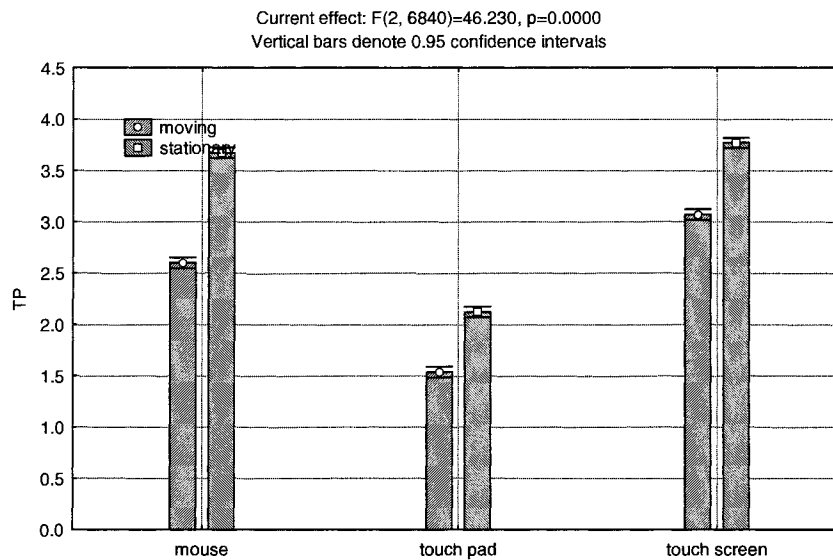
Figure 4. TP by pointing device and tractor state

Considering the interaction between typing device and age category of the participant there were significantly lower least square mean TP values found for age category 3 than or the other age categories across all tested devices (Figure 5).

Current effect: F(4, 6840)=15.962, p=.00000
Vertical bars denote 0.95 confidence intervals
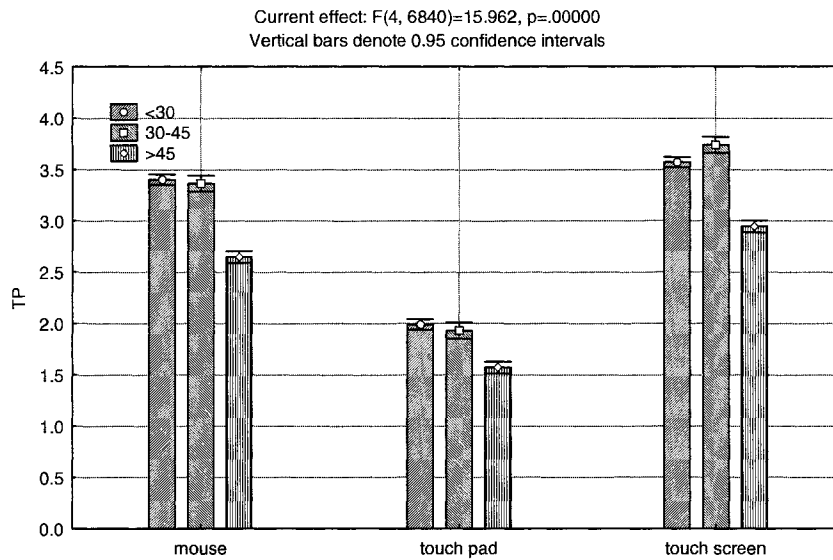
Figure 5. TP by pointing device and age categories

4.2 Tasks/second TS

The other performance measurement used in this study was the task/second TS value. This value was calculated by inverting the overall task time recorded by the test software. The overall task time was the total time it took the participants to correctly complete the task and includes the times to correct wrong selections. The TS values therefore represent a more complete measurement of performance than the TP.

For the TS values a significant difference was found between the tractor states (F=1830, p<0.0001), between the pointing devices (F=1101, p<0.0001), between the target sizes (F=395, p<0.0001), between age categories (F=473, p<0.0001), for the interaction of pointing device and tractor state (F=43.9, p<0.0001), the interaction of target size and pointing device (F=18.36, p<0.0001), the interaction of target size and tractor state (F=4.26, p=0.039) and the interaction of pointing device and age category (F=13.55, p<0.0001). There was no significant difference in TS values found between the different shapes, colors or menu numbers or for any other interaction between factors.

Between the different devices the best performance in TS was found when using the mouse with a least square mean TS of 0.21 tasks/sec followed by the touch screen with 0.19 tasks/sec and the touch pad with 0.13 tasks/sec.

Between the different tractor states significantly higher TS values were found in the stationary tractor with a least square mean of 0.21 tasks/sec compared to the moving tractor with a least square mean of 0.15 tasks/sec. For the different target sizes the better performance with a least square mean TS value of 0.19 tasks/sec was found for the big targets compared to 0.16 tasks/sec for the small targets.

Similar to the results for the TP values there was no significant difference in TS values found between age category 1 and age category 2 with a least square mean values for those two categories of 0.19 tasks/sec. Age category 3 had a significantly lower performance with a least square mean TS value of 0.15 tasks/sec.

For the interaction between pointing device and tractor state a significantly better performance in TS was found in the stationary tractor compared to the moving tractor for all devices. In the stationary tractor the best performance was found when using the mouse with a least square mean TS of 0.25 tasks/sec followed by the touch screen with 0.22 tasks/sec. and the touch pad with 0.16 tasks/sec. For the moving tractor there was no significant difference found between the mouse and the touch screen with least square mean TS values of 0.17 tasks/sec for both devices. A significantly lower TS value was found when using the touch pad with 0.10 tasks/sec (Figure 6).
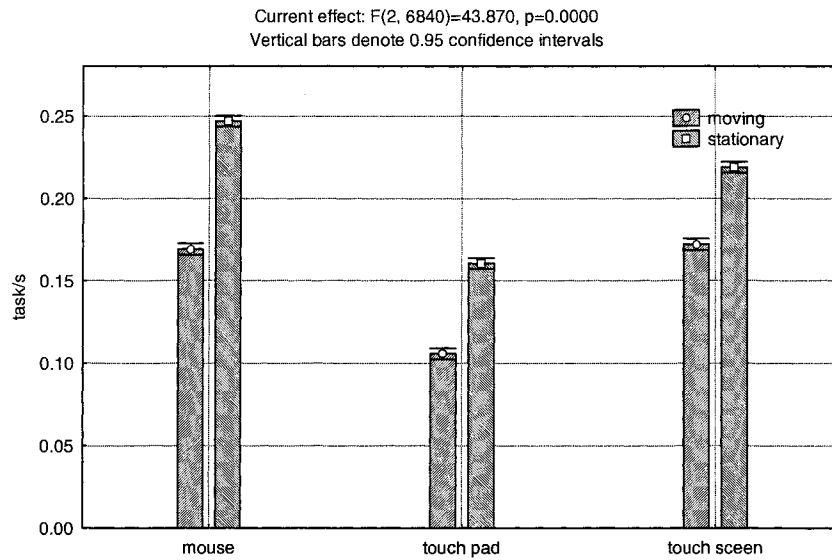
Figure 6. TS by pointing device and tractor state

For all devices a significantly better performance in TS was found for the big targets compared to the small targets. For the small targets the best performance was found when using the mouse with a least square mean TS value of 0.20 tasks/sec followed by the touch screen with 0.18 tasks/sec and the touch pad with 0.12 tasks/sec. For the big targets there was no significant difference found between the performance with the touch screen and the mouse and both had a least square mean TS value of 0.22 task/sec. With the touch pad the participants had a significantly lower performance with a least square mean TS of 0.14 tasks/sec (Figure 7).

Current effect: F(2, 6840)=18.358, p=.00000
Vertical bars denote 0.95 confidence intervals
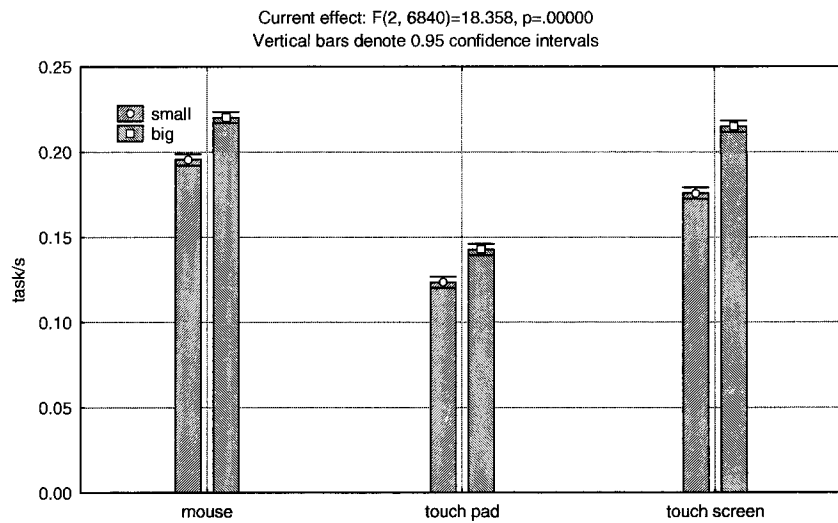
Figure 7. TS by pointing device and target size

For the interaction of tractor state and target size a better performance in the stationary tractor compared to the moving tractor was found for both target sizes. There was also a better performance found for the big targets compared to the small ones across both tractor states with the tractor state having the bigger influence on the performance than the target size (Figure 8).

Current effect: F(1, 6840)=4.2644, p=.03896
Vertical bars denote 0.95 confidence intervals
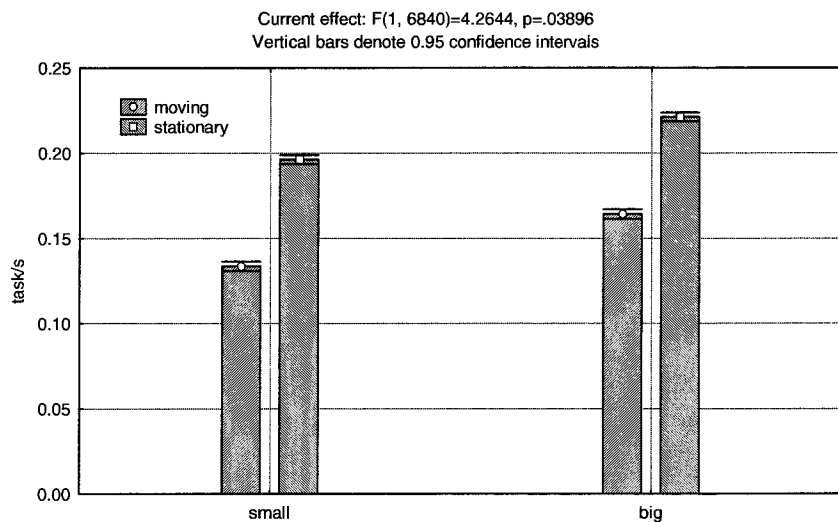
Figure 8. TS by tractor state and target size

For the interaction of pointing device and age category there was no significant difference in performance between the mouse and the touch screen for age category 2. Besides that across all age categories the best performance was achieved with the mouse followed by the touch screen and the touch pad. For all tested devices age category 3 had a significantly lower performance in TS compared to the other two categories (Figure 9).

Current effect: F(4, 6840)=13.552, p=.00000
Vertical bars denote 0.95 confidence intervals
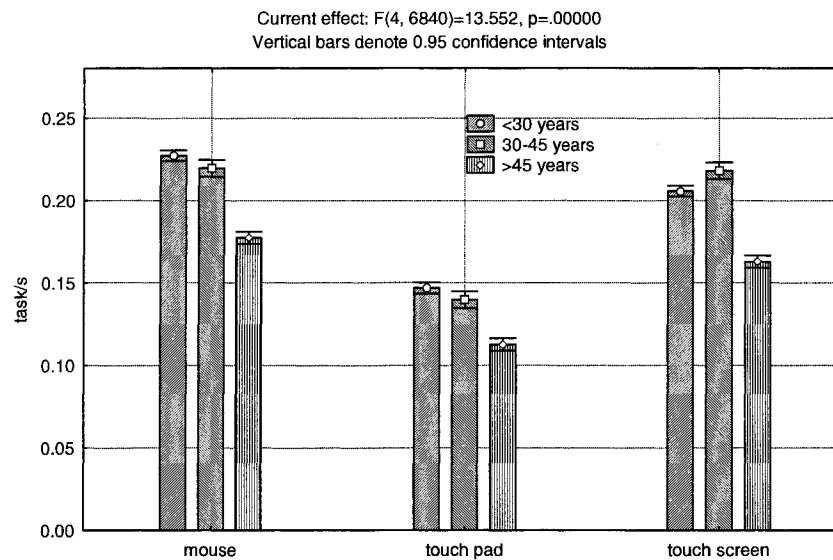
Figure 9. TS by pointing device and age category

4.3 Questionnaire

In addition to the objective measurements taken during this study a questionnaire was given to the participants to evaluate prior experience and their subjective impression of the tested devices. For every question of the subjects could give a score between 1 and 9. A Kruskal-Wallis nonparametric median test was used to analyze the responses for the different devices. For the purpose of this analysis the responses were categorized in low for ratings from 1 to 3, medium for 4 to 6 and high for 7 to 9.

Question 1: How Comfortable overall was the usage of the tested device in the given environment (moving tractor)?

For question 1 a significant difference was found between the responses for the different devices (Chi-Square=13.78, df=2, p=0.001). The multiple comparison found a significant difference between the mouse and the touch screen (p=0.0008). There were no significant difference found between the mouse and the touch pad or the touch screen and the touch pad. Overall the mouse achieved the highest rating amongst the participants with a mean of 7.22 followed by the touch pad with 5.61 and the touch screen with 4.44. The detailed results of the categorized responses to question 1 are shown in Table 2.

Device	low	medium	high	mean
mouse	1	4	13	7.22
touch screen	5	11	2	4.44
touch pad	4	6	8	5.61

Table 2. Categorized responses to question 1

Question 2: How would you rate the accuracy of the tested device in the given environment (moving tractor)?

For question 2 a significant difference was found between the responses for the different devices (Chi-Square=16.83, df=2, p=0.0002). The multiple comparison found a significant difference between the mouse and the touch screen (p=0.0007) and the mouse and the touch pad (p=0.007). There was no significant difference found between the touch screen and the touch pad. The order of the devices was the same as in question 1. The mouse achieved the highest ratings with a mean of 6.72 followed by the touch pad with 4.61 and the touch screen with 4.39. The detailed results of the categorized responses to question 2 are shown in Table 3.

Device	low	medium	high	mean
mouse	1	4	13	6.72
touch screen	4	13	1	4.39
touch pad	7	4	7	4.61

Table 3. Categorized responses to question 2

Question 3: How would you rate the operation speed of the tested device in the given environment (moving tractor)?

For question 3 a significant difference was found between the responses for the different devices (Chi-Square= 7.270, df=2, p=0.0264). The multiple comparisons found a significant difference between the mouse and the touch pad. There was no significant difference found between the mouse and the touch screen or between the touch pad and the touch screen. Overall the mouse again received the highest rating with a mean of 7.00 followed by the touch screen with 5.78 and the touch pad with a mean rating of 4.89. The detailed results of the categorized responses to question 3 are shown in Table 4.

Device	low	medium	high	mean
mouse	1	4	13	7.00
touch screen	3	7	8	5.78
touch pad	7	6	5	4.89

Table 4. Categorized responses to question 3

Question 4: How would you rate the influence of motion on the operation of the tested device in the given environment (moving tractor)?

For question 4 no significant difference was found between the responses for the different devices (Chi-Square= 0.596, df=2, p=0.7424) and motion had a big influence on the operation of all tested devices. The detailed results of the categorized responses to question 4 are shown in Table 5.

Device	low	medium	high	mean
Mouse	2	7	9	6.3
touch screen	2	7	9	6.4
touch pad	1	6	11	7.1

Table 5. Categorized responses to question 4

Question 5: How would you rate the influence of the button size on the operation of the tested device in the given environment (moving tractor)?

For question 5 no significant difference was found between the responses for the different devices (Chi-Square= 5.489, df=2, p=0.0634). The detailed results of the categorized responses to question 5 are shown in Table 6.

Device	low	medium	high	mean
mouse	6	4	8	5.4
touch screen	0	3	15	7.9
touch pad	0	6	12	6.9

Table 6. Categorized responses to question 5

Question 6: How would you overall rate the usability of the tested device in the given environment (moving tractor)?

For question 6 a significant difference was found between the responses for the different devices (Chi-Square= 6.38, df=2, p=0.0412). The mouse received the highest rating with a mean of 5.78 followed by the touch screen with 4.33 and the touch pad with 4.22 (Table 7).

Device	low	medium	high	mean
mouse	3	7	8	5.78
touch screen	8	5	5	4.33
touch pad	7	7	4	4.22

Table 7. Categorized responses to question 6

Question 7: Please rank the tested devices based on which one you would prefer to use in the tested environment (moving tractor)?

For question 7 a significant difference was found between the responses for the different devices (Chi-Square=10.500, df=2, p=0.0052). The multiple comparison found a significant difference between the mouse and the touch screen (p=0.022) and the mouse and the touch pad (p=0.0068). There was no significant difference found between the touch pad and the touch screen. The mouse received the best ranking with an average of 1.44 followed by the touch screen with 2.22 and the touch pad with 2.33 (Table 8).

Device	first	second	third	Mean
Mouse	11	6	1	1.444
touch screen	4	4	10	2.333
touch pad	3	8	7	2.222

Table 8. Responses to question 7

5.0 DISCUSSION

To determine which device provides the best usability to the user in the given environment both factors, the objective measurement of the performance a user can achieve with them and the subjective preference of the users have to be considered.

This study showed that consistent with the results of Rider et. al. (2003) all devices had a significantly worse performance in TP and in tasks/s in the moving tractor compared to the stationary tractor and also in using small targets compared to the larger targets.

It was also shown that for both performance measurements the age of the participants had a significant influence on their performance with the younger participants outperforming the older.

Concerning the TP values as a performance measure the participants had the same performance in the stationary tractor when using the mouse or the touch screen and a lower performance when using the touch pad. When looking at the moving tractor the tested devices had the same order that they had in the pilot study with the touch screen allowing the highest values followed by the mouse and the touch pad.

While according to Douglas et.al. (1999) throughput is the ISO recommended primary measurement of performance the TP values are calculated based on only a part of the

pointing task and only show one aspect of a device since they don't reflect any errors made while using the tested devices. Based on the results of his study Douglas (Douglas et. al. 1999) recommended that while throughput allows a single measure of both speed and accuracy to compute both movement time and error rate as separate dependent variables. The TS values on the other hand reflect the completion of the whole task and incorporate the errors the user made and the corrective action that was required and are therefore better suited to evaluate the usability and the overall performance a user can expect from the tested device. Concerning the overall performance measurement of tasks/s the mouse performed best in the stationary environment followed by the touch screen and the touch pad. In the moving tractor the mouse and the touch screen performed almost identically with the mouse having a small advantage when using the small targets while the touch screen has a small advantage when using the bigger targets. The touch pad performed significantly worse than the other devices for both target sizes and in both environments (moving, stationary).

It was also shown that overall the target size had the biggest influence on the performance of the touch screen followed by the mouse and the touch pad. This was most likely caused by the fact that the size of the human finger and the lack of sensing precision make precise touch screen interactions difficult.

Consistent with the results in previous studies (MacKenzie 1991, Bohan et.al. 2003) the participants favored the mouse in the subjective usability rating of the tested devices. The majority of the participants in the study gave it the best ratings concerning comfort, accuracy, speed and overall impression with no statistically significant difference between devices in the rating for the influence of the movement and the target size.

Between the Touch Pad and the Touch Screen the participants favored the Pad slightly in comfort and accuracy. The Touch Screen was slightly favored in speed and in the overall rating.

In the overall rating and ranking it was obvious that the mouse is the preferred device for the majority of the test participants. The touch pad and the touch screen were on average rated

the same with having big differences among individual participants depending on their personal preference.

This might be influenced by the fact that the majority of the participants indicated to have the most experience and prior exposure in using a mouse compared to the other tested devices.

6.0 CONCLUSION

The goal of this study was to determine which of the tested devices provides the best usability with standard applications in a moving off-road environment.

When looking at the objective performance data out of the tested devices overall the mouse and the touch screen seem to be best suited for this kind of environment. They are about equal in their overall performance with minor advantages for both devices in specific areas.

When looking at the subjective impression and rating of the test participants though the mouse looks like the clear winner. This would suggest that overall the mouse would be the best device to use.

When choosing a device for a certain application one has to consider other factors that might have a big influence on the usability. So could for example the choice of display size influence the performance because a bigger display would allow for bigger targets which would benefit the usability of the touch screen.

Another factor for consideration is that the mouse is the most commonly used device in an office environment and all of the participants were most familiar with the use of a mouse than the other devices. This most likely had an influence on the performance and the subjective rating of the mouse. It wasn't possible to get all participants the same amount of experience with all devices in the scale of this study. If in a real world application the user

would get more exposure to a touch screen he would get more used to it which could influence the performance and preference.

Out of the tested devices the mouse and the touch screen both seem to be well suited to be used in the described environment. To make a decision which one provides the best usability in a real world application a decision has to be based on the specific circumstances.

Future studies have to evaluate how environmental factors like dirt and dust influence the dependability and usability of the tested devices.

Besides that future studies have to evaluate the usability of different alphanumeric input devices and determine which one is suited best to provide a full user interface for standard applications in a moving off-road environment.

REFERENCES

Afifi, A., Clark, V.A., & May, S., (2004) Computer-aided multivariate analysis, Chapman & Hall/CRC, Boca Raton.

Bohan M., Solocum, J., Shaikh, A.D., Chaparro, B., (2003) Examining first-time usage of the RollerMouse, Usability News 5.2 2003

Douglas, S.A., Kirkpatrick, A.E. and MacKenzie I.S., (1999) Testing pointing device performance and user assessment with the ISO 9241, Part 9 Standard. Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI'99 pp.215-222 ACM: New York.

Fitts, P. M., (1954) The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 47 (1954), 381-391

Hochberg, Y., Tamhane, A.C., (1987) Multiple comparison procedures. Wiley, New York

ISO 9241-9, (1998) Ergonomic requirements for office work with visual display terminals, Non-keyboard input device requirements

ISO 9241-11, (1998) Ergonomic requirements for office work with visual display terminals, Guidance on usability

Jeffries, R., Miller, J.R., Wharton, C., Uyeda, K.M., (1991) User interface evaluation in the real world: a comparison of four techniques. Association for Computing Machinery, 1991

MacKenzie, I. S., (1991) Fitts' law as a performance model in human-computer interaction. Doctoral dissertation. University of Toronto: Toronto, Ontario, Canada.

MacKenzie, I.S., (2005) Within-subject vs. between-subjects design :which to use?. York University: Toronto, Ontario, Canada.

MacKenzie, I.S., & Buxton, W., (1993) A Tool for rapid evaluation of input devices using Fitts' law models. SIGCHI Bulletin, 25 (3), 58-63.

MacKenzie, I.S., & Buxton, W., (1994) The prediction of pointing and dragging times in graphical user interfaces. Interaction with computers, 6, 213-227.

MacKenzie, I. S., Sellen, A., & Buxton, W., (1991) A comparison of input devices in elemental pointing and dragging tasks. Proceedings of the CHI `91 Conference on Human Factors in Computing Systems, pp. 161-166. New York: ACM.

Salvucci, D.D., (2001) Predicting the effects of in-car interface use on driver performance: an integrated model approach. International Journal of Human Computer Studies vol. 55 , pp 85-107

Rider, K.A., Park, W., & Prohaska, M.P., (2003) Analyzing the effect of ride motion on the performance in-vehicle reaching tasks. Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society, Denver, Colorado, 2003.

Shneiderman, B., (1998) Designing the user interface: strategies for effective human-computer-interaction, 3rd ed. Addison Wesley Longman, Inc. 1998

Soukoreff, W., & MacKenzie, I. S., (1995). Generalized Fitts' law model builder. Companion Proceedings of the ACM Conference on Human Factors in Computing Systems – CHI '95, pp. 113-114. New York: ACM

Usability of alphanumeric input devices for office applications in a moving off-road environment

A paper to be submitted to *Applied Ergonomics*

Thorsten Baldus

ABSTRACT

In this study several alphanumeric input devices were evaluated to determine the one that provides the best usability with a standard application in a moving off-road environment. Out of the wide variety of alphanumeric input devices a focus group first identified five devices that had potential to be used in the described environment. An experiment was conducted with eighteen volunteers that performed a series of typing tasks in a moving tractor and rated the subjective usability of the tested devices. The standard QWERTY and the mini QWERTY keyboard are determined to be best suited for the tested environment and they achieved the best performance and the highest subjective ratings from the participants.

1.0 INTRODUCTION

The introduction of automated steering systems for agricultural machinery like tractors and combines has a big influence on the tasks an operator of such machinery has to perform. In the past the main task of the operator was to steer the vehicle and control the implement functions. With automated systems taking over the role of the operator switched to that of an observer that only monitors the system. For this reason many people believe the operator could utilize a personal computer to perform office tasks in the tractor cab while the automated system controls the machine. To do this effectively it is necessary to provide the operator with a usable computer interface.

An important part of the interface with a computer is alphanumeric input, as almost any interaction with a computer application requires the user to enter letters or numbers at some point. Depending on the kind of application these might be a couple of characters to name a

file or it can be the major part of the interaction such as when using text processing software or when writing e-mails.

Depending on the kind of application or the environment there are lots of options to realize alphanumeric input for a user. While in a typical office environment the QWERTY keyboard is the most common input device, other options to realize alphanumeric input are widely used for other applications. These options include, for example, the use of the number pad on cell phones for text entry, hand writing recognition on PDAs, or even voice recognition.

MacKenzie and Soukoreff (2002) studied text entry techniques for mobile computing. They did not evaluate the usability of any of the devices but instead discussed key factors for the evaluation of new technologies for mobile text entry including the methodology and the experimental design. They indicated that the most important issues for consideration when evaluating text entry techniques for mobile computing are focus of attention, text creation vs. text copy tasks, novice vs. expert performance, quantitative vs. qualitative measures, and the speed-accuracy trade off. They also discussed the optimization of different text input techniques by movement minimization and language prediction and they provided a survey of mobile text entry techniques both in research papers and in commercial products.

Cerney et.al. (2004) conducted a study to compare the usability of the standard QWERY keyboard with different alternatives of mobile text entry methods such as an onscreen QWERTY soft keyboard, a letter recognition system on a pocket PC, and a T9 text-input system on a cellular phone. Users were asked to perform common input tasks and user performance, accuracy and overall user preference for the four input methods were examined. They also compared the measured performance with the performance for the different input devices as predicted by Fitts' law. In this study, the participants had the best performance and the lowest cognitive load indicated by the users when using the standard QWERTY keyboard. This study also demonstrated that Fitts' law does not account for, nor predict, performance values including cognitive load or skill transfer and is only valid if the user's familiarity of the device placed interaction on a level of pure motor reflex.

MacKenzie (2002) evaluated different techniques for three key text entries for different mobile devices. Ten participants were asked to enter a series of phrases from the 'Phrase Sets for Evaluating Text Entry Techniques' (MacKenzie and Soukoreff 2003). Interaction issues that were examined in this study included the use of linguistic knowledge to accelerate input and the challenges in using auto repeat keying strategies to reduce the number of physical key presses.

Sears et al. (1993) investigated the relation between keyboard size and data entry rates for soft keyboards. As a result of this study it was shown that users made, and corrected, fewer errors when entering text on larger keyboards, with data entry speeds higher on larger keyboards, and the larger keyboards had better results in the subjective user preference.

Soukoreff and MacKenzie (2003) identified and described shortcomings in two statistics used to measure accuracy in text entry evaluations, minimum string distance (MSD) error rate and keystrokes per character (KSPC). In their study they developed a new framework and new measurements for evaluating text entry methods but fail to provide means to measure the bandwidth between humans and machines during text entry that includes both speed and accuracy.

MacKenzie and Soukoreff (2003) also describe and publish a phrase set for evaluating text entry techniques. The phrase set is a collection of 500 phrases of moderate length that are easy to remember and are representative for the English language. They also provide utility programs to calculate statistical properties like letter and word frequencies for this and other phrase sets.

ISO 9241, part 4 (1998) states the ergonomic requirements for linear detachable keyboards designed for stationary use. It provides guidance on the design of keyboards used for typical office tasks so that the limitations and capabilities of users are considered. It provides guidance based on ergonomic factors for keyboard layout arrangements, the physical characteristics of the individual keys and the overall design of the housing containing the

keys. It also includes recommendations for test methods to evaluate the usability of a product based on a user performance test and subjective user ratings (ISO 9241 1998).

ISO 9241, part 11 (1998) provides a guide for the evaluation of the usability of an office computer terminal. It states that in order to determine the level of usability achieved, it is necessary to measure the performance and the satisfaction of users working with a product. Measurement of usability is particularly important in view of the complexity of the interactions between the user the task characteristics and the other elements of the context of use. A product can have significantly different levels of usability when used in different contexts. The standard also states that measures of the performance and satisfaction of the user can provide a basis for the comparison of the relative usability of products with different technical characteristics which are used in the same context (ISO 9241, part 11, 1998).

This shows that lot of research was done to evaluate the performance and usability of different input devices for a wide variety of application and environments. Even though none of the prior research evaluated the use of input devices for standard applications in a moving vehicle the research done for other environments and tasks provide a good source of established evaluation techniques and testing methods which were adopted to be used in the course of this study.

The goal of this study was to identify the alphanumeric input device that is best suited to realize alphanumeric input with standard windows applications in a mobile off-road environment and identify the factors that influence the usability of the tested devices.

To evaluate the usability of the tested devices both the objective performance and the subjective impression of test participants will be considered. The standard QWERTY keyboard is the most commonly used device to realize alphanumeric input in an office environment and several studies showed the high usability and performance a user can expect from them. While standard QWERTY keyboards are proven to be well suited for an office environment it is expected that environmental factors like vibration and movement of the

environment have a negative influence on the performance and usability and other devices might be better suited for a mobile off-road environment.

2.0 METHOD

2.1 Selection of tested devices

Since there is a wide variety of devices available to realize alphanumeric input, not all available devices or methods could be evaluated in detail in the course of this study. The scope of this study had to be limited to representative devices having the potential for good usability in the given environment and that are likely to be accepted by the users. A focus group consisting of six John Deere employees with expert knowledge in the design of tractor user interfaces and/or a lot of experience in operating a tractor was formed to decide which of the commercially available input devices should be included in this study.

2.2 List of devices considered for this study and presented to the focus group

The following list represents the commercially available alphanumeric input devices and techniques that were presented to the focus group.

Standard QWERTY keyboard (104 keys)

> The standard QWERTY keyboard is the primary text entry device on most desktop systems. Their advantage is that they are widely used in office applications and almost everybody utilizing a computer is familiar with its use. The focus group decided that it should be included in this study since it is the most commonly used device and would serve as a good reference for all other tested devices.
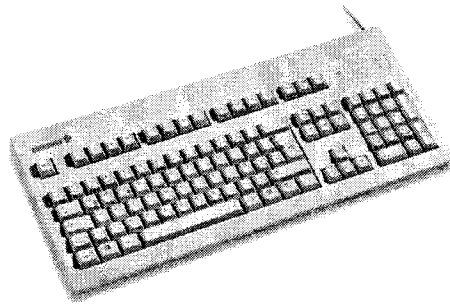
Figure 1. Standard QWERTY keyboard

Grandtec "Indestructible Keyboard"

This keyboard has the standard QWERTY layout and is available with either 109 or 85 keys that have similar size and spacing as the standard keyboard. The keyboard is made out of a rubber material and the key activation is realized through membrane switches. This results in different key travel and key activation forces compared to the standard QWERTY keyboard. This keyboard is fully sealed and factors like dirt and dust buildup shouldn't have an influence on its performance which makes it well suited for the tractor environment. The focus group therefore recommended including this keyboard in the study
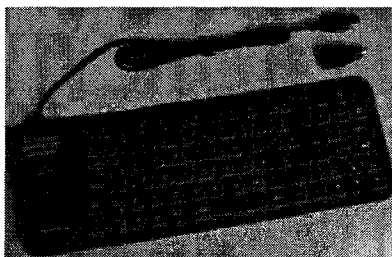
Figure 2. Grandtec indestructible keyboard

BIG Keys keyboard with 60 keys

This keyboard has a standard QWERTY layout with 60 oversized keys that support a 97 character set. The keys are bigger (20 x 20 mm) in size and have a different spacing (26 mm) compared to a standard keyboard. The oversized keys are expected to provide a benefit for non touch typists especially in the moving tractor since the keys should be easier to hit and movement of the tractor shouldn't influence the performance as much. This keyboard also had the feature that the key arrangement can easily be changed into an ABC-layout. The focus group recommended including this device in the study.
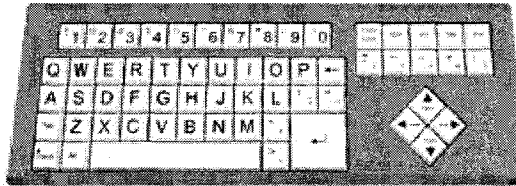
Figure 3. BIG Keys keyboard

Mini QWERTY keyboard

This keyboard features a standard QWERTY layout with 60 full size keys that support the full character set and functionality of standard keyboard in a smaller size. The key size, key spacing and key travel is similar to a standard keyboard but the overall size is smaller since there is no numeric pad and no assigned row for function keys included in the layout. Since none of this functionality is used in this study it is expected that the performance of this keyboard is very similar to that of the standard QWERTY keyboard. Because only limited space is available in a tractor cab this keyboard might be a good alternative to a standard keyboard and it was therefore recommended to be included in this study.
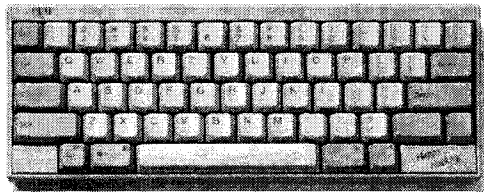
Figure 4. mini QWERTY keyboard

EasyReach TypeMatrix

This keyboard has a layout similar to the standard QWERTY with the keys in straight vertical columns rather than staggered and the "enter" and "backspace" keys in the center of the keyboard. Besides that some other functions like "delete", "insert", "home" and "end" are also located at differently compared to the standard keyboard. The focus group decided that this keyboard is very similar to the standard keyboard and that including it in the study doesn't provide any additional value. It was decided not to include it in the study.
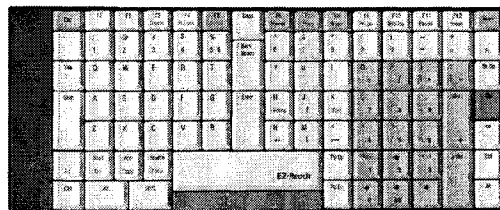
Figure 5. Easy reach TypeMatrix keyboard

Half QWERTY keyboard

This device has the same letter arrangement as does a standard QWERTY layout with only half the key set presented. The space bar is used to shift in a different mode to access the characters that would be located on the other side of the keyboard. Since the operation of this keyboard is very different from that of a standard keyboard a

long learning time is expected to become proficient with it. It was therefore decided
not to include it in the study since besides the reduced space it doesn't provide any
advantages over the standard keyboard.

Figure 6. Half QWERTY keyboard

Frog Pad

Similar to the half QWERTY this is a keyboard with a limited key set that utilizes
extra buttons to switch modes to represent all characters. Characters are not
represented in a QWERTY layout. For this keyboard there is also a long learning
period expected to become proficient with it and it was decided not to include it in the
study since it would offer no benefits besides the size.
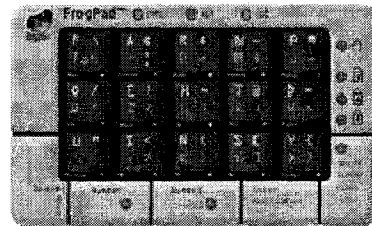
Figure 7. Frog Pad keyboard

Keiboard

The keiboard is a device with similar layout as a cell phone with additional keys for
computer specific functions. Text input with this device uses the Multi-press Input
Method the same principle that is used for text messaging on most cell phones
[Silfverberg et. al 2000]. A long learning period is expected for most users to achieve

expert status with this device and the maximum typing speed that can be expected from an expert user in an office environment lies around 21 wpm [Silfverberg et. al 2000]. Besides that the device doesn't have the full functionality of a standard keyboard. It was therefore decided not to include it in the study.
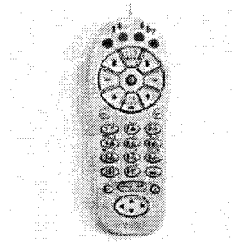
Figure 8. Keiboard

Orbitouch Keyless Keyboard:

The OrbiTouch™ keyless keyboard is an input device comprised of two domes upon which the hands rest. Each dome slides into one of eight positions from a central resting point, much like the eight major positions on a compass (N, NE, E, SE, S, SW, W & NW). To realize text entry the user has to slide the two domes each in a specific position and the combination of the two positions the domes are in determine the character entered. The operation of this keyboard is totally different than a standard keyboard. It would take a user a long time to get proficient with it and achieve any kind of performance. Except for users with limited use of their fingers there is no performance advantage over a standard keyboard expected even after the user reaches a certain proficiency level. It was therefore decided not to include this device in this study,
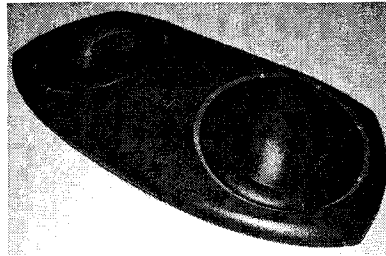
Figure 9. Orbitouch keyboard

DataHand Keyboard

The DataHand keyboard (Figure 10) offers a total of 132 keys through the use of five key switches clustered around the tips of each of the fingers and five modes that can be selected by the thumb. To realize input there is no hand movement required since the hands rest I none place and only the fingers move. This device requires a lot of precision in finger movement and is also expected to require extensive training to become proficient with it. It was decided not to include it in the study since the expected benefits don't outweigh the training requirements.
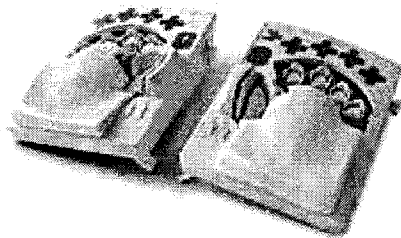
Figure 10. Data Hand keyboard

Fingerworks Touch Stream Mini Keyboard

This device is a small one handed keyboard with QWERTY layout and reduced key set. The keyboard has a flat smooth surface and works like a touchpad on a laptop.

This keyboard doesn't require any typing force and doesn't provide any tactile feedback to the user. The small size and flat surface that is easy to clean make it a well suited for the tractor environment where limited space and a dirty environment are an issue. It was therefore decided to include this keyboard in the study.
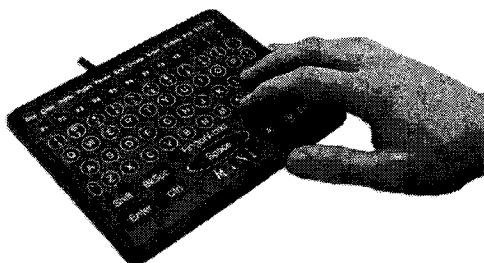
Figure 11. Fiingerworks Touch Stream Mini keyboard

Thumb Script

This device works like the graffiti text recognition in a PDA. To input a character the user has to slide his finger over a touch pad to realize gestures that represent the character and that are interpreted by driver software. This device doesn't provide full keyboard functionality and besides that would require the user to learn the necessary gestures for the operation of the device. Besides that multiple motions are required to represent a character which would limit the performance that can be achieved. It was decided not to include the device in the study.
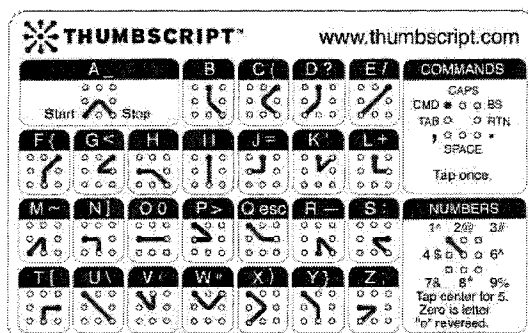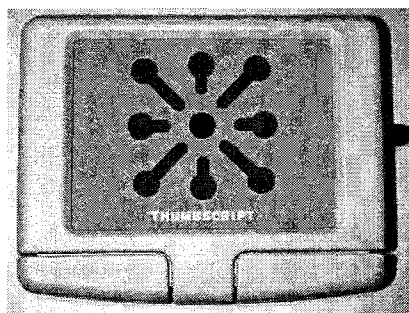
Figure 12. Thumb Script

X-Keys

This is a keyboard with reconfigurable keys in that allow customizing the layout and what functionality provided. The advantage over a standard keyboard is the ability to realize different keyboard layouts that could be specific to one application. The back draw is that all keys are identical so there is no way to provide different key sizes for important functions like "shift" or "return". At this point there was no real use for the additional features this keyboard provides and the lack of having special keys for important functions might hinder performance when using a standard windows application. It was decided not to include this device in the study.
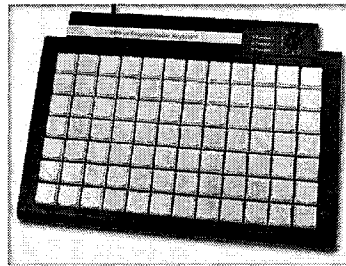
Figure 13. X-Keys keyboard

Ergo Dex

This is a keyboard with reconfigurable keys similar to the x-keys keyboard. The keys are programmable and can realize any function from a simple character to a macro in an application. Additionally this device allows for the free positioning of the keys on the surface of the device to generate any kind of layout wanted for a certain application. This device is similar to the X-keys keyboard and wasn't included in the study for the same reasons.
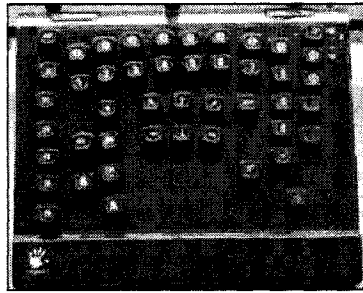
Figure 14. Ergo Dex keyboard

Chording keyboards (Twiddler/Bat keyboard)

Chording keyboards are smaller and have fewer keys than the typical keyboard, typically one for each finger and possibly the thumbs. Instead of the usual sequential, one-at-a-time key presses, chording requires simultaneous key presses for each character typed, similar to playing a musical chord on a piano.

The primary advantage of the chording keyboard is that it requires far fewer keys than a conventional keyboard. For example, with five keys there are 31 chord combinations that may represent letters, numbers, words, commands, or other strings. With fewer keys, finger travel is minimized because the fingers always remain on the same keys. The back draw is that the operation of them is complicated and requires a lot of training to achieve a satisfying performance. It was therefore decided not to include any of them in this study.
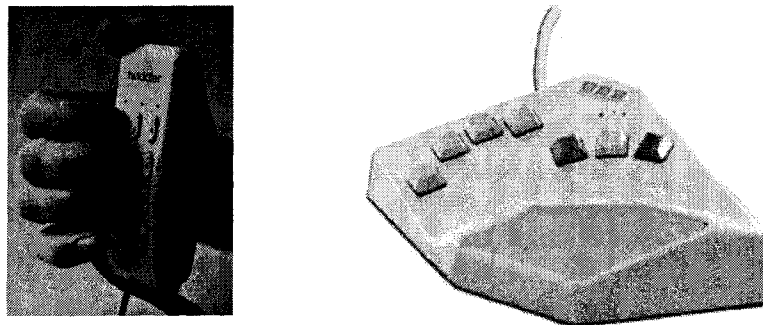
Figure 15. Twiddler keyboard and the Bat Keyboard

On Screen soft keyboard

On screen soft keyboard could feature different layout and key set options and could be used with either a touch screen or any other kind of pointing device. The focus group decided that an on screen keyboard is not well suited for extended text entry like mail client or text processing software. Besides that some of the already limited screen space would have to be taken away from the application to realize a soft keyboard. It was decided not to include it in this study.

Speech Recognition

Speech recognition was not considered as an alternative to realize alphanumeric input in this study. Some limited amount of speech recognition can be used when selecting from a limited selection of pre-programmed entries like names in an address book of a cell phone but with the technology that is available today it is not usable for general purpose text input [MacKenzie et. al. 2002]. Besides that the focus group pointed out that even with speech recognition being available for use in an office environment the high amount of ambient noise in a tractor cab might cause additional problems. It was decided not to consider any speech recognition systems in the course of this study.

2.3 Devices included in this study

Overall the focus group recommended to only including conventional keyboards in different varieties in this study. Other devices that have a different way of realizing the input require a fair amount of training for the user to become proficient with them and for some of the devices it is questionable if there will be any performance or usability advantage over a traditional keyboard once proficiency is achieved. The devices that were chosen to be included in the study represent not only the specific device that was tested they also give some insight in the impact of different keyboard characteristics (like key size, spacing, tactile feedback and activation force).

The devices included are:

- Standard QWERTY keyboard
- Big Key keyboard (in QWERTY and ABC configuration)
- Grandtech Virtually Indestructible Keyboard
- Fingerworks mini keyboard
- Mini QWERTY keyboard

The key specifications of the included devices are shown in Table 1.

Keyboard	switch actuation		key travel	key size	key spacing	size
Standard QWERTY	55 g	0.539 N	4mm	13mm x 14mm	19mm	160x455 mm
Big Key (QWERTY/ABC)	55g	0.539 N	4mm	20mm x 20mm	26mm	175X475 mm
Indestructible	80 g	0.7848 N	2mm	13mm x 13mm	19mm	130x 345 mm
Fingerworks	touch	touch	0 mm	12mm x 12mm	15mm	140x180 mm
mini QWERTY	55 g	0.539 N	4mm	13mm x 14mm	19mm	120x290 mm

Table 1. Key specifications for included devices

2.4 Participants

The study was conducted with eighteen volunteers (16 male, 2 female) among the employees of the John Deere PEC in Waterloo Iowa. The participants were required to have some experience in operating desktop computers as well as experience in operating a tractor. Nine of the subject were touch typists and nine of them were non touch typists. In a self assessed typing experience rating all of the participants that were touch typists rated them self as very experienced. Among the non touch typists there were 3 participant that rated themselves as

very experienced , 4 with medium typing experience and 2 who rated themselves having little typing experience. The age of the participants ranged from 21 to 58 years (mean = 39.4). For analytical purposes the participants were divided in different age categories. 7 of the participants were younger than 30 years (category 1), 5 were between 30 and 45 years of age (category 2) and 6 participants were older than 45 years of age (category 3). Since for this test common phrases in the English language were used, only participants whose first language is English were considered.

A matrix of the characteristics for all participants is shown in Table 2.

	Age Category	Low Exp.	Med Exp.	High Exp.
Touch Typists	1	-	-	6
	2	-	-	2
	3	-	-	1
Non Touch Typists	1	-	-	1
	2	-	2	1
	3	4	-	1

Table 2. Participants summary

2.5 Apparatus

The tractor used for this study was a John Deere 8020 series tractor equipped with auto steering capability and path planning software. This allowed putting the tractor in auto steering mode during the test so it would travel on the same predetermined path at a consistent speed during every test without any input from the test subject or the investigator.

A Fujitsu lifebook tablet pc with an integrated 12" LCD display was mounted into tractor cab. Additionally a 50x30 cm lab tray was used to provide a surface to rest the tested devices. A picture of the test setup in the cab is shown in Figure 17 and Figure 18.
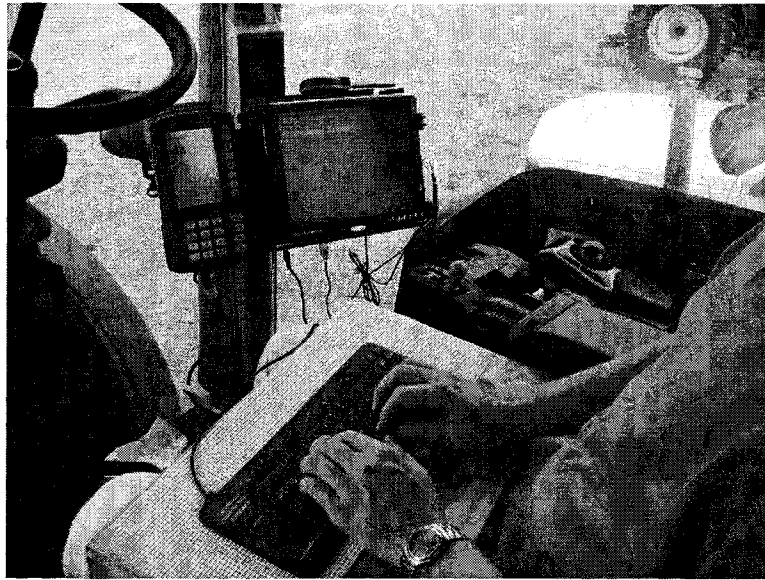
Figure 17. Test setup in tractor cab
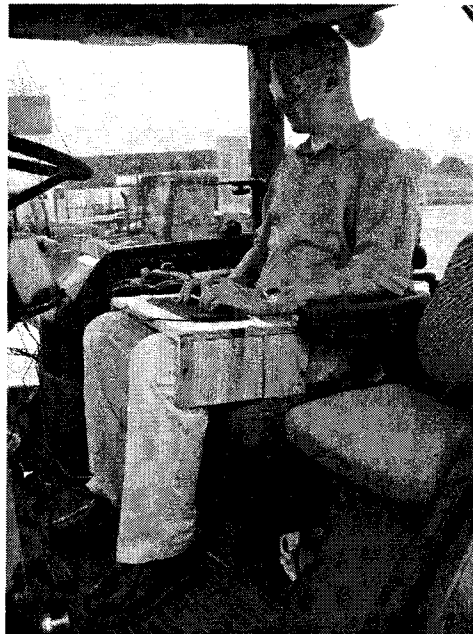
Figure 18. Test setup in tractor cab (side view)

The experimental software was a self developed MetaCard application for text entry evaluation. When launched the program reads a file containing a series of text phrases of about 25 characters in length. The phrases used were part of the "Phrase Set for Evaluating Text Entry techniques" by MacKenzie and Soukoreff (2003). The phrase set contains a total of 500 phrases that were chosen to be representative of English and easy to remember. The phrase set was tested for its correlation with common English using the frequency count in Mayzner and Tresselt's corpus (Mayzner and Tresselt, 1965). The result was r=.9845 for the single letter correlation and r =.9418 for the digraph correlation. The phrases ranging from 16 to 43 characters (mean 28.6). There were 2712 total words, including 1163 unique words. The words ranged from 1 to 13 characters (mean = 4.46) (MacKenzie et. al. 2002).

For each task the software randomly picked one phrase out of the phrase set while ensuring that the same phrase wasn't presented more than once to a test subject within one session.

During the text entry the phrase was displayed at the screen and the test software highlighted and put a pointer on the next character it expected so the participants could refer to the displayed phrase in case they lost track after performing a false character entry. There was also a visual and audio signal given by the test software if a false character was entered. A typical screen of the test software is shown in Figure 19. The top line shows the presented text phrase while the line below shows the progress of input. Between both lines an arrow is shown that points to the next character that is expected by the software.
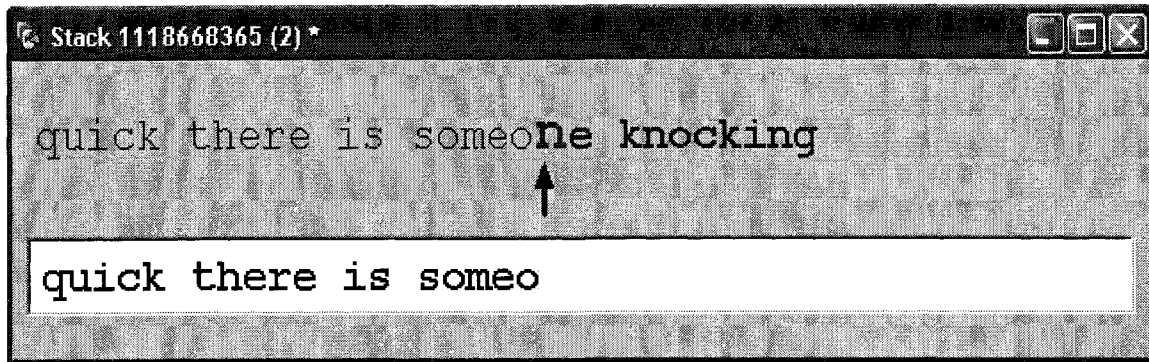
Figure 19. Screen shot test software

Besides presenting the typing task to the participant the test software also recorded performance data during the test. For each phrase entered the software recorded the following information

- Text phrase presented to the participant.
- Character string entered by the participant
- Time to enter character string (time started with the first character entered)
- Number of errors between presented phrase and character string entered by test subject(character by character comparison)

As in previous studies (MacKenzie and Zang 2001) the starting point for the time measurement was set to the entry of the first character in each string. This was necessary since there is no reference from which to time the entry of the first character. The first character of each phrase was also excluded from the calculation of correlation of the letter frequency in the test phrases with common English.

2.6 Experimental Design

According to MacKenzie (2005) a within subject design was used for this study with the following independent variables:

- Keyboard alternatives (list tested keyboards)
- Session (2 sessions)
- Tractor state (moving stationary)
- Typing status of participant (touch typist/non touch typist)

2.7 Procedure

The test was conducted in two sessions with three devices being evaluated in each session. During the test the participants in this study were placed in the driver seat of a tractor and asked to perform a series of typing tasks presented to them by test software utilizing different input devices. The participants first had a chance to familiarize themselves with the tested devices and the test software. For this the participant performed a series of 3 test phrases with each tested device before any data was recorded to make sure the testing procedure and the operation of each device was understood

During the actual test the software presented one phrase at the time to the participants. The participants were instructed to remember the phrase and type it from their memory if possible rather than copying it. This approach simulates the text creation task in that the user knows exactly what to enter. This is in contrast to a text copy task wherein the user's focus of attention continually switches between the source text and the keyboard [MacKenzie and Sourkoreff 2002]. The participants were instructed to aim for text entry speed and accuracy and enter the phrase "as quickly and accurately as possible". Participants were instructed to ignore mistakes and to continue with the rest of the phrase in the event of an error.

To guard against any effect the test order might have on the outcome a balanced design was used to test the different devices and tractor states in a different order for each participant. The test order details of the balanced design are shown in Appendix C.

Within one session the participant used 3 different devices in both a moving and a stationary environment. When using the device in the moving tractor the participant only had to

perform the typing task. All tractor operations were performed by an auto steering system and monitored by the investigator.

After the completion of each test session the participants were asked to answer a questionnaire to investigate their subjective opinion about the tested devices. Every participant was asked to rate the performance of the tested devices based on there usability in the tested environment and to pick their favorite device which provided them with the highest usability in the tested environment.

Some of the tested devices were new to all participants. Even if their operation was similar to that of the standard keyboard everybody used before they didn't have a chance to reach the same level of experience with all of the tested devices. But rather than focusing on the potential or expert text entry rate of a particular device the "immediate usability" of the tested devices is important. In other words, it may be a moot point to establish the expert, or potential, text entry rate for an input device, if prolonged practice is required to achieve it. Consumers may be "turned off" by their initial experience and frustration (McKenzie and Zang 2001).

2.8 Moving Vehicle Environment

The test was conducted over the same track and under similar conditions as a previous study. The tractor was traveling on a predetermined path at a constant speed of 7.5 km/h with an automated steering system monitored by the investigator handling all tractor controls. During the previous study acceleration data of the tractor frame and cab was recorded and it was decided by experts at John Deere that the conditions are representative for field work conditions. It was determined that it was not necessary to record any acceleration data for this study since the test conditions were very similar.

3.0 ANALYSIS

There are two primary evaluation metrics for text input: speed and accuracy. It is well established that there is a tradeoff between speed an accuracy and It is possible to achieve higher gross text entry speed if someone is willing to sacrifice accuracy and vice versa (MacKenzie and Sourkoreff 2002). It is therefore necessary to measure both speed and accuracy and to evaluate them together to find the device with the overall best performance. One way to evaluate speed and accuracy together in one measurement is to calculate the net typing speed. Net typing speed is adjusted compared to the gross typing speed by reducing the speed depending on the amount of errors made during the typing task.

For this study the Net Typing Speed (NTS) and Gross Typing Speed (GTS) were calculated according to the 'Australian standard for keyboarding speed tests' (AS2708-2001) using the following formula:

$$NTS = \frac{NH}{t} \tag{1}$$

$$GTS = \frac{GH}{t} \tag{2}$$

Where NH are the Net Hits, GH are the Gross Hits, which means the number of keystrokes, and t is the time.

$$NH = \frac{GH}{EH} \tag{3}$$

EH are the Error Hits.

$$EH = WE * SWL \qquad\qquad (4)$$

Where WE are the word errors which means the number of words typed incorrectly and SWL is the standard word length. To get the Word Errors post test analysis had to be done to identify how many wrong words were typed incorrectly for every text entry string during the test.

In the AS2708-2001 standard the commonly accepted 5-stroke standard word which is also consistent with the finding is some additional research (Gentner, Grudin, Larochelle, Norman & Rumelhart 1983) is used.

In contrast to this according to the research conducted by L.J. West (1968), D.J. Perry (1968) and B.S. Ober (1983) the average word, including spaces and punctuation marks, in written American business language contained respectively 5.97, 5.83 and 6.13 keystrokes. This would support the assumption that for the evaluation of typing performance a 6-keystroke per word system would be a better choice since it is closer to the actual written language (AS2708-2001).

If a speed test was intended to measure actual keyboarding speed, than the standard word should resemble as closely as possible the average word in current business language, i.e. 6 keystrokes. On the other hand, it is argued that the main purpose of the standard speed test was not to provide an accurate measurement of speed, but rather to provide an effective ranking method. Therefore the unit of measurement used is not as important as the fact that the unit should remain constant (AS2708-2001). For this reason the 5-character standard word length (SWL) was used for the calculation of the typing performance in this study.

To calculate the Accuracy the following formula was used according to AS2708-2001

$$Accuracy = \frac{NH}{GH} \qquad\qquad (5)$$

Additionally all typing speeds were transformed from characters per second (cps) to the common unit for text entry speed words per minute using the Standard Word Length of 5 characters. The formula used for this was:

$$wpm = \frac{cps * 60}{5}$$
(6)

4.0 RESULTS

4.1 Gross typing speed

The gross typing speed GTS represents the number of key presses in a given time and represents the most direct evaluation of typing speed.

A multivariate GLM analysis found significantly higher GTS values for all devices in the stationary tractor compared to the moving tractor (F=23.74, p< 0.0001). A significant difference was found between the tested devices (F=529.11, p<0.0001), typing styles (F=140.12, p<0.0001) and between age category of the participant (F=193.68, p<0.001)

The mean GTS in the moving tractor was 31.9 wpm compared to 34.2 wpm in the stationary tractor

With the standard QWERTY keyboard the participants had the highest mean GTS with 48.2 wpm followed by the mini QWERTY with 46.6 wpm and the BIG Keys QWERTY with 34.6 wpm. The indestructible keyboard allowed a mean GTS of 31.6 wpm, the Fingerworks keyboard 24.9 wpm and the BIG Keys ABC 12.6 wpm. For detailed results see Figure 20.

Current effect: F(5, 3168)=529.11, p=0.0000
Vertical bars denote 0.95 confidence intervals
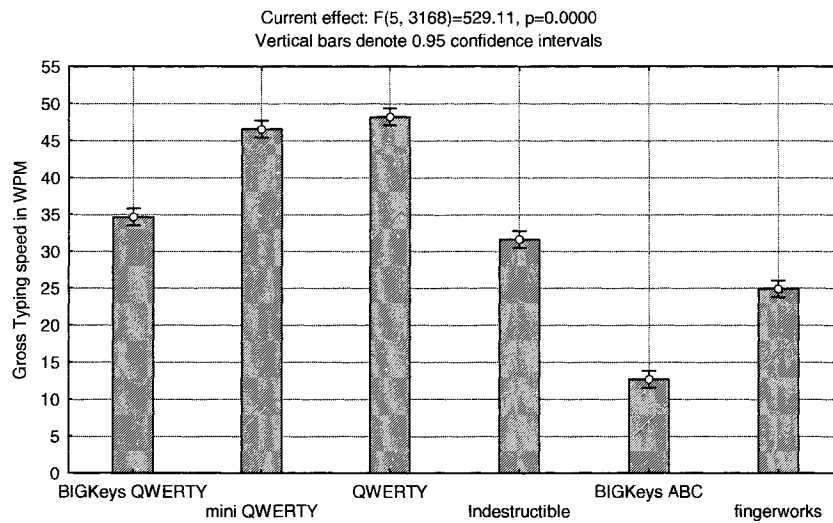
Figure 20. GTS by tested devices

The mean GTS for the touch typists across all devices was 35.9 wpm and the non touch typists 30.3 wpm

The highest mean GTS amongst age categories was achieved by category 1 (30 and younger) with 38.7 wpm followed by category 2 (30-45) with 33.8 wpm and category 3 (older than 45) with 26.7 wpm.

There was also a significant difference found for the interaction of tractor movement and typing device (F=2.38, p=.0364), typing style and typing device (F=94.01, p<0.0001, age category and typing style (F=53.33, p<0.0001) and age category and typing device (F=10.32, p<0.0001).

For all devices but the BIG Keys ABC a higher GTS was achieved in the stationary tractor compared to the moving tractor. The biggest difference in GTS was found for the Mini QWERTY (4.5 wpm or 9.6%) followed by the Fingerworks (3.6 wpm or 13%) and the standard keyboard (3 wpm or 6%). Detailed results are shown in Figure 21.
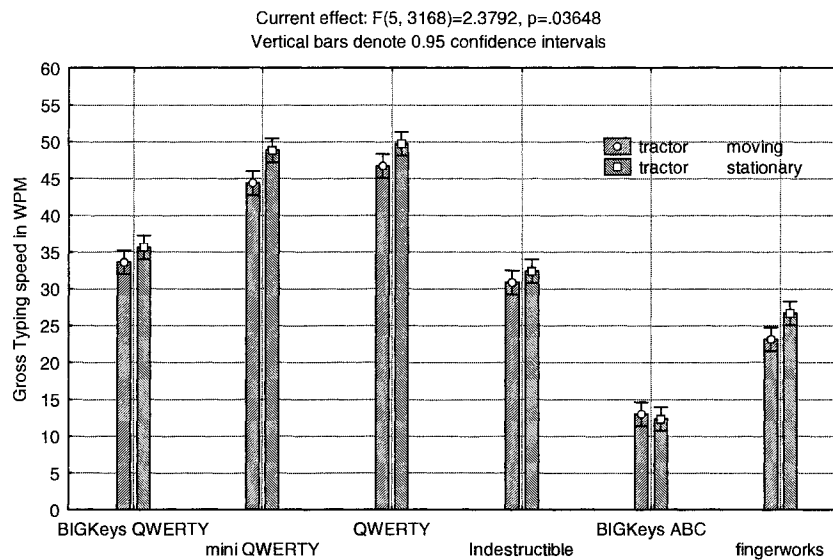
Figure 21. GTS by tractor movement and device

The best performance in GTS for the interaction between device and typing style was found for touch typists using the standard QWERTY or the mini QWERTY keyboard with a GTS of 57.6 wpm and 56.8 wpm respectively. For the non touch typists the best performance was found with the standard QWERTY with 38.8 wpm followed by the BIG Keys QWERTY with 36.7 wpm and the mini QWERTY with 36.4 wpm. With the standard QWERTY and the mini QWERTY there was also the biggest difference in performance between the touch typists and the non touch typists with a difference of 33% or 38 wpm for the standard QWERTY and 36% or 36 wpm for the mini QWERTY keyboard. With all other devices there was only little difference between the typing styles with the non touch typists outperforming the touch typists with the Fingerworks (5.4 wpm or 19 %). For detailed results check Figure 22.

Current effect: F(5, 3168)=94.014, p=0.0000
Vertical bars denote 0.95 confidence intervals
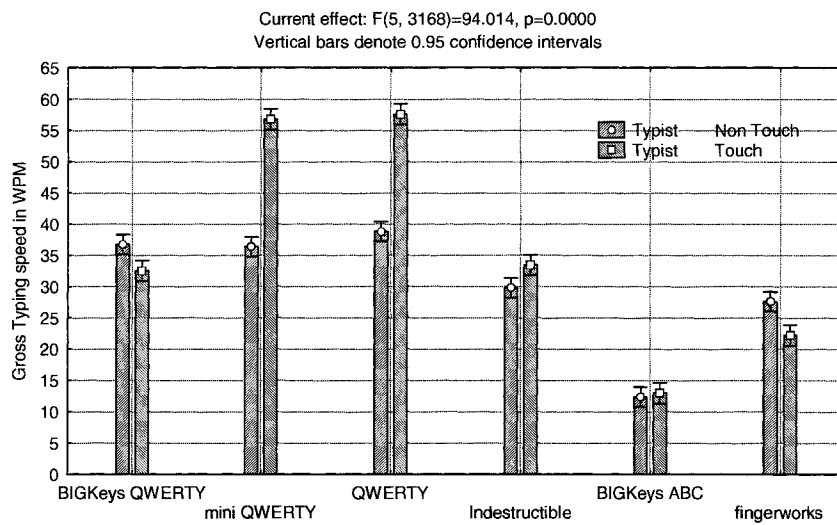
Figure 22. GTS by device and typing style

For the interaction between age category and typing style there was no significant difference between touch typists and non touch typists for age category 1 (30 and younger) but a significant difference for age categories 2 (30-45) and age category3 (45 and older) with higher GTS values for the touch typists in both cases. For detailed results check Figure 23.

Current effect: F(2, 3168)=53.329, p=0.0000
Vertical bars denote 0.95 confidence intervals
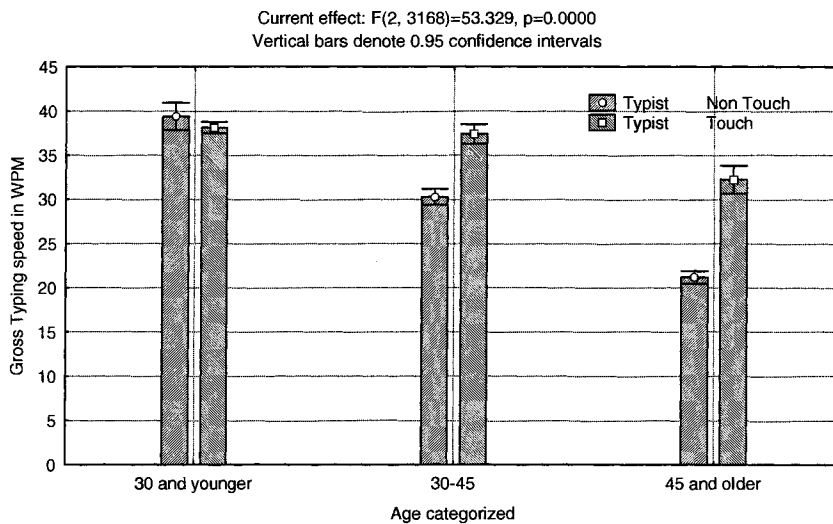
Figure 23. GTS by age category and typing style

Concerning the interaction between age category and device used, age category 1 performed best followed by age category 2 and age category 3 for most devices. The only two exceptions are the indestructible keyboard which had no significant difference between age categories one and two which both had a higher mean GTS than category 3 and the BIG Keys ABC keyboard that had no significant difference between any age categories. For detailed results see Figure 24.

Current effect: $F(10, 3168)=10.325$, $p=0.0000$
Vertical bars denote 0.95 confidence intervals
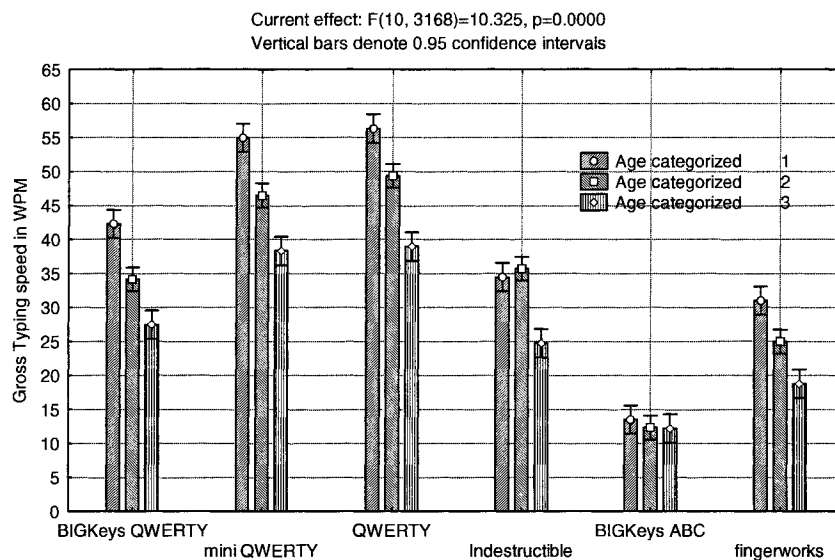
Figure 24. GTS by device and age category

All of the touch typing participants categorized themselves as having a lot of typing experience and a Kendall's Tau analysis found a significant correlation of 0.67 between the two factors. The amount of typing experience therefore wasn't a factor that influenced the performance amongst them. When only looking at the participants that were 'non touch' typists there was a significant difference ($F=506$, $p<0.0001$) in GTS between the different experience levels and also for the interaction of experience level and typing device used ($F=14.7$, $p<0.0001$). The detailed results are shown in Figure 25.
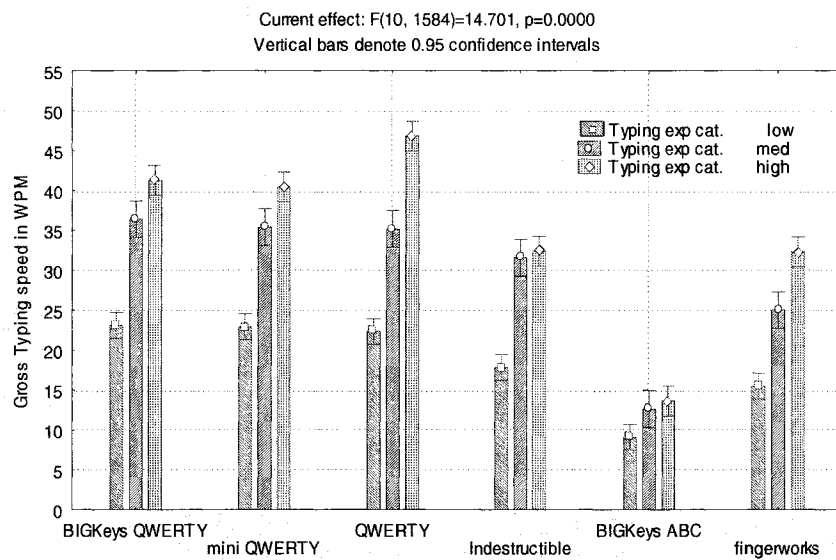
Figure 25. GTS by device and typing experience

The participants that indicated little typing experience had the significantly lower performance in GTS than the other two groups for all tested devices. Between the participants that indicated medium typing experience and those that indicated a high level of typing experience there was no significant difference in mean GTS found when using the indestructible and the BIG Keys ABC keyboard. For all other devices there was a significantly higher mean GTS found for the participants with a high level of typing performance with the biggest difference in performance found when using the standard QWERTY keyboard.

4.2 Accuracy

Besides the typing speed the accuracy that can be achieved with a given device is an important factor to evaluate its performance.

A multivariate GLM analysis found significantly higher accuracy values (F=22.93, p<0.0001) for all devices in the stationary tractor compared to the moving. Significant

differences were also found between the tested devices (F=47.59, p<0.0001) and different
age categories (F=6.23, p=0.00199).

The mean accuracy in the moving tractor was 83.6% and in the stationary tractor it was
87.2%

Between the different devices the highest accuracy was found with the BIG Keys ABC
keyboard with 92.8% followed by the standard QWERTY with 89.5%,the mini QWERTY
with 88.2%, the BIG Keys QWWERTY with 86.2%, the indestructible keyboard with 78.8%
and the Fingerworks with 77%. The detailed results are shown in Figure 26.
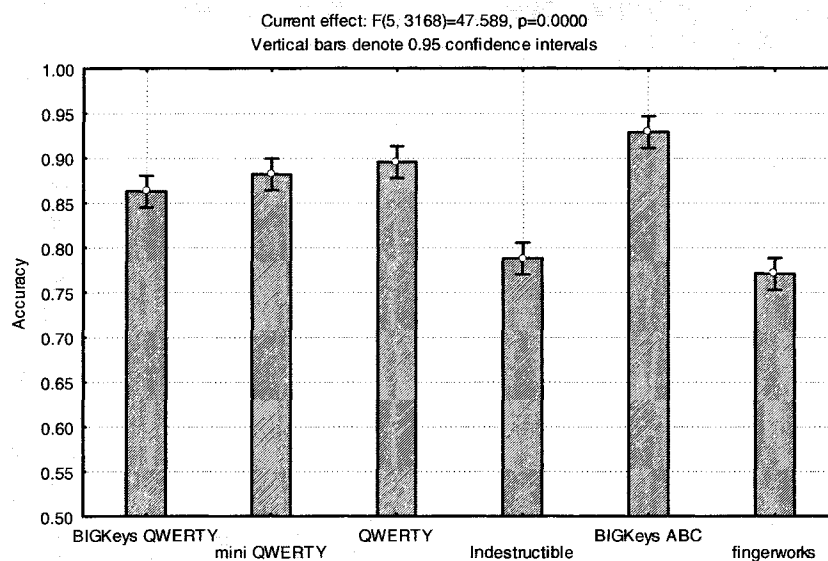
Figure 26. Accuracy by devices

Between the different age categories the highest mean accuracy was achieved by age
category 3 (45 and older) with 86.7 % followed by age category 2 (35 to 45) with 86% and
age category one (35 and younger) with 83.6%.

There was also a significant difference found for the interaction between typing style and tractor movement (F=4.328, p=0.0377), typing style and input device (F=5.152, p=0.0001) and tractor movement and input device (F=4.929, p=0.0002).

For the interaction between typing style and tractor movement there was a bigger difference in accuracy found between the moving and the stationary tractor for non-touch typists (5.2%) than for touch typists (2%) (Figure 27).
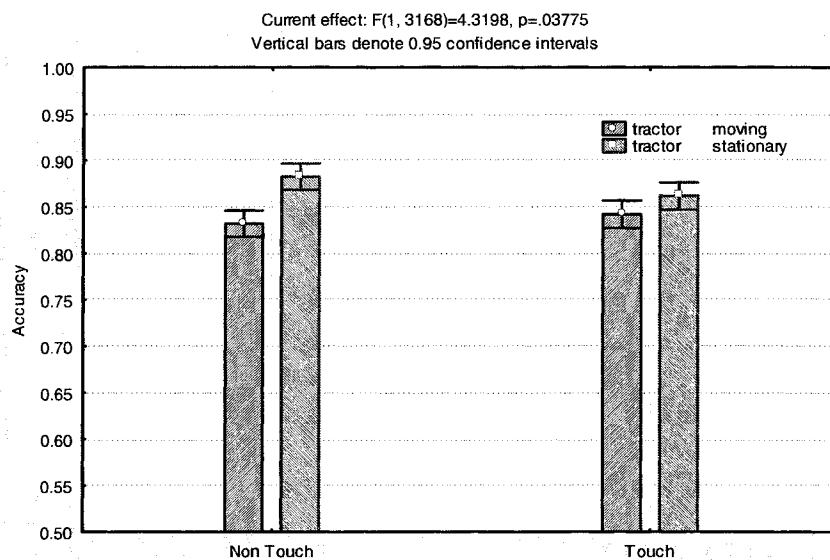
Figure 27. Accuracy by typing style and tractor movement

For the interaction between typing style and input device there was a higher accuracy found for touch typists when using the QWERTY and the mini QWERTY keyboard while the non-touch typists had a higher accuracy when using the indestructible keyboard. For all other devices there was no real difference in accuracy between the two typing styles. See Figure 28 for details.
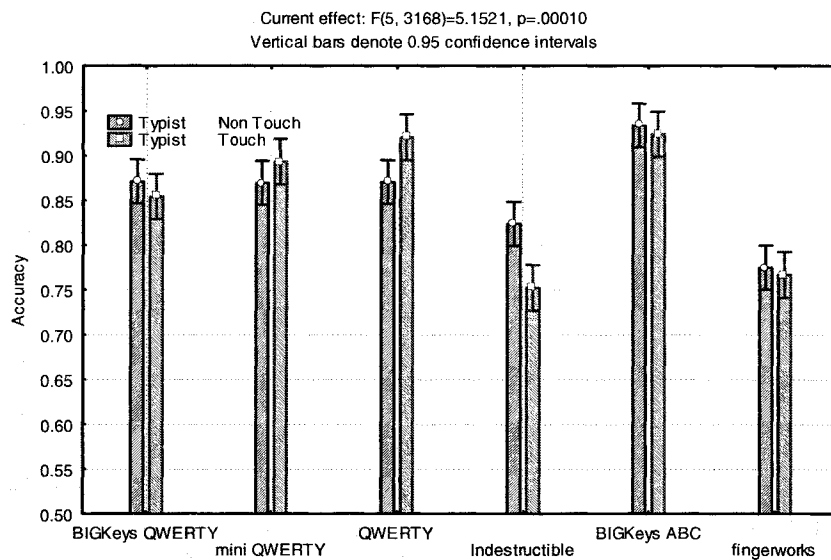
Figure 28. Accuracy by typing style and input device

Concerning the interaction of tractor state and device a significant difference in accuracy between the moving and the stationary tractor was found when using the Fingerworks keyboard (10.4%). For all other devices there was no significant difference found between the stationary and the moving tractor.

When just looking at the 'non touch' typists there was a significant difference found between different levels of typing experience (F= 4.60, p=0.0101) and for the interaction of typing experience and input device (F= 5.46, p<0.0001) used. See Figure 29 for the detailed results.
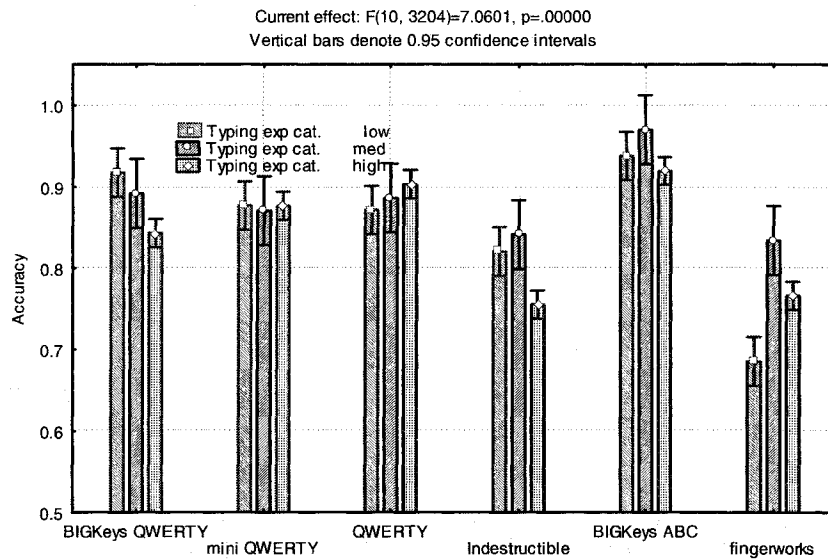
Figure 29. Accuracy by device and typing experience

4.3 Net typing speed NTS

To combine the typing speed and accuracy measurements into one variable the net typing speed (NTS) was calculated according to AS2808.

A multivariate GLM analysis found significant differences in NTS between the moving and stationary tractor (F=28.18, p<0.0001), between, between typing styles (F=86.20, P<0.0001), between the different age categories (F=97.47, P<0.0001) and typing devices used (F=351.38, p<0.0001).

The mean NTS in the moving tractor was 27.3 wpm compares to the stationary tractor with 30.2 wpm. This is a decrease of 4.6 wpm in the moving and 4.0 wpm in the stationary tractor compared to the GTS.

The mean NTS across all devices for touch typists was 31.3 wpm and 26.2 wpm for non touch typists. Compared to the GTS this means a decrease of 4.1 wpm for the non touch typists and 4.6 wpm for the touch typists.

Concerning the influence of the age category on the NTS the highest mean NTS was found for age category 1 (30 and younger) with a NTS of 33.3 wpm followed by age category 2 (30 -45) with 29.4 wpm and category 3 (45 and older) with 23.5 wpm.

Between the different devices the highest mean NTS was found when using the standard QWERTY keyboard with 43.9 wpm followed by the mini QWERTY with 41.9 wpm, the BIG Keys QWERTY with 30.0 wpm, the indestructible with 25.2 wpm, the Fingerworks keyboard with 19.7 and the BIG Keys ABC with 11.8 wpm. See Figure 30 for the detailed results.
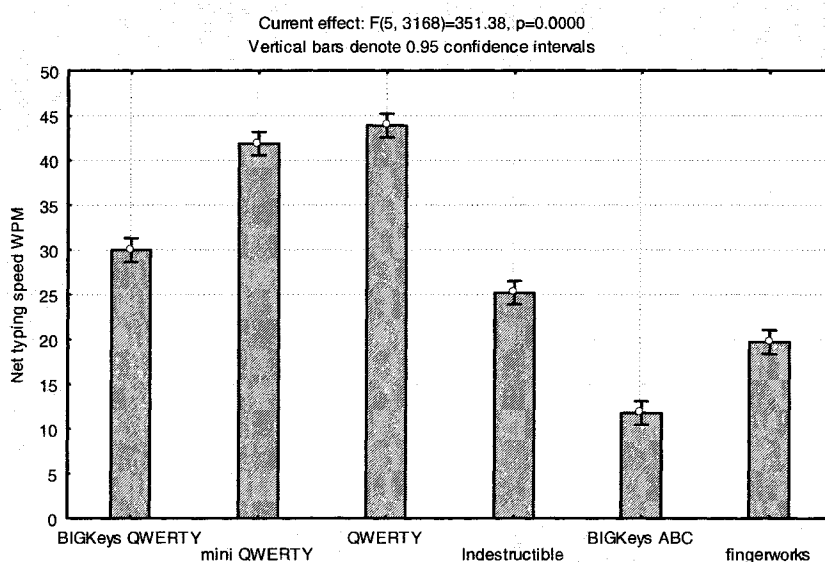
Figure 30. NTS by typing device

There was also a significant difference in NTS found for the interaction of typing style and typing device used (F=73.37, p<0.0001), for the interaction of age category and typing style (F=33.44, p<0.0001) and between age category and typing device used (F=5.710, p<0.0001).

The best performance in NTS for the interaction between typing style and device used was found for the touch typists using the standard QWERTY (53.7 wpm) or the mini QWERTY (51.6 wpm) keyboard. For non touch typists the best performance was found with the standard QWERTY with 34 wpm followed by the mini QWERTY with 32.1 wpm and the BIG Keys QWERTY with 31.9 wpm. The biggest difference between the performance of non touch typists versus touch typists was also found for the standard QWERTY and the mini QWERTY keyboard with 19.7 wpm (37%) and 19.5 wpm (38%) respectively. See Figure 31 for details.
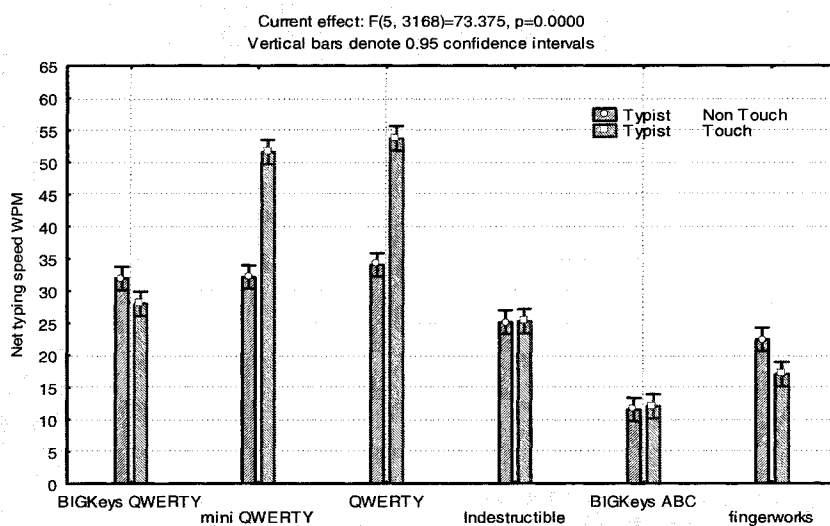
Figure 31. NTS by typing style and input device

For the interaction of age category and typing style similar results than in GTS were found for NTS. For age category 1 there was no significant difference found in mean NTS between touch typists and non touch typists. For age category 2 there was a difference of 4.9 wpm in mean NTS and for age category 3 a difference of 11 wpm in mean NTS between touch typists and non touch typists. See Figure 32 for detailed results.

Current effect: F(2, 3168)=33.445, p=.00000
Vertical bars denote 0.95 confidence intervals
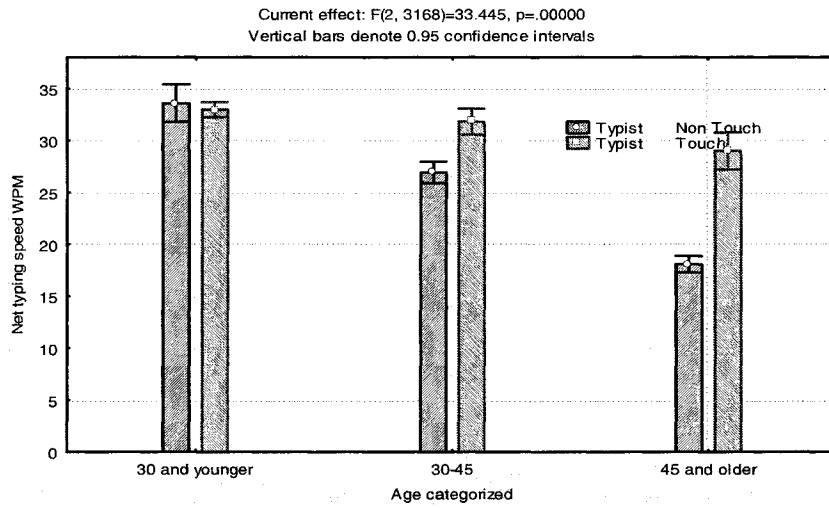
Figure 32. NTS by typing style and age category

For the interaction between age category and device used the same results than those for GTS were found for NTS. Age category 1 performed best followed by age category 2 and age category 3 for most devices. The only two exceptions are the indestructible keyboard which had no significant difference between age categories one and two which both had a higher mean NTS than category 3 and the BIG Keys ABC keyboard that had no significant difference between any age categories (Figure 33).

Current effect: F(10, 3168)=5.7106, p=.00000
Vertical bars denote 0.95 confidence intervals
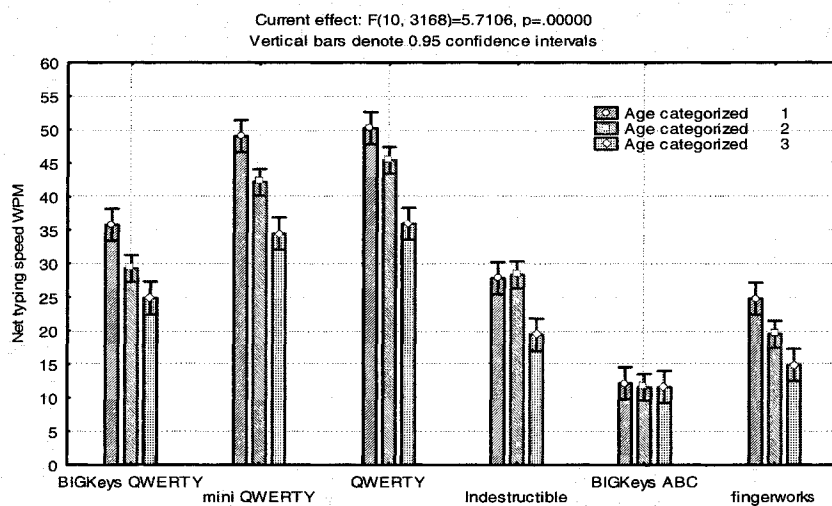
Figure 33. NTS by age category and input device

When just looking at the 'non touch' typists there was a significant difference found between different levels of typing experience (F= 348.43, p=0.0001) and for the interaction of typing experience and input device (F= 10.50, p<0.0001) used (Figure 34).
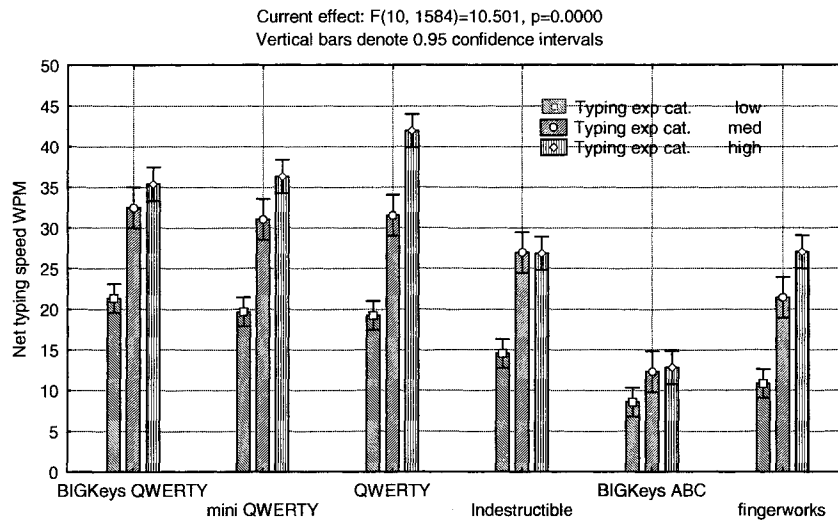
Figure 34. NTS by typing exp. and input device

4.4 Post test questionnaire

Besides the objective performance measurements the subjective rating of usability for the tested devices was evaluated. The participants in this study were first asked to rate the usability of each device in the tested environment and than pick which device they would prefer to use. The rating scale was from 1 to 9 with 1 being the most difficult to use and 9 being the easiest to use. For analytical purposes the scores from 1 to 3 were combined as a low rating 4-6 as a medium rating and 7-9 as a high rating.

A Kruskal-Wallis nonparametric median test found a significant difference between the responses for the different devices (Chi-Square=51.24, df=5, p<0.0001).

A multiple comparison found several significant differences between the tested devices. The detailed results of the multiple comparison analysis are shown in Table 3.

	BIG QWERTY	BIG ABC	QWERTY	Mini QWERTY	Indestructible	Fingerworks
BIG QWERTY	-	0.0009	0.6856	1.0000	0.3813	0.0392
BIG ABC	0.0009	-	0.0001	0.0001	1.0000	1.0000
QWERTY	0.6856	0.0001	-	1.0000	0.0003	0.0001
Mini QWERTY	1.0000	0.0001	1.0000	-	0.0053	0.0002
Indestructible	0.3819	1.0000	0.0003	0.0053	-	1.0000
Fingerworks	0.0392	1.0000	0.0001	0.0002	1.0000	-

Table 3. Multiple comparison p values

The standard QWERTY keyboard got overall the best ratings with an average score of 7.28. It had 14 participants rating it high with three medium ratings and only one low rating. The mini QWERTY had the second best average with a score of 6.61. It had twelve high ratings with four mediums and two low ratings. The third best rated device was the BIG Key QWERTY keyboard with an average rating of 5.39 or seven high ratings and eleven mediums. The indestructible keyboard had an average score of 3.39 with one high rating seven medium ratings and ten low. The next best was the Fingerworks keyboard with an average of 2.83 or seven medium and eleven low. The worst rating was given to the BIG Keys ABC with an average rating of 2.11 or 2 medium ratings and 16 low ratings. The detailed results of the subjective evaluation of the tested devices are shown in Figure 35 and Table 4.
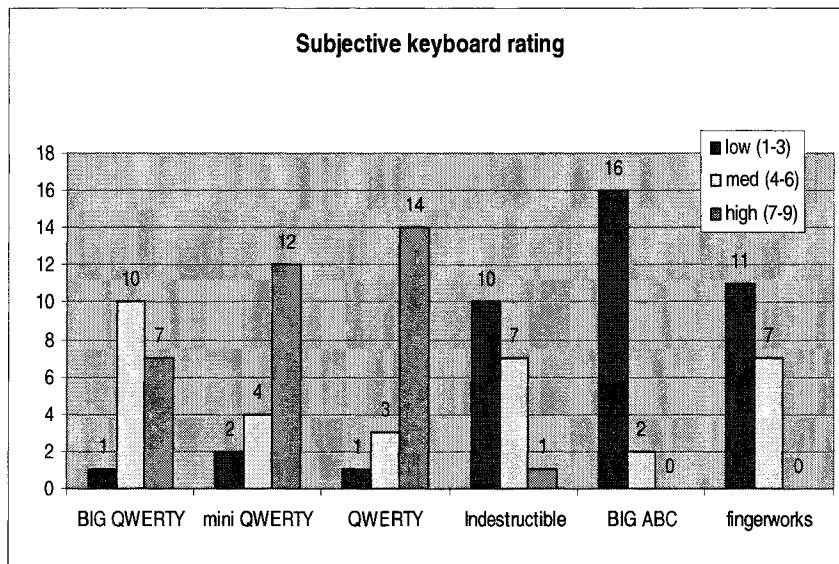
Figure 35. Subjective keyboard rating

	BIG QWERTY	mini QWERTY	QWERTY	Indestructible	BIG ABC	Fingerworks
low (1-3)	1	2	1	10	16	11
med (4-6)	10	4	3	7	2	7
high (7-9)	7	12	14	1	0	0
Average	5.39	6.61	7.28	3.39	2.11	2.83

Table 4. Subjective keyboard rating

Besides rating the usability of all devices the participants were also asked to choose the device they would prefer to use in the tested environment. Overall the standard QWERTY keyboard was the favorite device and was chosen by nine participants, the second favorite was the mini QWERTY with six votes followed by the BIG Key QWERTY with 3 votes. None of the other tested devices were chosen by any of the participants as their favorite device. When braking down the selection of the favorite device by typing style the following results were found. Of the nine touch typists five voted for the standard QWERTY and four for the mini QWERTY. Amongst the non touch typists the standard QWERTY keyboard was

still the most favorite with four votes followed by the BIG Key QWERTY with three votes and the mini QWERTY with two votes. The detailed results are shown in Figure 36 and Figure 37.

Favorite keyboard
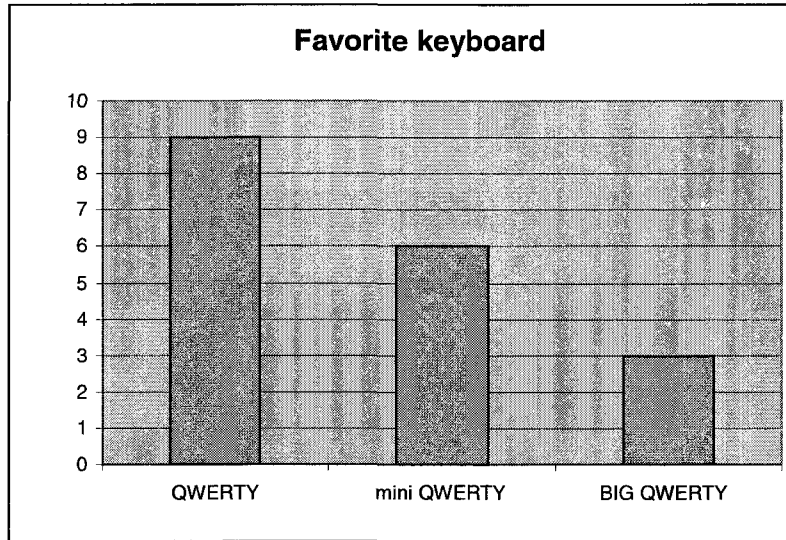
Figure 36. Subjective keyboard rating overall

Favorite keyboard by typing style
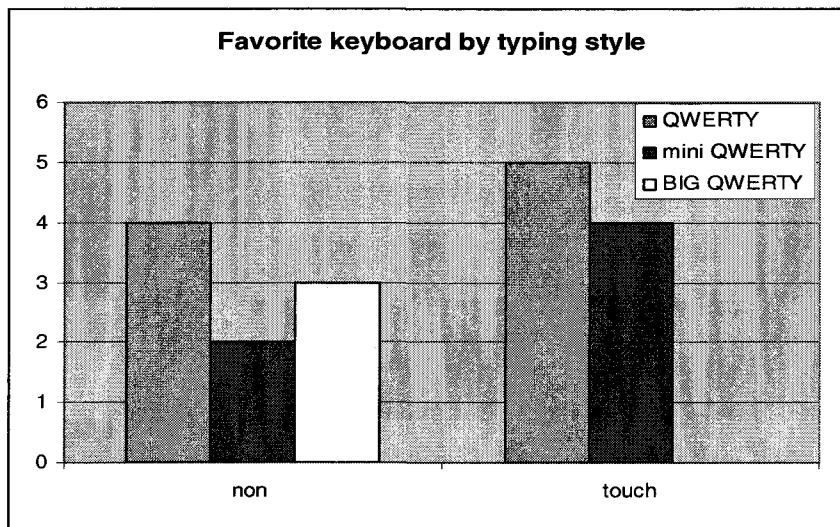
Figure 37. Subjective keyboard rating based on typing style

5.0 DISCUSSION

To determine which of the tested devices provides the best usability in the tested environment the performance a user can achieve with it and the subjective user evaluation have to be considered.

The evaluation of the performance measurements showed that as expected all devices allowed a significantly better performance in the stationary tractor compared to the moving tractor for all three of the performance indicators GTS, NTS and accuracy.

It was also shown that overall as expected the touch typists had a better performance in GTS and NTS compares to the non touch typists while there was no significant difference in accuracy between the typing styles. The typing speeds found in this study for the non touch typists using the standard QWERTY keyboard were slightly higher than the average speed for hunt and peck typing on a standard keyboard fund by Wikund et. al. (1987). This can be explained by the fact that most of the non touch typists in this study had some typing experience and utilized more than two fingers.

For the different age categories in GTS and the NTS the best performance was found for age category one (30 and younger) followed by category two (30-45) and age category three (45 and older). With the differences in performance in GTS and NTS between the different age categories being more eminent for the non touch typists than it was for the touch typists. This can be contributed to the fact that the younger participants tent to be the more experienced typists and the more frequent computer users.

When looking at the accuracy as a performance measurement the order was the opposite. Age category three had the best performance followed by age category two and age category one which was most likely caused by the established trade off between accuracy and speed (MacKenzie and Soukoreff 2002).

It has to be mentioned that the distribution of typing styles and typing experience amongst the different age groups was not equal. Six of the participants in age group one were touch typists with only one non touch typist in that age group. In age group two there were two touch typists and three non touch typists and in age group three there were five non touch typists and only one touch typist. Besides this all of the participants that were touch typist indicated a high level of typing experience while there was a variety of experience levels amongst the non touch typists. A Kendall Tau correlation analysis showed a statistically significant correlation between Typing style and typing experience of 0.67, a statistically significant correlation between typing style and age category of - 0.55 and a statistically significant correlation between typing experience and age category of -0.45. Because of the correlation between these three factors the differences caused in the performance measurement by one is also influenced by the other factors.

Concerning the influence of the tested devices on the performance the overall best performance in GTS and NTS was found when using the standard QWERTY keyboard which also ranked second best in the overall accuracy. It was followed in both typing speed and accuracy by the mini QWERTY keyboard and the BIG Keys QWERTY keyboard.

The BIG Key ABC keyboard was the device that allowed the highest accuracy ratings but the lowest performance in typing speeds. The poor performance of the ABC keyboard in typing speed was most likely caused by the fact that none of the participants had used an ABC layout before and all participants used this keyboard in a "hunt and peck" typing style. This caused a low performance in GTS and NTS while allowing a high accuracy. When using the ABC keyboard there was also no significant difference found between the participants with different typing styles, age categories or typing experience levels.

It was also shown that the typing style, the typing experience and the age in combination with the device used had a significant influence on the performance the participants achieved with the tested devices.

Besides the high performance in accuracy for the BIG Key ABC the touch typists and the non touch typists achieved the best performance in GTS, NTS and accuracy when using the standard QWERTY, the mini QWERTY and the BIG Key QWERTY keyboard.

For non touch typists there was no significant difference between these three devices while there was a significant difference found in all three performance measurements for the touch typists. The touch typists had the best performance with the standard Keyboard followed by the mini QWERTY and the Big Key QWERTY.

Besides that the touch typists were able to outperform the non touch typists in all thee performance measurements when using the standard QWERTY and the mini QWERTY. When using the BIG Key QWERTY keyboard the non touch typists had a better performance than the touch typists in all three performance measurements.

When considering the influence of the different age categories on the performance it was shown that age category 1 (30 and younger) had the best performance in GTS and NTS followed by age category 2 (30-45) and age category 3 (45 an older) when using all devices other than BIG Key ABC and the indestructible keyboard. When using the indestructible keyboard there was no significant difference between categories 1 and 2 with both of them performing better than category 3.When using the Big Key ABC keyboard there was no significant difference between any of the 3 age groups.

Besides that there was no difference amongst the age categories in which device allowed the best performance in GTS and NTS. All 3 age categories had the best performance when using the standard QWERTY followed by the mini QWERTY, the BIG Key QWERTY, the indestructible, the finger works and the BIG Key ABC keyboard.

Only looking at the non touch typists there was a significant difference in NTS and GTS between different experience levels, with the order in performance being according to the experience levels. For all experience levels the three devices that allowed the best performance were the standard QWERTY, the mini QWERTY and the BIG Key QWERTY

keyboard. While there was no significant difference between those three keyboards for the participants with medium and low typing experience the participants with high typing experience had a significantly better performance when using the standard QWERTY.

In the subjective ranking the standard QWERTY keyboard was raked highest followed by the mini QWERTY and the BIG Key QWERTY. Besides having the highest averages these three keyboards also had most of the participants rate them high or medium and only had two or less low ratings. The standard QWERTY was also the device that was picked as the favorite keyboard by the majority of the participants. Both the majority of the touch typists and the non touch typists picked this device as their favorite. The mini QWERTY was the second favorite amongst the touch typists and the third favorite amongst the non touch typists which picked the BIG Key QWERTY as their second. This is consistent with Cerney et. al. (2004) where the participants also preferred the standard QWERTY keyboard.

The indestructible, Fingerworks and BIG Key ABC keyboards had only a low average and only 1 participant gave the indestructible a high rating with no high ratings for the other two keyboards. Those keyboards also weren't picked as the favorite keyboard by any of the participants.

6.0 CONCLUSION

The goal of this study was to determine which of the tested keyboards provided the best usability with standard windows applications in a moving off-road vehicle.

The results of the objective performance measurements showed that the standard QWERTY and the mini QWERTY had the overall best performance across the different user groups and were also the two devices that received the highest ratings in subjective user evaluation.

Both devices are identical in most of the key specifications an represent the keyboard characteristics like layout, key size key spacing and key activation force of the standard keyboard the majority of computer users are familiar with.

None of the devices that had different characteristics were able to provide a better usability in the tested environment. In the scope of this study it was not possible to give the participants the same amount of exposure they had with the standard keyboard with all tested devices which might have influenced their performance and the subjective rating.

When looking at the overall population though a bias towards the standard keyboard can be expected. Especially for touch typists who rely on their motor memory to utilize a keyboard. Using different keyboard characteristics would most certainly mean a decrease in performance unless the same motor memory is gained with the different layout (MacKenzie and Soukoreff 2002). Besides that not even the non touch typists who don't rely on motor memory and therefore are less biased towards the standard keyboard were able to have a significant increase in performance when using on of the other keyboards.

Out of the devices considered in this study both the mini QWERTY and the standard QWERTY keyboard provide good usability and seem to be well suited to be used in the tested environment. To make definite decision which of the two is better suited for an actual application the specific circumstances of the situation have to be considered. Future studies have to show how factors like space constrains or which pointing device is used in combination with the keyboard have an influence on what is best in a certain situation.

REFERENCES

Afifi, A., Clark, V.A., & May, S., (2004) Computer-aided multivariate analysis, Chapman & Hall/CRC, Boca Raton.

AS2708-2001 (2001) Australian Standard: Keyboarding speed tests, Standards Australia International 2001

Cerney M.M.,Mila, B.D.,Hill, L.C.,(2004) Comparison of mobile text entry methods, Proceedings of the Human Factors and Ergonomic Society 48th annual meeting 2004

Fitts, P. M., (1954) The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 47 (1954), 381-391

Hochberg, Y., Tamhane, A.C., (1987) Multiple comparison procedures. Wiley, New York

ISO 9241-4 , (1998) Ergonomic requirements for office work with visual display terminals, Keyboard requirements

ISO 9241-11, (1998) Ergonomic requirements for office work with visual display terminals, Guidance on usability

Jeffries, R., Miller, J.R., Wharton, C., Uyeda, K.M., (1991) User interface evaluation in the real world: a comparison of four techniques. Association for Computing Machinery, 1991

Lueder, R & Grant, C., (1997) Alternative Keyboard Designs written for Ergonomics Branch the National Institute of Occupational Safety & Health

MacKenzie, I. S., (1991) Fitts' law as a performance model in human-computer interaction. Doctoral dissertation. University of Toronto: Toronto, Ontario, Canada.

MacKenzie, I.S., (2005) Within-subject vs. between-subjects design :which to use?. York University: Toronto, Ontario, Canada.

MacKenzie, I.S., Nonnecke, B. Riddersma, S., McQueeen, C. & Meltz, M.,(1994) Alphanumeric entry on pen-based computers. International Journal of Human-Computer Studies, 41, 775-792.

MacKenzie, I.S., Sourkoreff, R.W., (2002) Text entry for mobile computing: Models and methods, theory and practice. Human-Computer Interaction, 17, 147-198.

MacKenzie, I.S., Sourkoreff, R.W., (2003) Phrase sets for evaluating text entry techniques. Extended Abstract of the ACM Conference on Human Factorsin Computing Systems- CHI 2003, pp. 754-755

MacKenzie, I.S., Zhang, S.X. & Sourkoreff, R.W., (1999) Text entry using soft keyboards. Behavior & Information Technology, 18, 235-244

MacKenzie, I.S.& Zhang, S.X., (1999) The design and evaluation of a high performance soft keyboard. Proceedings of the ACM Conference on Human Factors in Computing Systems- CHI 99, pp. 25-31

MacKenzie, I.S.& Zhang, S.X., (2001) An empirical investigation of the novice experience with soft keyboards. Behavior& Information Technology, 20,411-418

Ober, B.S., (1983) The difficulty level of typewritten copy in industry. Delta Pi Epsilon Journal, 1983, vol. 25, no 1, pp 1-8

Pavlovych A. and Stuerzlinger W., An analysis of novice text entry performance on large interactive wall surfaces, Dept. of Computer Science, York University Toronto

Pavlovych A. and Stuerzlinger W., Less Tap: A fast and easy to learn text input technique for phones. , Dept. of Computer Science, York University Toronto

Sears, A., (1991) Improving touch screen keyboards: design issues and a comparison with other devices. Interacting with Computers vol. 3, 253-269

Sears A., (2003) Data Entry for mobile devices using soft keyboards: understanding the effects of keyboard size and user task. International journal of human-computer interaction, 16 (2), 163-184

Shneiderman, B., (1998) Designing the user interface: strategies for effective human-computer-interaction, 3rd ed. Addison Wesley Longman, Inc. 1998

Silverberg, M., MacKenzie, I.S., & Korhonen, P., (2000) Predicting text entry speed on mobile phones. Proceedings of the ACM Conference on Human Factors in Computing Systems – CHI 2000, pp. 9-16. New York: ACM.

Perry, D.J., (1968) An analytical comparison of the relative word-combination frequency of business correspondence with phrase frequency of selected shorthand textbooks, vols I and II, University of North Dakota 1968

Sears, A., Revis, D., Swatski, J., Crittenden, R., Schneiderman, B., (1993) Investigating touchscreen typing: the effect of keyboard size on typing speed. Behaviour & Information Technology, vol. 12, 17-22

Sears, A.& Ying, Z., (2003) Data entry for mobile devices using soft keyboards. International Journal of Human-Computer Interaction 16(2), 163-184

Soukoreff, W., & MacKenzie, I. S., (2004).Recent development in text-entry error rate measurement, Extended abstract of the ACM conference on Human Factors in Computing Systems- CHI 2004

Wikund, M.E., Dumas, J.S., &Hoffmann, L.R. (1987). Optimizing a portable terminal keyboard for combined one-handed and two-handed use. Proceedings of the Human Factor Society—31st Annual meeting—1987(pp. 585-589). Santa Monica, CA: Human Factor Society

West, L.J., (1968) The vocabulary of instructional materials for typing and stenographic training-Research findings and implications. Delta Pi Epsilon Journal, 1968, vol. 10, no 3, pp 13-25

Zhai, S., (2000) An exploration of quantitative techniques for virtual keyboard Design ACM Symposium on User Interface Software and Technology (UIST 2000) pp 119-128

Evaluation of input devices with an office application in a moving off-road environment

A paper to be submitted to *Applied Ergonomics*

Thorsten Baldus

ABSTRACT

Based on earlier studies, three user interfaces to be used with standard Windows™ application were installed in a tractor. An experiment was conducted to evaluate their usability in the tested environment. The objective performance compared to an office environment as well as the subjective user rating of their usability and performance was evaluated. Out of the tested interfaces the mouse/mini QWERTY and touch screen stylus/standard QWERTY interface showed the best performance and highest user rating and therefore provide the highest usability to the user.

1.0 INTRODUCTION

With computers taking over more control functions in modern vehicles it becomes important to create better computer interfaces in those environments.

The first step of implementing a useful computer interface in a vehicle is to evaluate its use for simple vehicle control functions. In the future it is possible that personal computers will become available in vehicles that will allow the operator to perform all kinds of computer tasks while an automated system is controlling the vehicle. While these types of systems are still in the development phase for on-road vehicles, semi-autonomous controls are already available for use in tractors. One example for such a system is the John Deere's Green-Star Auto-Trac System™, which is a hands free steering system for straight-line driving applications.

When using a tractor that is equipped with one of these systems, the work load of the operator is reduced and the operator has the opportunity for performance of other tasks while doing fieldwork. For this reason many people believe a tractor cab could become a mobile office in the near future in which the operator could uses a personal computer to take care of all types of office tasks while working in the field. With the advances in mobile technology, it is easy to imagine that in the near future a farmer could browse the Internet to find weather data or market prices or sell a harvest while sitting in the tractor doing fieldwork. To enable the operator to do this kind of work it is necessary to provide a usable computer interface that includes a pointing device and an alphanumeric input device. This interface has to enable the operator to use the office application in the moving tractor efficiently and with satisfaction or he would have little motivation to utilize such a system.

The ISO 9241, part 11 (1998) standard provides a guide for the evaluation of the usability of an office computer terminal. It states that in order to determine the level of usability achieved, it is necessary to measure the performance and the satisfaction of users working with a system. Measurement of usability is particularly important in view of the complexity of the interactions between the user, the task characteristics and the other elements of the context of use. A product can have significantly different levels of usability when used in different contexts. The standard also states that measures of the performance and satisfaction of the user can provide a basis for the comparison of the relative usability of products with different technical characteristics which are used in the same context (ISO 9241, part 11, 1998).

The goal of this study was to identify the interface that provides the best usability with standard applications in a moving environment and to identify the factors that influence the usability of the tested devices. In two previous studies a variety of commercially available pointing and alphanumeric input devices were evaluated to identify the devices that provide the best usability with a standard application in a moving off-road environment. These only focused on the single devices and didn't evaluate the usability of the whole interface. In the course of these studies the standard and mini QWERTY keyboard were identified as the

alphanumeric input devices and the mouse and the touch screen as the pointing devices that are best suited for the tested environment.

Several studies (MacKenzie 1991, Bohan et.al. 2003, Cerney et.al 2004, Pavlovych and Stuerzlinger 2004) established the dominance of the mouse and the standard QWERTY keyboard as the interface devices that provide the highest usability in an office environment. It is expected that due to factors like movement and vibration the performance achievable with a computer system in a mobile off-road environment will be lower than the performance achievable in an office environment. To determine the performance a user can achieve with the tested interfaces both the absolute performance they can achieve and the relative performance of each participant compared to an office environment will be evaluated. To determine if the use of a computer interface in a mobile off-road environment is worthwhile the overall perceived performance of the user and the subjective perception of the usability also have to be evaluated. If the productivity and usability a user can achieve with such a system are not satisfying to the user he will get frustrated and have little reason to utilize the system.

2.0 METHOD

An experiment was designed to evaluate the performance of a computer interface in a moving tractor as compared to the performance of a standard office interface. The interface devices tested in this study were the ones that showed the best results in the earlier studies and were therefore expected to have the highest potential to provide a good usability in the tested environment. The devices identified in the earlier study include the mouse and the touch screen for pointing and the standard and mini QWERTY keyboard as alphanumeric input devices. Since the size of the human finger and the lack of sensing precision can make precise touch screen interaction difficult, this study will also evaluate the usability and performance and user can achieve with the touch screen utilizing a stylus. All interfaces evaluated in this study consisted of a pointing device and an alphanumeric input device.

The interfaces evaluated were:

Interface 1:	Mouse/mini QWERTY
Interface 2:	Touch screen with finger/ standard QWERTY
Interface 3:	Touch screen with stylus/ standard QWERTY

Both of the keyboards included in this study are very similar in their operation and most of their features. It was therefore decided to only test specific combinations of keyboards and pointing devices that make sense for the tested environment. Since for the operation of a mouse some space on the work surface is required the mouse was tested in combination with the mini QWERTY keyboard. The operation of a touch screen on the other hand doesn't require any additional space and will therefore be tested in combination with a standard QWERTY keyboard. The characteristics of the two tested keyboards are shown in Table 1. For the further progress of the study the interfaces will be identified by the pointing device used since this is unique for each interface.

Keyboard	Switch actuation		Key travel	Key size	Key spacing	Size
Standard QWERTY	55 g	0.539 N	4mm	13mm x 14mm	19mm	160x455 mm
mini QWERTY	55 g	0.539 N	4mm	13mm x 14mm	19mm	120x290 mm

Table 1. Keyboard characteristics

To get some benchmark data on the performance each participant is able to achieve with a computer interface they were asked to perform one session completing a set of tasks in an office workplace before performing the same set of tasks in a moving tractor. For the office interface session the devices the participants normally use for their day to day computer work were used.

2.1 Participants

The study was conducted with 12 volunteers recruited by John Deere. The participants were all Owner/Operators of John Deere tractors between 26 and 57 years of age (average age 42 years) who were also required to have some experience in operating desktop computers. Four of the participants were left handed with eight being right handed. Since as in an earlier study common phrases in the English language were used to evaluate text entry only participants with English as their first language were considered.

The participants were also asked to evaluate their prior experience with the tested input devices and indicate their average computer usage. Five of the participants were touch typists and seven were non touch typists. When asked about their average weekly use of a computer one participant indicated an average computer usage of 1 hour or less, two indicated an average computer usage of 1to 4 hours, 5 indicated an average usage of 4 to 10 hours, two subject indicated 10 to 20 hours and 2 indicated to use a computer more than 20 hours in a week.

A Kendall's Tau Correlation analysis was performed for the characteristics of the participants and found a significant correlation between the typing style and the typing experience (0.68), the dominant hand and computer usage (0.52), computer usage and typing style (0.44), computer usage and typing experience (0.44). and between the touch screen and the mouse experience (0.42). The detailed distribution of the test participants amongst the different characteristics can be seen in Table 2.

	Age Category	Typing Exp			Mouse Exp			Touch screen exp.		
		Low LH/RH	Med LH/RH	High LH/RH	Low LH/RH	Med LH/RH	High LH/RH	Low LH/RH	Med LH/RH	High LH/RH
Touch Typists	1	-	1/0	-	-	-	1/0	-	1/0	-
	2	-	-	0/2	-	-	0/2	0/1	0/1	-
	3	-	0/1	1/0	-	0/1	1/0	-	0/1	1/0
Non Touch Typists	1	-	0/1	-	-	0/1	-	-	0/1	-
	2	2/0	0/1	-	-	1/2	0/0	0/1	1/1	-
	3	1/0	0/2	-	-	1/1	0/1	1/1	0/1	-

		Average computer Usage				
		<1h	1-4h	4-10h	10-20h	>20h
Average computer Usage	1	-	1	1	-	-
	2	1	-	2	1	1
	3	-	1	2	1	1

Table 2. Categorization of test participants.

Besides this it has to be mentioned that one additional participant started the study but withdrew from it during the first session because he was not feeling well. It was not possible to determine weather this was caused by participating in the study or if other factors were responsible.

2.2 Apparatus

The test setup was similar to the one in a previous study. The same tractor with an automated steering system was used. Additionally a John Deere GSD 2600 display with integrated touch screen was installed in the tractor. The display had a screen diagonal of 260 mm and a

maximum screen resolution of 640x480 pixels. It was mounted on the front right corner post of the tractor cab with an adjustable arm so that every participant could adjust the position of the screen to their preference (Figure 1). There was also a lab tray (500x300 mm) provided to the participants to serve as a surface for the keyboard and the mouse.
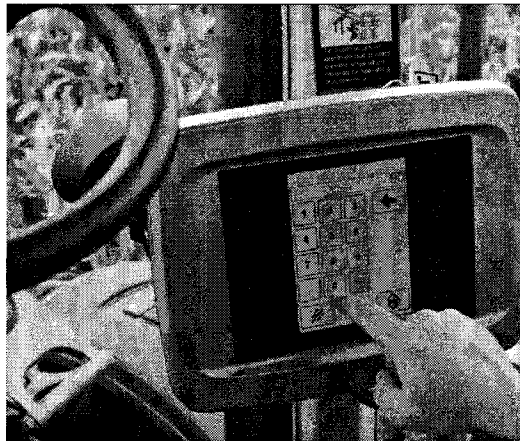
Figure 1. Tractor test setup with GSD 2600

All participants stated that they normally use a standard QWERTY keyboard and a mouse for their day to day computer work. For the office computer workplace a standard DELL QWERTY keyboard and a Microsoft IntelliMouse was provided to all participants. The office test setup is shown in Figure 2.
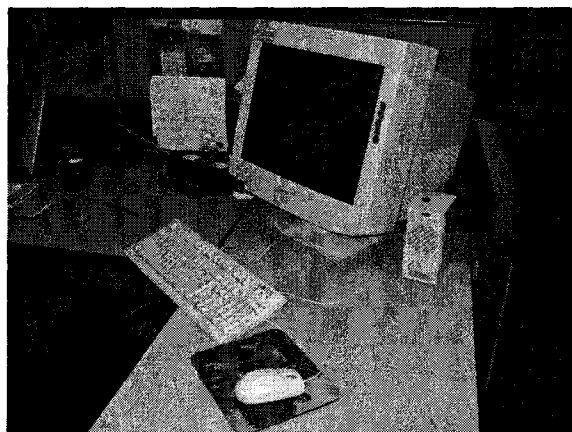
Figure 2. Office test setup

For both work test setups the gain of any pointing devices used was set to the middle value of the standard driver software for pointing device sensitivity for all devices (Douglas at al 1999).

The test software used for the objective performance measurements was a self developed Meta Card application. The software presented a series of tasks to the participants and recorded the performance they could achieve with different devices. All tasks represented different elements a user would find when interacting with a standard application in a windows environment.

Task 1 Toolbar selection and typing:

This task was chosen to simulate the selection of a small target like an icon in a toolbar and the switch from mouse operation to typing and back.
In the first step of this task the participants were asked to select a highlighted button out of a series of 9 buttons. Each button was 25x25 pixels (8.125mm x 8.125mm) big which is equivalent to the size of the icons in the toolbars of windows applications at the tested resolution (Figure 3). When selecting the wrong button a pop up window appears that informs the participant that he did a wrong selection and asks him to try it again after confirming the message. Figure 3 shows a typical screen of this task.
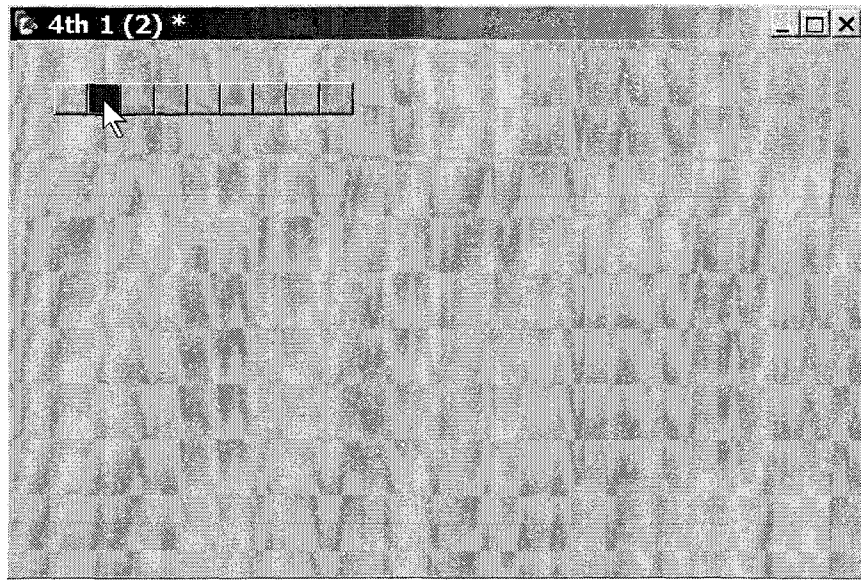
Figure 3. Screenshot of test software task 1 pointing part

After the correct selection of the highlighted button the screen would change and a text
phrase was presented to the participant and asked them to switch to the keyboard and enter
the phrase. The phrases presented were part of the "Phrase Set for Evaluating Text Entry
techniques" by MacKenzie and Soukoreff (2003). The test subjects were instructed to correct
any mistakes they realize and basically use the same procedure for this text entry they would
use when entering text in a standard application. This part of the task was very similar to the
alphanumeric input study done earlier and the same phrase set and test setup was used. To
have a better representation of a real application though the test software didn't indicate
when a typing error occurred and the participants were instructed to check and correct there
errors in the same way they do during normal computer operation. When finished with the
typing the participant had to hit a button with the pointer to confirm the entry. Figure 4 shows
a typical screen for this task.
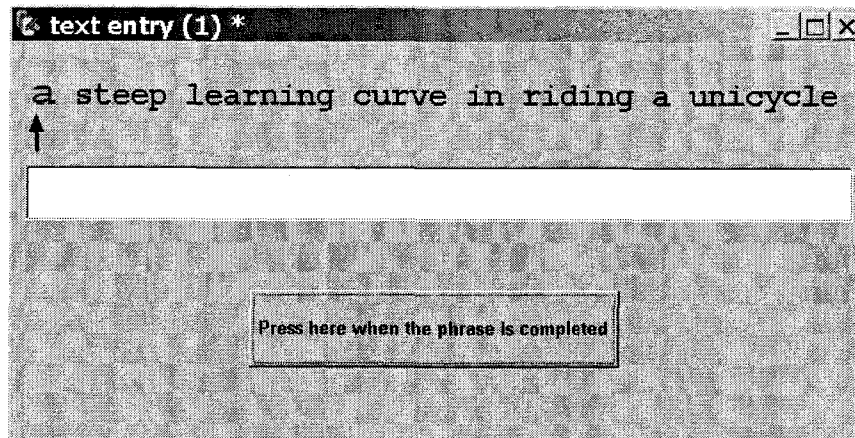
Figure 4. Screenshot of test software task 1 typing part

For the first part of the task the test software recorded time it took for pointer selection, the number of wrong selections in case they occurred and the location of the pointer when a selection was made. For the second part the text entry time, the number of corrections and the characters entered were recorded. Besides that the overall task time and the time for switching from pointing to typing and back was recorded.

Task 2 Pull down/Cascading menu selection:

This task was simulating the use of a cascading pull down menu in a windows application. To complete the task the participants first had to select a specific button out of a menu bar with five buttons. This opened a pull down menu where they had to move the pointer over a colored area which opened the cascade menu where they had to select the colored button to complete the task. If at anytime in the task a wrong button was selected, a pop up window would appear that informed the participants that a wrong selection was made. The participants were asked to confirm the message and then try again to make the correct selection. During the completion of this task the test software recorded the number of wrong selections and the total time it took to complete the task correctly. For these tasks all buttons had a size of 60 x 20 pixels (19.5mm x 6.5mm) which is equivalent of the size of a pull

down/cascade menu button in a standard windows application at the tested resolution. A typical screen for this task is shown in Figure 5.
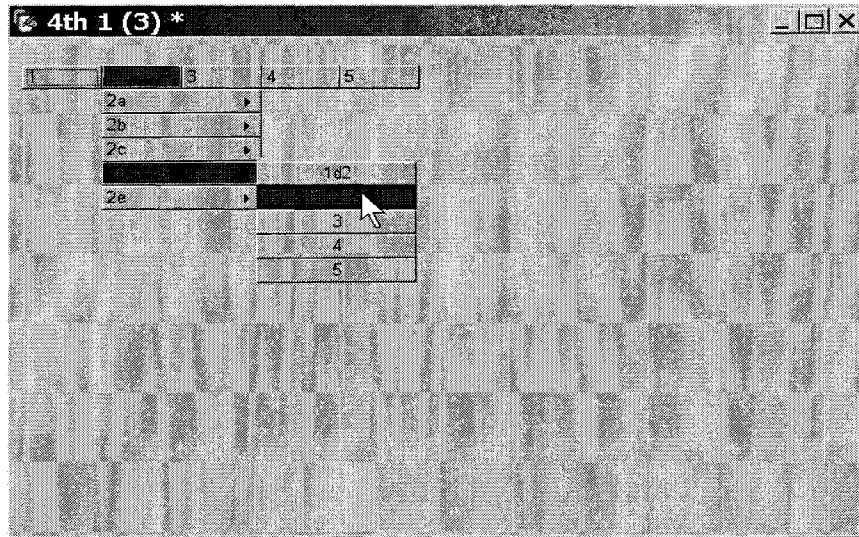
Figure 5. Screenshot test software task 2

Task 3 Scroll bar operation:

The third task was simulating the use of the scroll bar in a windows application. The participants were presented with a number from 1-50 and were than ask to navigate to the same number in a list using a scroll bar. They than had to select the number in the list to complete the task. The test software recorded the total time it took to select the number and if there were any wrong selections made. To make sure the direction of navigation didn't have an influence on the performance the software alternated the starting position on the list and the number that had to be selected. A typical screen for this task is shown in Figure 6.
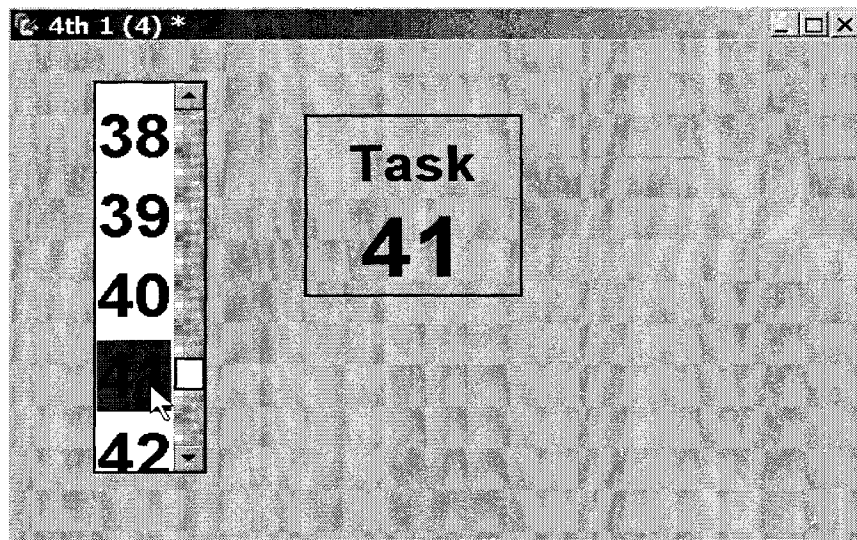
Figure 6. Screenshot test software task 3

Task 4 Pointer drag and drop operation:

This task was simulating the drag and drop operation of an item on a windows desktop. At the beginning of the task the square was located in the center of the screen. The participants were instructed to click on the square and drag it to the frame which was randomly located along the edge of the screen. The color of the frame changed to black to indicate when the square was inside. The size of the square was 40x40 pixels (13 mm x 13 mm) and the target frame was 60x60 pixels (19.5mm x 19.5mm) which is equivalent to the size of an icon on a windows desktop or the virtual frame around an icon for a drag and drop operation at the tested resolution. For this task the test software recorded the total time it took to complete the task and the location of the target square. A typical screen for this task is shown in Figure 7.

Figure 7. Screenshot test software task 4

2.3 Experimental Design

According to the recommendation of MacKenzie (2005) a within subject design with the following independent variables was used for this study:

- Tested interface
  - Interface 1: Mouse/mini QWERTY
  - Interface 2: Touch screen with finger/ standard QWERTY
  - Interface 3: Touch screen with stylus/ standard QWERTY

- Session (office, 3 tractor sessions)

2.4 Procedure

Each participant was asked to perform multiple sessions with each of the tested interfaces. This allowed testing the "immediate usability" of the tested devices as well as the performance possible by a more "experienced" user. It is understood that after a few sessions not all participants will have reached their performance maximum with the tested interfaces but it was possible to get an insight on the performance that a user can achieve with the tested interfaces.

For the first session all participants were ask to perform a series of tasks in an office environment. For this session the devices the participants normally use when working with a standard application were used, which was the mouse and the standard QWERTY keyboard for all of them. At the beginning of the test the participants had a chance to familiarize themselves with the test software and the testing procedure by completing a couple of task sets without any data being recorded. They were than asked to complete 15 task sets with data recording. The data from this session was used as a benchmark of their basic performance with a computer. There were no additional sessions necessary with this setup since it was assumed that all participants had prior experience with the devices used in this environment and the performance they could achieve wouldn't change significantly.

Additionally there were three sessions conducted in a moving tractor. In each of these sessions the participants first were given some time to familiarize themselves with the testes devices by "playing around" with windows and any of the standard applications and they were asked to perform tasks they usually perform when using standard applications. There was no objective performance data measured and the main purpose was to provide the participants with a chance to get a feel for how the devices work and how usable they are with a real application in a moving environment. The participants were also provided with two different stylus options to be used with the touch screen, a pen stylus (Figure 8) and a finger stylus (Figure 9). The subjects had a chance to use both of them with the standard

application and the test software and were than asked to pick there preferred stylus to complete the second part of the in tractor evaluation.
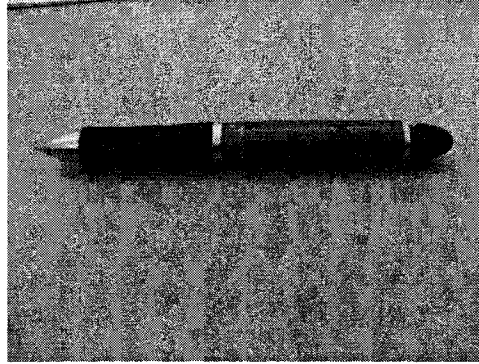
Figure 8. Pen stylus
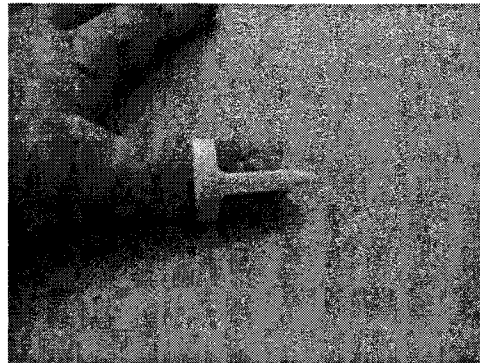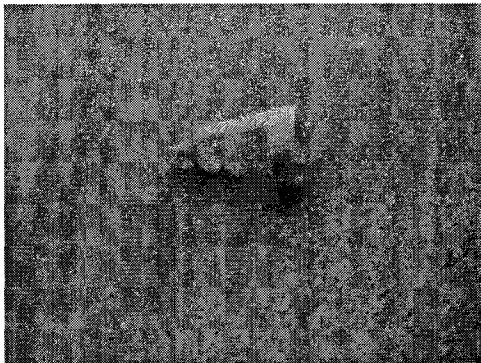
Figure 9. Finger stylus

In the second part of each tractor session the participants were asked to complete a series of tasks described above presented to them by test software. All tasks were presented to the participants in a fixed sequence and the participants had to complete the whole sequence of 4 tasks fifteen times with each device in each session.

To guard against any effect the test order could have on the performance the participants were able to achieve with the tested devices, a balanced design was used that determined the

order in which the interfaces were presented to the participants. The test order details of the balanced design are shown in Appendix D.

After the completion of each tractor session the participants were asked to complete a questionnaire. In the questionnaire they were asked to document their computer experience and experience with the tested devices and evaluate their subjective impression of the usability of the tested devices.

2.5 Moving Vehicle Environment

The test was conducted over the same track and under similar conditions as a previous study. The tractor was traveling on a predetermined path at a constant speed of 7.5 km/h with an automated steering system monitored by the investigator handling all tractor controls. During a previous study acceleration data of the tractor frame and cab was recorded and it was decided by experts at John Deere that the conditions are representative for field work conditions. It was determined that it was not necessary to record any acceleration data for this study since the test conditions were very similar.

3.0 ANALYSIS

The goal of this study was to determine weather the tested interfaces are capable of providing usable interface with an acceptable level of productivity utilizing standard software applications in a moving tractor. The objective performance measurements taken during this study allowed an insight on the influence of different factors on the performance and a comparison between the productivity achieved in the moving tractor and an office workplace. To calculate the relative performance in the tractor compared to an office environment the mean value of the different performance measurements for each tractor session and interface was calculated and compared to the mean value of the performance measurement for each subject in the office session. The performance in the office environment served as a base line and represented 100%. Independent from the measurement values (time, text entry speed

etc.) a relative performance value above 100% represented an increase in performance and values under 100% a decrease in performance compared to the office environment.

The first task presented to the participant was a combination of a pointing and a typing task. Since the participants were presented with different phrases in the typing part the overall task time is not a good measurement to evaluate the performance the participants were able to achieve. This task was divided into different parts for evaluation purposes. To evaluate the performance of the pointer selection the time to do the pointer selection was recorded. Besides that the time it took the participants to switch from pointing to typing and back was recorded. This can be used to evaluate the influence of different factors on the time it takes to switch from pointing to typing and back.

To evaluate the typing part of this task the NTS was calculated according to AS 2708-2001. Since not all aspects of the testing procedures for this study are conforming to AS 2708-2001 the results can not be compared to other studies and should only be used to compare the different devices within this study.

The second, third, and fourth task in the sequence represent different pointer operations and the overall task time was a good measurement to evaluate the performance for the different pointer operations. Because of the distribution of the test participant's characteristics it was not possible to evaluate all interactions of the different factors. Only interactions between factors that were fully represented were considered.

4.0 RESULTS

4.1 Objective performance measurement

<u>**Task 1:**</u>

Selection time

When only looking at the pointing part of the task the "selection time", the time it took to select the highlighted button to start the text entry, is a direct measurement of the pointing performance. For this parameter a total of 1620 measurements were taken with a mean of 2269 ms, a minimum value of 742 ms and a maximum of 14514 ms.

A multivariate GLM analysis found a significant difference between the pointing devices used (F=21.77, p<0.0001), between the age categories (F=7.12, p=0.0008), between the different levels of computer usage (F= 4.47, p=0.001), and between the dominant hand (F=-15.7, p<0.0001). There was no significant difference in click time found between sessions, mouse experience or for any interaction between the factors.

Between the different pointing devices the lowest least square mean selection time was found when using the mouse with 1624 ms followed by the touch screen/stylus with 1931 ms and the touch screen/finger with 2274 ms. For the different age categories there was no significant difference found between category two (30-45 years) and category 3 (> 45years) with a least square mean selection time of 2153 ms and 2108 ms respectively. Age category 1 (< 30 years) had a significantly lower selection time than the other two with a least square mean of 1569 ms.

Concerning the dominant hand the right hand dominant test participants performed better and had a lower least square mean value in selection time with 1667 ms compared to the left hand dominant participants with 2220 ms.

For the computer usage the worst performance of a least square mean selection time of 2700 ms was found for the participants of computer usage category 1 (<1h). The best performance with a least square mean of 1980ms was found for the participants of category 4 (10-20h) (Figure 10).

It has to be mentioned that because of the high correlation (0.52) between the dominant hand and the computer usage amongst the participants in this study it can't be determined which of these two factors actually has the biggest influence on the performance and they have to be viewed together. Because of the distribution of these factors amongst the test participants it was not possible to evaluate if there was a significant difference for the interaction of those two factors.
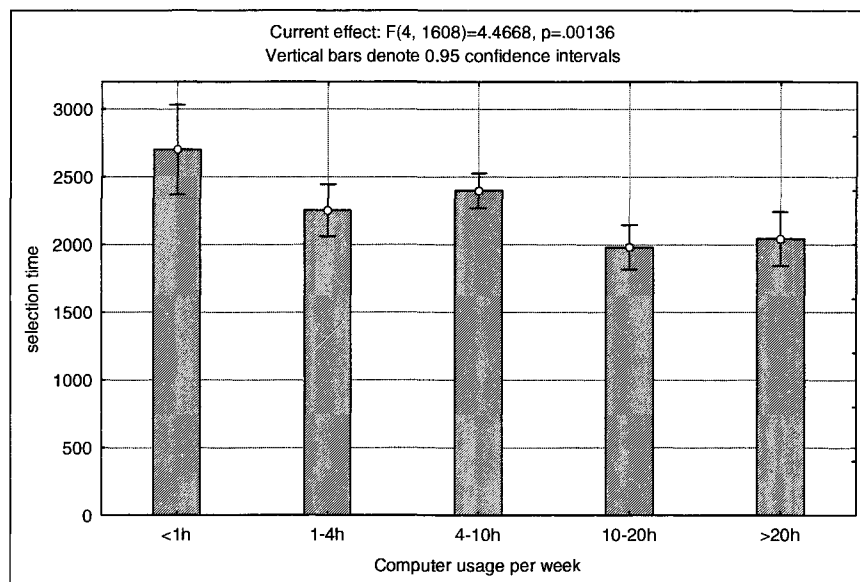
Figure 10. Selection time by average computer usage

For the relative performance in selection time compared to the office workstation 108 samples were taken with a mean value of 62%, a minimum value of 30%, and a maximum value of 97%. A significant difference was found between different pointing devices

(F=7.092, p=0.0014), age categories (F=4.994, p=0.0090), and between different levels of computer usage (F=6.257, p=0.0002).

Between the different pointing devices the best relative performance in selection time was found for the mouse with 69.7% followed by the touch screen/stylus with 62.0 % and the touch screen/finger with 53.7%.

Between the different age categories age category 3 had the highest relative performance with 62.8 % followed by category 2 with 57.2% and category 1 with 50.2 %.

For the different levels of computer usage the best performance was found for category 4 (10-20 h) with a least square mean of 69.1% followed by category 2 (1-4 h) with 67.5 %, category 3 (4-10 h) with 61.5%, category 1 (<1 h) with 55.8%, and category 5 (>20 h) with 52.9 %.

Change time 1

Change 1 represents the time it took the participants to switch from the pointing part of task 1 to the typing part. This part of the task lasted from the completion of the pointer selection to the input of the first character. For the absolute performance in change 1 a significant difference was found between age categories (F=26.279, p<0.0001), between the dominant hand (F=144.40, p<0.0001), the computer usage (F=28.699, p<0.0001) and between typing styles (F=24.221, p<0.0001).

Between the different age categories the best performance was found for category 1 (<30) with a least square mean change time of 2027 ms followed by category 3 (>45) with 2283 ms and category 2 (30-45) with 2510 ms.

Between the dominant hands a significantly better performance was found for the right handed participants with a least square mean change 1 time of 1851 ms compared to the left handed participants with 2696 ms.

Between the different levels of computer usage the participants with the higher levels of computer usage had a significantly higher performance in change 1 time than the participants with lower computer usage. The best performance was found for category 5 (>20 h) with a least square mean change 1 time of 1639 ms followed by category 4 (10-20h) with 2161 ms, category 3 (4-10h) with 2266 ms, category 2 (1-4h) with 2444 ms, and category 1 (<1h) with 2893 ms.

Between the different typing styles the touch typists had a significantly better performance in change 1 time with a least square mean value of 2105 ms compared to the non touch typists with 2442 ms.

For the relative performance in change 1 time compared to the office environment the sample size was 108 and the mean value found was 95%, a minimum value of 49% and a maximum value of 140%. A significant difference was found between the different age categories (F=13.64, p<0.0001), and levels of computer usage (F=12.84, p<0.0001).
There was no significant difference found in relative performance between age category 2 (30-45) and age category 3 (>45) which had a least square mean of 94.6% and 96.6 % respectively. Age category 1 (<30) had a significantly lower relative performance in change 1 time than the other two with a least square mean of 68.5 %.

Between the different levels of computer usage the best performance was found for category 5 (>20h) with 109.6 % followed by category 2 (1-4h) with 106.6 %, category 4 (10-20h) with 96.9%, category 1 (<1h) with 89.2 %, and category 3 (4-10h) with 84.9 %.

Net typing speed NTS

Concerning the typing part of task 1 the net typing speed NTS was the best performance measurement. The sample size for this measurement was 1620 with a mean value of 28.045 wpm, a standard deviation of 11.58 wpm and extreme values of 3.728 wpm and 74.29 wpm. There was a significant difference in Net Typing Speed (NTS) found between the typing

styles (F=128.31, p<0.0001), the typing experience (F=45.26, p<0.0001) and the computer usage of the participants (F=74.61, p<0.0001). There was no significant difference found for the keyboard used, the age category, and the dominant hand or for the interaction of any factors.

As expected there was a higher least square mean NTS found for the touch typists (32.0 wpm) compared to the non touch typists (21.5 wpm).

When looking at the influence of the typing experience on the NTS we found as expected a higher NTS for the participants with more typing experience. Least square mean NTS found were between 32.6 wpm for the participants that indicated high typing experience, 28.2 wpm for the participants with medium typing experience, and 23.6 wpm for the participants with low typing experience.

Since there was a high correlation between typing style and typing experience (0.68) found for the participants in this study it is not sure which of these two factors has more influence on the NTS and they have to be viewed together.

For the influence of the average computer usage of the participants on the typing performance there was as expected a higher mean typing speed found for the participants with more frequent computer usage. The mean NTS was between 21.5 wpm for the participants of average computer usage category 1 (<1h) and 35.4 wpm for the participants of average computer usage category 5 (>20h) (Figure 11).
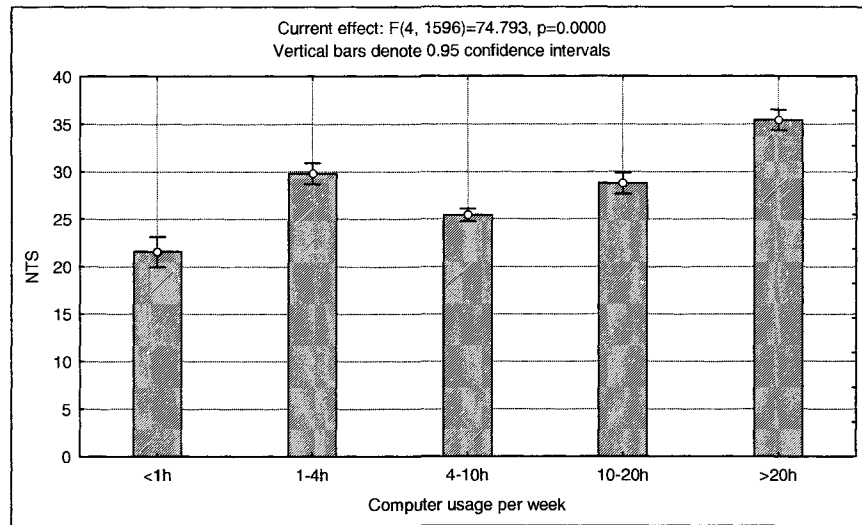
Figure 11. NTS by average computer usage

When looking at the relative performance in NTS a mean value of 86% with a minimum value of 45% and a maximum value of 122% was found for a sample size if 108 measurements. A significant difference in the relative performance for the NTS was found between age categories ($F=33.359$, $p<0.0001$), different levels of computer usage ($F=25.822$; $p<0.0001$), dominant hand of the participant ($F=8.741$, $p=0.0039$), and different levels of typing experience ($F=3.16$, $p=0.0470$).

For the different age categories the best performance was found for age category 1 with a least square mean value of 101.0% followed by category 2 with 91.9 % and category 3 with 78.0 %.

Between the different levels of computer usage the best performance was found for category 4 (10-20h) with a least square mean value of 95.4 %followed by category 5 (>20h) with 93.8 %, category 2 (1-4h) with 87.8%, category 1 (<1h) with 82.8 %, and category 3 (4-10h) with 78.0 %.

Between the dominant hands the higher performance was achieved by the left handed participants with a least square mean value of 96.6% compared to 84.0% by the right handed participants.

Between the different levels of typing experience there was no significant difference found between the individuals with medium and high experience and the least square mean values were 94.0 % and 92.1%. The subjects with low typing experience had a significantly lower performance with a lease square mean value of 84.9 %.

Change time 2

Change time 2 represents the time it took the participants to switch back from typing to the pointer operation. This part of the task lasted from the entry of the last character to the confirmation of the text entry by pointer selection of the confirmation button. The sample size for this measurement was 1620 with a mean of 1942 ms, a minimum value of 595 ms and a maximum of 8887 ms. For the performance in change time 2 a significant difference was found between pointing devices (F=88.172, p<0.0001), different levels of computer usage (F=27.425, p<0.0001), and between typing styles (F=9.428, p=0.0022).

Between the different interfaces there was no significant difference found for the touch screen/stylus with least square mean change time of 1846 ms and the touch screen finger with 1750 ms. The mouse had a significantly lower performance than the other two interfaces and a least square mean change time 2 of 2393 ms.

Between the different typing styles the touch typists had a significantly better performance with a least square mean change time 2 of 1788 compared to the non touch typists with 2205 ms.

For the different levels of computer usage the best performance was found for category 5(>20h) with 1534 ms followed by category 2 (1-4 h) with 1608 ms, category 4 (10-20h) with 1857 ms, category 3 (4-10h) with 2074 ms and category 1 (<1h) with 2911 ms.

For the relative performance in change time 2 a significant difference was found between different interfaces (F=9.708, p=0.0002) and for different levels of computer usage (F=7.140, p<0.0001). The sample size for this measurement was 108 with a mean value of 98%, a minimum of 40% and a maximum value of 142%.

For the different interface devices there was no significant difference found between the touch screen/finger and the touch screen/stylus. The touch screen finger had a least square mean performance of 101.1 % compared to the touch screen stylus with 97.9%. The performance of the mouse was significantly lower with a least square mean value of 78.6%.

For the different levels of computer usage there was no significant difference found between category 2 (1-4h) with a least square mean of 103.0%, category 3 (4-10h) with 99.9 %, category 4 (10-20h) with 99.4% and category 5 (>20h) with 96.9%. Category 1 (<1h) had a significantly lower performance than all other categories with a least square mean value of 63.5%.

Task 2:

For the second task the total task time T2 was the best indicator of the performance. It had a mean value of 4813 ms with a standard deviation of 2483 ms. Amongst the 1620 measurements the minimum value for T2 was 1769 ms with a maximum value of 22614 ms. A significant difference in total task time was found between devices (F=6.015, p=0.002), between different levels of computer usage (F=6.242, p<0.0001), between the dominant hand of the participants (F=44.9, p<0.0001), between age categories (F=13.383, p<0.0001), between different experience levels with the touch screen (F=26.059, p<0.0001), and between the buttons the participants had to select in the pull down menu (F=5.633, p=0.0002).

Between the different devices the best performance was found for the touch screen/stylus with a least square mean task time of 4085 ms followed by the mouse with 4239 ms and the touch screen/finger with 4561 ms.

Concerning the dominant hand the right hand dominant test participants performed better and had a lower least square mean value in selection time with 3632 ms compared to the left hand dominant participants with 4957 ms.

Between the different levels of computer usage the best performance was found for category 5 (>20h) with a least square mean task time of 3817 ms and the worst performance was found for the participants of category 1 (<1h) and task time of 5182 ms.

The subjects with high touch screen experience performed significantly better than the others and had a least square mean task time of 2944 ms compared to 5251 ms for the ones with medium touch screen experience and 4691 ms for the subjects with low touch screen experience.

Consistent with the results for the selection time in the first task there was no significant difference found between age category 2 (30-45) and category 3 (> 45) with a least square mean task time of 4610 ms and 4679 ms respectively. Age category 1 (< 30) had a significantly lower task time than the other two with a least square mean of 3597 ms.

Between the different buttons the shortest least square mean task time of 3778 ms was found for button one and the longest with 4560ms for button 5. Between those two buttons the least square mean task time increases with every button. Since the button number represents the position of the button in the pull down menu the differences in performance between the buttons can be explained by the increased distance between the button and the origin for the selection.

For the relative performance in T2 the sample size was 108 with a mean value of 69% a minimum value of 37% and a maximum value of 111%. A significant difference in relative T2 performance was found between different pointing devices (F=3.997, p=0.0216), between age categories (F=31.350, p<0.0001), between different levels of computer usage (F=7.928, p<0.0001), and between the dominant hand of the participant (F=10.868, p=0.00013).

For the different pointing devices there was no significant difference found between the mouse and the touch screen/stylus. The best performance with a least square mean value of 69.8% was found with the touch screen/stylus followed by the mouse with 65.4%. The touch screen finger had a significantly lower performance with a least square mean value of 62.2%.

Between the different age categories the best performance was found for age category 3 (>45) with a least square mean value of 78.4% followed by category 2 (30-45) with 61.8% and category 1 (<30) with 56.4%.

Between the different levels of computer usage the best performance was found for category 2 (1-4h) with a least square mean of 79.1% followed by category 4 (10-20h) with 72.4%, category 5 (>20h) with 70.5%, category 3 (4-10h) with 64.8% and category 1 (<1h) with 59.6% (Figure 12).
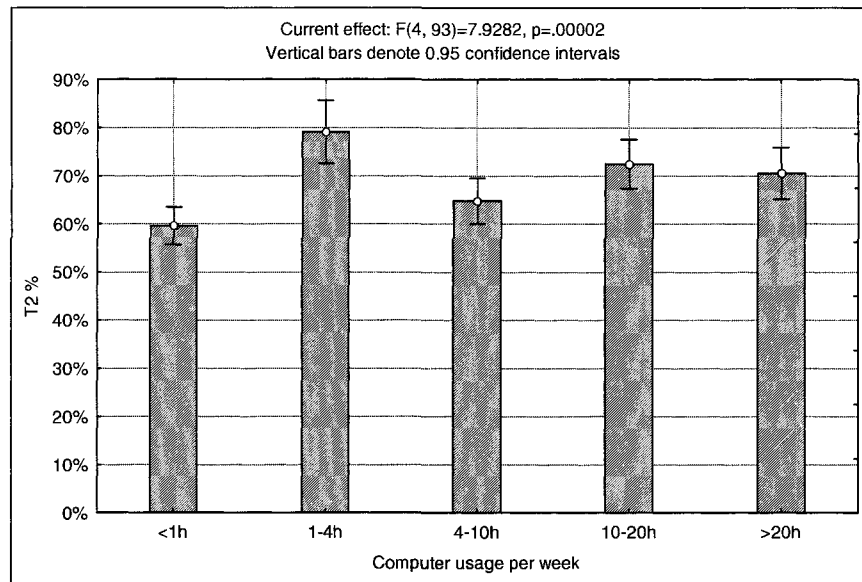
Figure 12. Relative performance in T2 by computer usage

Between the dominant hands the right handed participants had the higher performance with a least square mean value of 71.5% compared to the left handed participants with 59.6%.

Task 3:

For the third task the overall task time T3 was also the best measurement of performance. The sample size for this measurement was 1620 with a mean value of 4487 ms, a standard deviation of 2330 ms and values between 943 ms and 18603 ms. For this task the same factors as for the second task were found to have a significant influence on the task time. There was a significant difference between pointing devices ($F=31.3$, $p<0.0001$), between dominant hands ($F=98.1$, $p<0.0001$), between different levels of computer usage ($F=18.5$, $p<0.0001$), between different age categories ($F=39.3$, $p<0.0001$) and between different levels of touch screen experience ($F=26.3$, $p<0.0001$).

For the pointing devices there was no significant difference found between the mouse and the touch screen/stylus with a least square mean task time of 3776 ms and 3711 ms respectively.

With the touch screen/finger the participants had a significantly lower performance with a least square mean task time of 4589 ms.

Similar as in the other tasks the right hand dominant participants had a significantly lower task time than the left hand dominant participants. The least square mean task times were 4890 ms for the left handed and 3162 ms for the right handed participants.

Between the different levels of computer usage the best performance with a least square mean task time of 3285 ms was found for the participants of category 5 (>20h). The worst performance with a least square mean task time of 6283 ms was found for the participant of category 1 (<1h) with a steady increase in task time for the categories between those two (Figure 13).
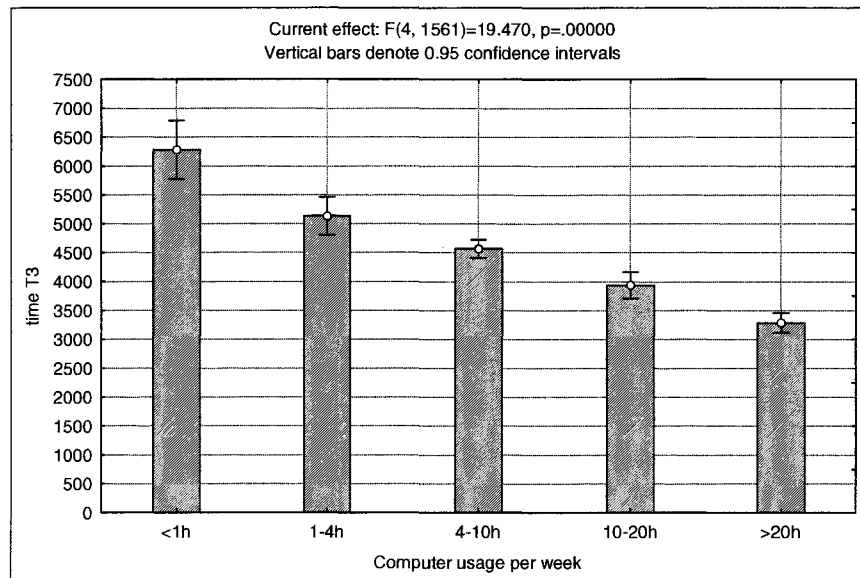
Figure 13. Task time T3 by computer usage

The results between the different levels of touch screen experience are very similar to those for the second task. The subjects with high touch screen experience performed significantly better than the others and had a least square mean task time of 2956 ms compared to 4320 ms

for the ones with medium touch screen experience and 4801 ms for the subjects with low touch screen experience.

There was no significant difference found between age category two (30-45) and category 3 (> 45) with a least square mean task time of 4441 ms and 4661 ms respectively. Age category 1 (< 30) had a significantly lower task time than the other two with a least square mean of 2975 ms.

For the relative performance of T3 the sample size was 108 with a mean value of 76%, a minimum value of 45% and a maximum of 118%. A significant difference in relative performance was found between different pointing devices (F=4.6302, p=0.0129), between different age categories (F=10.225, p=0.0001), between different levels of computer usage (F=9.778, p<0.0001), and between the dominant hand of the participants (F=30.410, p<0.0001).

Between the different pointing devices the best performance was found with the touch screen/stylus with a least square mean value of 86.6% followed by the mouse with 80.6% and the touch screen/finger with 68.0%

Between the different age categories age category 3 (>45) had the highest relative performance with a least square mean value of 85.4%, followed by age category 1 (<30) with 80.3%, and age category 2 (30-45) with 69.7%.

Between the different levels of computer usage the best performance was found for category 1 (<1h) with a least square mean value of 90.2% followed by category 4 (20-20h) with 82.7%, category 5 (>20h) with 78.4%, category 2 (1-4h) with 74.1 % and category 3 (4-10h) with 71.4 %.

Between the dominant hands the right handed participants had the higher performance with a least square mean value of 91.2% compared to the left handed participants with 65.7%.

Task 4:

For the fourth task the overall task time T4 was the best measurement of performance. For the sample size of 1620 a mean value of 3556 ms with a standard deviation of 1865 ms and values between 1159 ms and 18125ms were found. A significant difference in task time was found between pointing devices (F=12.43, p<0.0001), between dominant hands (F=8.17, p=0.004), between different levels of computer usage (F=15.20, p<0.0001), between different age categories (F= 41.4, p<0.0001), between different level of touch screen experience (F=19.8, p<0.0001) and between different locations of the target box (F=3.87, p=0.0004). There was also a significant difference in task time found for the interaction of pointing device and target box location (F=2.23, p=0.0056), the interaction of pointing device and computer usage (F=3.94, p=0.0001) and the interaction of pointer and dominant hand (F=3.04, p=0.0481)

For the pointing devices there was no significant difference found between the mouse and the touch screen/stylus with least square mean task times of 2767ms and 2903 ms. With the touch screen finger the participants had a significantly higher least square mean task time of 3742 ms.

As for the other tasks the right hand dominant participants had a significantly lower task time than the left hand dominant participants. The least square mean task times were 3359 ms for the left handed and 2916 ms for the right handed participants.

The best performance with a least square mean task time of 2437 ms was found for the participants of computer usage category 5 (>20h). The worst performance with a least square mean task time of 3671 ms was found for the participant of computer usage category 1 (<1h). For all categories between those two the performance decreased constantly.

Between the different levels of touch screen experience the participants with high touch screen experience performed significantly better and had a least square mean task time of

2319 ms compares to 3654 ms for the participants with medium experience and 3440 ms for those with little prior touch screen experience.

For the different age categories the best performance was achieved by the participants of age category 1 (<30) with a least square mean task time of 2587 ms followed by category 2 (30-45) with 3051 ms and category 3 (>45) with 3774 ms.

Between the different locations of the target box the best performance with a least square mean task time of 2454 ms was achieved when the target box was located in the top center. The highest least square mean task time with 3564 ms was found when the target box was located in the bottom left. Overall it can be seen that the task time increases for the target box being located further to the bottom and to the left (Figure 14).
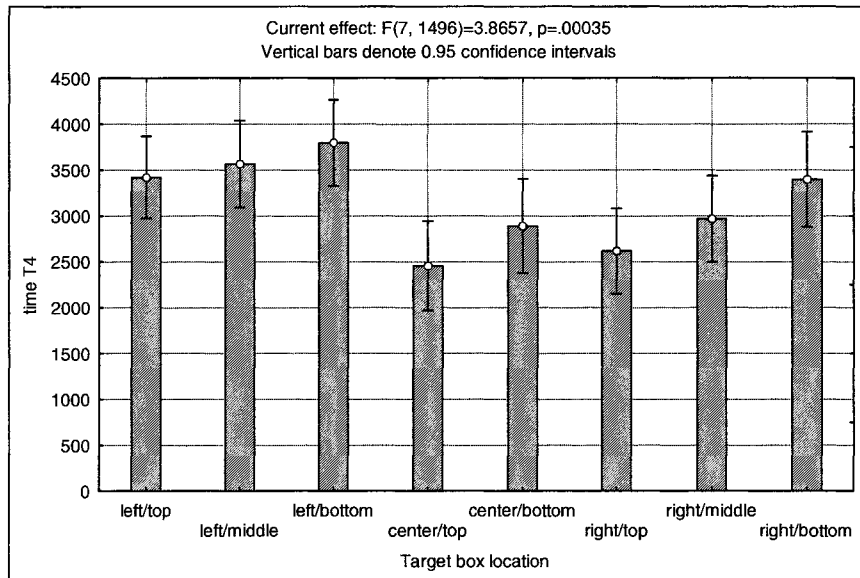
Figure 14. Task time T4 by target box location

When looking at the interaction of target box location and pointing device there was no significant difference between target box locations found with the mouse and only moderate impact on the performance with the touch screen/stylus. With the touch screen/finger the

influence of the target box location on the performance was the biggest amongst all devices (Figure 15).

It was shown that especially for the touch screen a location of the target to the top and the right is favorable. This can be explained by the fact that in the moving tractor the majority of the users rested their hand on the frame of the screen to gain stability while using their thumb for selection on the touch screen. With this strategy the targets on the lower and left side of the screen require more reaching or even to move the hand off the frame which makes the selection of targets in this part of the screen significantly harder.
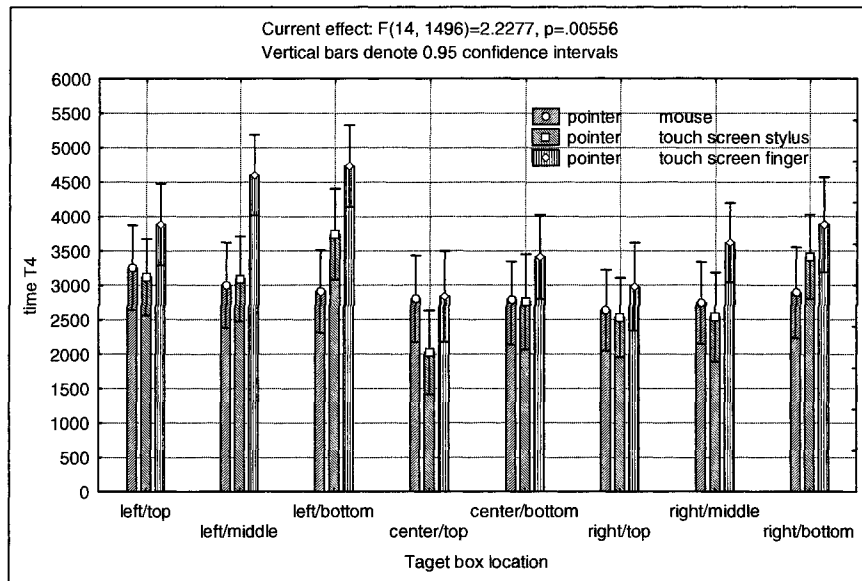
Figure 15. Task time T4 by Interaction of pointing device and target box location

For the interaction of pointing device and computer usage it was found that there is a smaller difference between devices for the participants with a more frequent computer usage than for the lower computer usage. The biggest difference between the levels of computer usage was found for the touch screen/finger followed by the touch screen/stylus and the mouse (Figure 16).
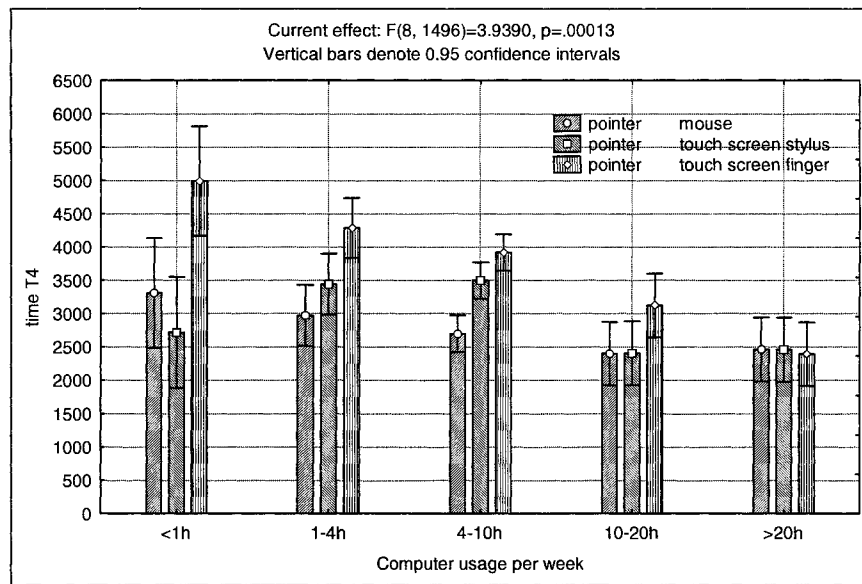
Figure 16. Task time T4 by Interaction of pointing device and computer usage

For the relative performance in T4 the sample size was 108 with a mean value of 63% a minimum of 31% and a maximum of 152%. A significant difference was found between pointing devices (F=6.834, p=0.0020), computer usage (F=2.999, p=0.0242), dominant hand (F=5.676, p=0.0200), and the interaction of computer usage and pointing device (F=2.626, p=0.0143).

For the different pointing devices there was no significant difference found between the mouse and the touch screen/stylus. With the mouse the participants had the highest performance with a least square mean of 68.8% followed by the touch screen/stylus with 67.2%. With the touch screen/finger they had a significantly lower performance with a least square mean value of 50.8%.

Between the different levels of computer usage the highest performance was found for category 4 (10-20h) with a least square mean value of 68.0% followed by category 1 (<1h) with 66.1%, category 2 (1-4h) with 61.3%, category 3 (4-10h) with 59.3%, and category 5(>20 h) with 56.6%.

Between the different dominant hands the right handed participants had a significantly higher performance in T4 with a least square mean value of 67.0% compared to the left handed participants with 57.6%.

For the interaction between pointing device and computer usage there was no significant difference in relative T4 performance between different pointing devices found for the participants with computer usage category 5. For all other levels of computer usage the mouse and the touch screen/stylus had a significantly better performance than the touch screen/finger (Figure 17).
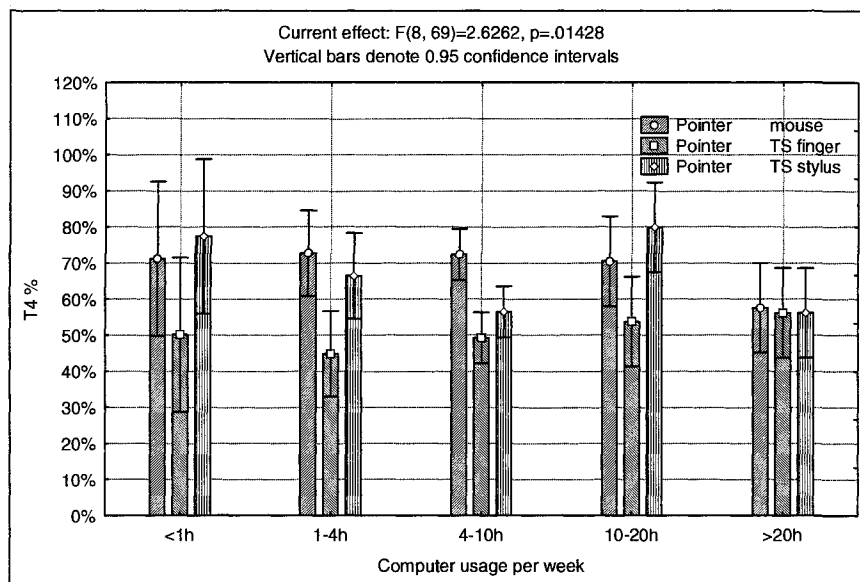
Figure 17. Relative performance in task time T4 by pointing device and computer usage

Overall relative performance:

To evaluate the overall relative performance the mean value off all performance measurements was calculated for each participant and each session. The mean value of the

108 samples for overall performance was 78% with a minimum value of 58% and a maximum value of 100%.

When comparing the relative performance of each task the highest performance compared to the office environment was found for the typing part of task 1 with 86% followed by task 3 with 76% task 2 with 69%, task 4 with 63% and the pointer part of part 1 with 62 % (Table 3).

	Valid N	Mean	Minimum	Maximum	Std.Dev.
Selection time %	108	62%	30%	97%	15%
Change 1 %	108	95%	49%	140%	19%
NTS %	108	86%	45%	122%	13%
Change 2 %	108	98%	40%	142%	24%
T2 %	108	69%	37%	111%	14%
T3 %	108	76%	45%	118%	17%
T4 %	108	63%	31%	122%	16%
Overall %	108	72%	50%	92%	9%

Table 3. Summary of relative performance by task

For the overall relative performance a significant difference was found between different interfaces (F=4.203, p=0.0182), levels of computer usage (F=8.347, p<0.0001), and different age categories (F=3.947, p=0.0230).

For the different interfaces there was no significant difference found between the mouse with a least square mean of 78.9% and the touch screen/stylus with 79.8%. With the touch screen/finger the participants had a significantly lower relative performance with a least square mean of 74.5 %.

Between the different levels of computer usage the best performance was found for category 4 with a least square mean value of 83.8% followed by category 2 (1-4h) with 83.5%, category 5 (>20h) with 79.8%, category 3 (4-10h) with 74.5%, and category 1 (<1h) with 47.0%.

Between the different age categories the best relative performance was found for category 3 (>45) with a least square mean of 80.9%, followed by category 2 (30-45) with 77.6%, and category 1 (<30) with 74.7%.

A significant difference was also found between the overall measured performance and the subjective performance rating of the test subjects for all three interfaces (mouse $p<0.0001$, TS/finger $p=0.0021$, TS/stylus $p<0.0001$). For each interface the measured performance was significantly higher than the subjective rating of the participants

The average measured performance across all performance measurements and the participant's subjective performance ratings for the different interfaces are shown in Table 4 and Figure 18, Figure 19, and Figure 20.

	Valid N	Mean	Minimum	Maximum	Std.Dev.
mouse measured	36	79%	60%	94%	8%
Mouse questionnaire	36	49%	20%	90%	20%
TS/finger measured	36	75%	61%	87%	7%
TS/finger questionnaire	36	55%	10%	90%	21%
TS/stylus measured	36	81%	58%	100%	10%
TS/stylus questionnaire	36	52%	20%	80%	18%

Table 4. Summary relative performance by interface
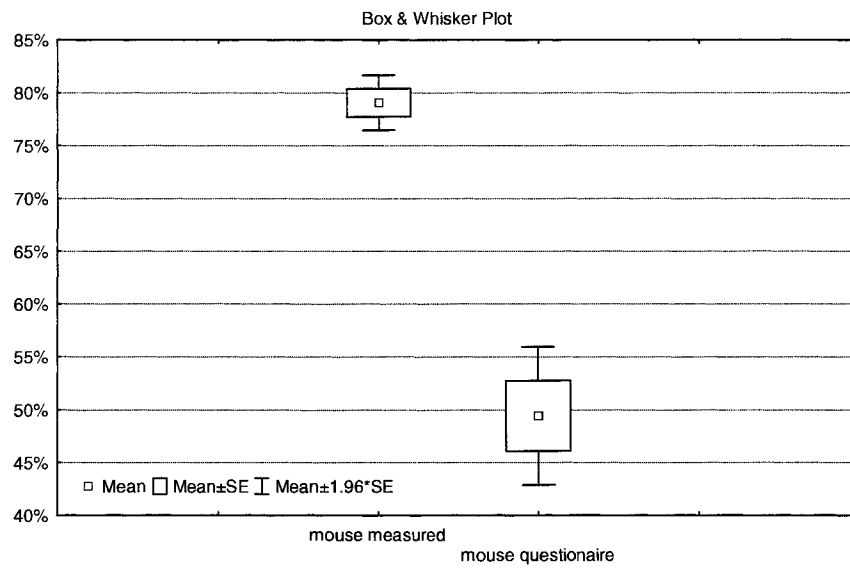
Figure 18. Mouse performance measured vs. subjective rating
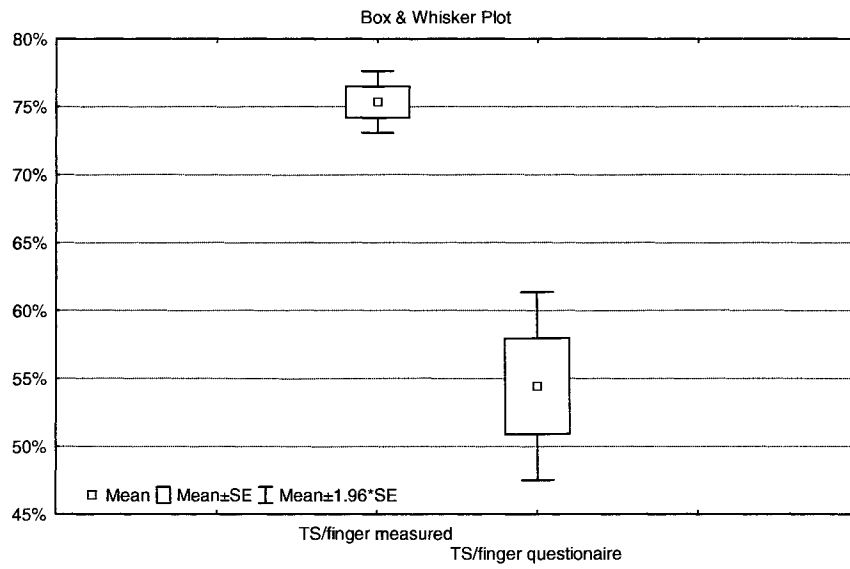
Figure 19. TS/finger performance measured vs. subjective rating

Box & Whisker Plot
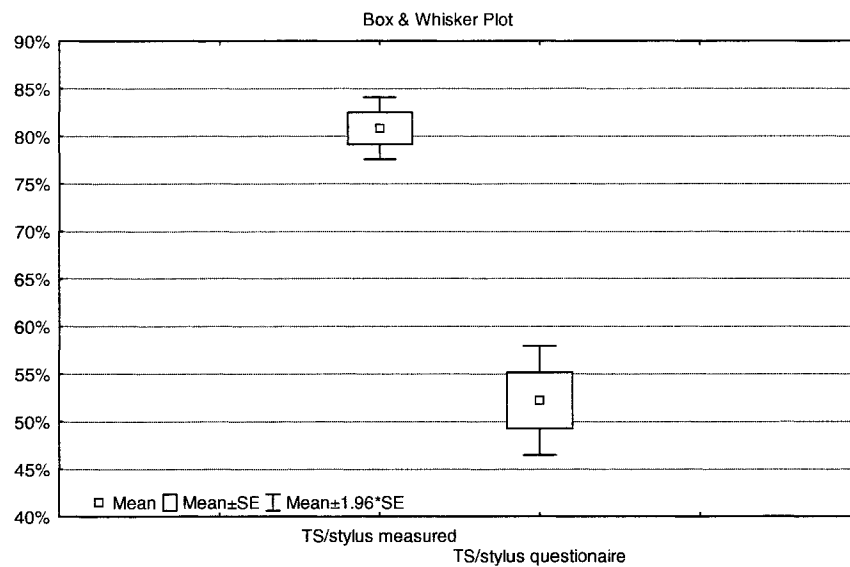
Figure 20. TS/stylus performance measured vs. subjective rating

Subjective interface evaluation

Besides the objective measurement of the performance the participants could achieve with the tested interfaces the subjective perception of performance is an important factor. To evaluate the perception of performance the participant were asked to complete a questionnaire after each session. In the questionnaire they were asked to rate the usability of the pointing and typing device for each interface and rate the overall usability of tested interface based on the experience they gathered when using the interfaces with the standard windows application and during the objective performance measurements. For every question the subjects could give a rating between 1 and 9 were for the purpose of this analysis the responses were categorized in low for ratings from 1-3, medium for 4-6 and high for 7-9.

Question 1: How would you rate the usability of this interface pointer for a standard windows application in a moving tractor

A Krusk-Wallis nonparametric median test found a significant difference in the responses for the different interfaces (Chi-square=9.687, df=2, p=0.0079). The multiple comparison found a significant difference between the mouse and the TS/finger (p=0.0488) and between the TS/finger and the TS/stylus (p=0.0290). There was no significant difference found between the mouse and the TS/stylus. Overall the TS/stylus got the highest rating with a mean of 5.861 followed by the mouse with 5.833 and the TS/finger with 5.000. The detailed results of the categorized responses to question 1 are shown in Table 5.

Device	low	medium	high	mean
mouse	3	18	15	5.833
TS/stylus	4	18	14	5.861
TS/finger	4	28	4	5.000

Table 5. Categorized responses to question 1

Question 2: How would you rate the usability of this interface keyboard for a standard windows application in a moving tractor

No significant difference (Chi-square=2.167, df=2, p= 0.338) in the responses for the different interfaces was found. The overall best results had the standard QWERTY used with the S/finger with a mean rating of 6.167 followed by the standard QWERTY with the TS/stylus with 5.917 and the mini QWERTY with the mouse with 5.417. The detailed results of the categorized responses to question 2 are shown in Table 6.

Device	low	medium	High	mean
mouse	4	22	10	5.417
TS/stylus	1	22	13	5.917
TS/finger	0	20	16	6.167

Table 6. Categorized responses to question 2

Question 3: How would you rate the overall usability of this interface for a standard windows application in a moving tractor?

No significant difference (Chi-square=4.146, df=2, p= 0.1258) in the responses for the different interfaces was found. The overall best rating was given to the TS/stylus interface with a mean rating of 5.805 followed by the mouse with a mean of 5.583 and the TS/finger with 5.194. The detailed results of the categorized responses to question 3 are shown in Table 7.

Device	low	medium	high	mean
mouse	1	24	11	5.583
TS/stylus	1	23	12	5.805
TS/finger	3	28	5	5.194

Table 7. Categorized responses to question 3

Question 4: Which of the tested interfaces allows the best productivity for a standard windows application in a moving tractor and how would you rate the productivity with this device in the tested environment?

When asked to select the interface that would provide the best productivity both the mouse and the TS/stylus got 15 votes and the TS/finger 6 votes (Table8)

There was no significant (Chi-Square=1.33, df=2, p=0.5134) difference found between interfaces when the participants were asked to rate the device they picked for being most productive. The mouse got the best rating and had a mean value of 6.80 followed by the TS/stylus with 6.20 and the TS/finger with 5.83. The detailed results of the categorized responses to question 4 are shown in Table 8.

	favorite	unsatisfactory	neutral	satisfactory	mean
mouse	15	0	5	10	6.80
TS/stylus	15	0	5	10	6.20
TS/finger	6	0	3	3	5.83

Table 8. Categorized responses to question 4

For the subjective evaluation of the tested interfaces a Kendall's Tau correlation analysis was performed to evaluate if any of the participants characteristics had a significant effect on the ratings. A at a p level of 0.05 some significant but weak correlation were found between the questionnaire responses and some of the participant's characteristics. The correlation coefficients are shown in Table 9.

Kendall Tau Correlations MD pairwise deleted; Marked correlations are significant at p <.05000

Question	age category	dominant hand	comp. usage	typing style	typing exp.	mouse exp.	TS exp.
1 mouse	-0.372	0.267	-0.248	-0.361	-	-	-
2 mouse	-	-	-0.326	-0.332	-	-	-
3 mouse	-0.328	0.256	-	-	-	-	-
1 stylus	-	-	-	-	-	0.280	0.344
2 stylus	-0.305	0.243	-	-	0.323	0.379	-
3 stylus	-	-	-0.288	-	-	-	-
1 finger	-	-0.243	-	-	-	-	-
2 finger	-0.467	0.277	-	-	-	0.286	-
3 finger	-	-	-	-	-	-	-
4a	0.262	-	-	-	-	-	-
4b	-0.290	0.236	-	-0.282	-	-	-

Table 9. Significant correlation coefficients for questionnaire results

An ordinal logistic regression analysis was performed to identify if the results of the performance measurements for the different interfaces had a significant effect on how the participants rated the interfaces in the questionnaire. A significant influence was found for the performance of the mouse/mini QWERTY interface on question 1 (p=0.0026), and question 3 (p=0.020) and for the TS stylus/standard QWERTY interface on question 3 (p=0.0360). There was no significant influence found for the TS finger/standard QWERTY interface or any of the other questions for the other two devices.

Additional comments on questionnaires

Besides rating the tested devices the participants had a chance to leave additional comments on the questionnaire. One comment made was that the location of the touch screen should be optional between the left and the right side of the cab. Even though the movable arm enabled the participants to adjust the screen to their preference one participant felt that the range of adjustment wasn't big enough to adjust it perfectly for a left handed user

One participant stated that when concentrating on the computer task he wasn't able to pay any attention to his environment. This caused a loss of his orientation after completing the computer task which might cause problems in a real world application.

Concerning the test environment one participant stated that the roughness of the field and the kind of application performed would be a major factor in the performance that could be achieved. While under relatively smooth working conditions a high level of performance compared to an office environment can be achieved, the performance would drop significantly under rough conditions.

There was also a concern that concentrating on a computer screen in a moving tractor for an extended period of time puts lot of stress on the user and might cause a decrease in performance and motion sickness. The occurrence of this effect depends on the susceptibility of the user but is also directly related to the roughness of the driving conditions and the suspension features of the vehicle.

It also has to be mentioned that one of the participants withdrew from the study in the middle of first session because he was not feeling well. It wasn't possible to determine if this was caused by participating in the study or if other factors were responsible. Two other subjects reported some discomfort when concentrating on the screen for an extended period of time but none of them to the extent that they felt sick and had to stop the test.

5.0 DISCUSSION

To determine which of the tested interfaces provided the best usability and satisfaction a user can achieve when using a standard windows application in a moving off-road vehicle both the objective performance measurements and the subjective preferences of the users have to be considered.

For the evaluation of the performance with the different interfaces a series of tasks were presented to the test participants and task specific performance measurements recorded. The tasks presented to the participants included pointing and typing tasks that were designed to represent the operations that are required to operate a standard windows application. To evaluate the performance a user can achieve with the tested interfaces both the absolute performance values that were measured and the relative performance of each participant compared to his performance in an office environment was analyzed.

For the absolute performance of the participants when doing pointing tasks the same factors were found to make a significant different for all tasks. The factors found were the pointing device used, the dominant hand of the participant, the computer usage of the participant, the categorized age of the participant and the touch screen experience of the participant.

Between the pointing devices the participants had the best performance with the mouse and the touch screen/stylus. With both of these devices the performance the participants could achieve was very similar with small advantages for either of the devices for specific tasks. When using the touch screen/finger the performance the participants could achieve was significantly lower for all pointing related tasks.

There was also a significant difference in absolute performance found between the dominant hand of the participants and between the different levels of computer usage. For all pointing related tasks the right handed participants had a significantly better absolute performance than the left handed participants and the participants with a higher computer usage had a

significantly better absolute performance than those with less frequent computer usage. A Kendall's Tau analysis found a significant correlation between the dominant hand and the computer usage of the participant. It is therefore not possible to determine if the dominant hand or the computer is causing the differences in performance or if these differences are caused by the differences in computer usage or the combination of both

There was also a significant difference in the pointing performance found between the different age categories. The participants of age category one (<30) had a significantly better absolute performance in all pointing related tasks than the participants of the other two age categories. There was no significant difference in performance found between the participants of age category two (30-45) and age category three (>45).

Besides the factors that caused significant differences for all pointing related tasks there were a couple of factors that were task specific.

For the two performance measurements of the switching between pointing and typing in task 1, change 1 (switching to typing) and change 2 (switching to pointing), the average computer usage and the typing style had a significant influence on the performance. For both measurements the touch typists and the individuals with more frequent computer use had a significantly better performance. For change 2 a significant difference was found between pointing devices with both touch screen interfaces having a better performance than the mouse interface. This can be explained by the fact that the operation of the mouse requires first to reach for the mouse and than to move the pointer while with the touch screen the user directly selects the target.

Concerning the typing related performance there was a significant difference found between different typing styles, different levels of typing experience and different levels of computer usage. As expected a significantly higher level of typing speed was found for touch typists compared to non touch typists and for the participants with a higher level of typing experience and compute usage. A Kendall's Tau analysis also found a significant correlation

between those three factors which makes it impossible to determine which of the factors has the biggest influence on the performance. For the typing performance there was no significant difference between devices found.

Besides the absolute performance values the relative performance of each interface and task compared to an office workstation was evaluated.

For all pointing related measurements there was a significant difference found between the pointing devices. As with the absolute performance measurements the mouse and the touch screen/stylus had the best relative performance with either one having minor advantages for some of the tasks. The touch screen/finger had the lowest relative performance for all tasks.

Concerning the influence of the dominant hand on the relative performance in the pointing related tasks there was no significant difference found for the selection time. For all other pointing related tasks (T2,T3,T4) the relative performance for the right handed participant was significantly higher than for the left handed ones which is consistent with the results for the absolute performance measurements.

There was also a significant difference in relative performance in all pointing related tasks found between different levels of computer usage. However unlike the absolute performance measurement where a clear trend to higher performance was found for the more frequent computer users, there is no clear trend found for the relative performance. For each of the different tasks the relative performance for the different levels differs without any obvious trends.

Between the different age levels of the participants there was also a significant difference found in relative performance for all pointing related tasks. In contrast to the results for the absolute performance the participants of age category 3 (>45) had the highest relative performance for all pointing tasks except T4 were no significant difference between the participants of the different age categories was found.

For the relative performance in change 1 and change 2 a significant difference was found between the different levels of computer usage with the more frequent computer users having a better performance than the participants with less computer usage. Besides that a significant difference between age categories was found for change 1 with a better performance for the older participants. A significant difference in relative performance between pointing devices was found for change 2 with both touch screen interfaces performing better than the mouse interface which is consistent with the results for the absolute performance.

Concerning the relative performance for the typing task there was a significant difference found between age categories with the younger participants outperforming the older ones and between typing styles with the touch typists outperforming the non touch typists.

When comparing the overall relative performance for the different tasks it is shown that for all tasks the performance in the office environment was significantly better. The relative performances that could be achieved for the different tasks were between 86% and 62% with a mean value of 72%.

For the overall relative performance across all tasks a significant difference was found between age categories, different levels of computer usage, and between the different interfaces. Concerning the performance of the different age categories similar to the results for the individual tasks the older participants had a higher average relative performance than the younger participants. Concerning the different interfaces the touch screen/stylus had the best overall relative performance followed by the mouse and the touch screen/finger.

In the subjective ratings of the usability for the pointer part of the tested interfaces the touch screen/finger was rated significantly lower than the other two devices. The mean rating of the mouse was slightly higher than that of the touch screen/stylus but the difference between the two interfaces was not significant. Both interfaces received the vast majority of their votes in the medium to high rating

For the typing part of the interfaces the standard QWERTY keyboard that was used with the touch screen/stylus and touch screen/finger interface had a slightly higher mean rating than the mini QWERTY that was used with the mouse. The difference in the rating between the two devices was not statistically significant though. For both keyboards the majority of the votes were in the medium to high range with only a few low ratings.

For the rating of the overall usability there was no significant difference found between any of the interfaces. For all interfaced the majority of the ratings were in the medium to high part of the scale with only a few low ratings. Overall the touch screen/stylus had the highest mean rating followed by the mouse and the touch screen/finger.

When asked to pick the interface that would provide them with the best productivity there was again a tie between the mouse and the touch screen/stylus. The touch screen finger was selected significantly less as the favorite device based on productivity.

Even though the participants rated the perceived relative productivity of the tested interfaces lower than they were measured, all ratings for the overall productivity with the selected device were in the medium to high part of the scale and none of the devices received any low ratings.

6.0 CONCLUSION

The goal of this study was to determine which of the tested interfaces provides the best usability with a standard application in a moving off-road environment. Besides that we wanted to evaluate the objective and subjective performance a user can expect with such an interface in the tested environment compared to the performance he can expect with an office workplace.

For both the absolute and the relative performance of the tested interfaces the mouse and the touch screen/stylus achieved the best results and seem to be best suited for the tested

environment. They are about equal in what performance a user can achieve with them with minor advantages for both devices for specific tasks. The touch screen/finger seems to be least suited for the tested environment since both the absolute and the relative performance the subjects could achieve with it were the lowest for all tested tasks.

When looking at the keyboard part of the tested interfaces only minor differences in the performance were found so that when just looking at the objective performance measurement both of the tested devices seem to be equally well suited.

In the subjective evaluation of the usability the mouse and the touch screen/stylus interface achieved the best ratings in the pointing tasks with only minor differences between them and the touch screen/finger was rated the lowest. For the typing tasks there was no difference in the user rating between the tested interfaces.

Based on the results of the subjective and objective performance evaluation the mouse and the touch screen/stylus interface provided the best usability and seem to be best suited out of the tested devices. To determine which of those two should be installed in a real world application other factors like ease of integration in the cab, expected durability under real world conditions, and tasks performed have to be considered.

For the productivity that can be achieved with office applications in the tested environment the relative performance measurement showed average values of 60% to 98% depending an interface and type of task. The mean relative performance between all participants and across all tasks was 72%. Most of the participants rated their own performance lower in their subjective rating which means the subjective impression of performance was lower than the objective performance measurements showed. Nevertheless even with underestimating the performance they can achieve in the tested environment, all of the participants rated the satisfaction of the overall productivity they could achieve with their favorite of the tested interfaces neutral or positive with non of the participants giving it an unsatisfactory rating

It was also shown that the different factors like age, average computer usage and dominant hand have different influence on the absolute and relative performance of the tested devices. While for the absolute performance the participants with higher average computer usage performed significantly better there was no such trend found for the relative performance compared to an office environment. It has to be considered that especially those individuals that have a high frequency in computer usage would be most likely to utilize a computer in the tested environment. For this kind of operator, a system like the one tested could mean an increase in overall productivity since computer operation and field work could be done at the same time. It is expected that future generations of operators will be more familiar with computers and therefore the interest in utilizing such a system will grow.

This study has shown that it is possible to provide a usable computer interface to work with standard office application in a moving off road vehicle and that a satisfying performance can be achieved by the user.

Besides determining which devices are best suited there are a lot of other factors that have to be considered when implementing a computer workplace in a mobile off-road environment.

Future research has to determine how much different field conditions and operation speeds influence the performance and what difference ride comfort improvements like a suspended cab or rear axle can influence the productivity that can be achieved with such a system.

Besides that the overall safety of such a system has to be evaluated. When using a semi autonomous steering system like John Deere's Green-Star Auto-Trac System™ obstacle avoidance is still the responsibility of the operator. Future research has to determine if it is possible for the operator to split his attention and do productive work on the computer while still being able to provide a safe operation of the farming equipment. This includes awareness of the machine's periphery to avoid accidents as well as awareness of the machine operation and the implement to avoid damaging the equipment

REFERENCES

Afifi, A., Clark, V.A., & May, S., (2004) Computer-aided multivariate analysis, Chapman & Hall/CRC, Boca Raton.

AS2708-2001 (2001) Australian Standard: Keyboarding speed tests, Standards Australia International 2001

Douglas, S.A., Kirkpatrick, A.E. and MacKenzie I.S., (1999) Testing pointing device performance and user assessment with the ISO 9241, Part 9 Standard. Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI'99 pp.215-222 ACM: New York.

ISO 9241-4 , (1998) Ergonomic requirements for office work with visual display terminals, Keyboard requirements

ISO 9241-9, (1998) Ergonomic requirements for office work with visual display terminals, Non-keyboard input device requirements

ISO 9241-11, (1998) Ergonomic requirements for office work with visual display terminals, Guidance on usability

Jeffries, R., Miller, J.R., Wharton, C., Uyeda, K.M., (1991) User interface evaluation in the real world: a comparison of four techniques. Association for Computing Machinery, 1991

Rider, K.A., Park, W., & Prohaska, M.P., (2003) Analyzing the effect of ride motion on the performance in-vehicle reaching tasks. Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society, Denver, Colorado, 2003.

MacKenzie, I. S., (1991) Fitts' law as a performance model in human-computer interaction. Doctoral dissertation. University of Toronto: Toronto, Ontario, Canada.

MacKenzie, I.S., (2005) Within-subject vs. between-subjects design :which to use?. York University: Toronto, Ontario, Canada.

MacKenzie, I.S., & Buxton, W., (1993) A Tool for rapid evaluation of input devices using Fitts' law models. SIGCHI Bulletin, 25 (3), 58-63.

MacKenzie, I.S., & Buxton, W., (1994) The prediction of pointing and dragging times in graphical user interfaces. Interaction with computers, 6, 213-227.

MacKenzie, I. S., Sellen, A., & Buxton, W., (1991) A comparison of input devices in elemental pointing and dragging tasks. Proceedings of the CHI `91 Conference on Human Factors in Computing Systems, pp. 161-166. New York: ACM.

MacKenzie, I.S., Sourkoreff, R.W., (2003) Phrase sets for evaluating text entry techniques. Extended Abstract of the ACM Conference on Human Factorsin Computing Systems- CHI 2003, pp. 754-755

Salvucci, D.D., (2001) Predicting the effects of in-car interface use on driver performance: an integrated model approach. International Journal of Human Computer Studies vol. 55 , pp 85-107

Shneiderman, B., (1998) Designing the user interface: strategies for effective human-computer-interaction, 3rd ed. Addison Wesley Longman, Inc. 1998

Society—31st Annual meeting—1987(pp. 585-589). Santa Monica, CA: Human Factors Society

CONCLUSION

6.1 GENERAL DISCUSSION

Four studies were conducted to investigate the feasibility of using a computer, running standard office application in a moving off-road environment. First it was determined which commercially available interface devices for pointing and for alphanumeric input provided the best usability and were therefore best suited for this environment. The performance a user can achieve in a mobile tractor utilizing these devices was than compared to the performance a user can achieve in an office environment.

The first study that was conducted was the pilot pointer study. In this study the performance a user can achieve with five commercially available pointing devices in a moving tractor was evaluated. To evaluate the performance fourteen participants were presented with a series of multidirectional discrete pointing tasks and the task time the throughput and the error rate were recorded. Based on these performance measurements it was shown that all tested devices had a significantly worse performance in the moving tractor compared to the stationary tractor and also that the target size had a significant influence on the performance of all devices. The influence of the target size was found to be more substantial in the moving tractor compared to the stationary. Out of the tested devices the track ball and the touch stick performed poorly in all aspects of this study and it was determined that they are not well suited to be used in the tested environment.

The touch pad, the mouse and the touch screen showed a good performance in some aspects of the study. While the touch screen had the lowest movement time and the highest throughput it also was the device with the highest error rate. The touch pad on the other hand had the lowest error rate amongst the tested devices but didn't allow a good performance in movement time and throughput. The mouse was the device with the most balanced performance and had the second best performance in error rate, throughput and in movement time. Based on these results the mouse, the touch pad, and the touch screen were determined to provide good usability for some aspects and have potential to be well suited for the tested

environment. None of the tested devices was the clear winner since none of them provided superior usability in all aspects of the study. Besides that it was found that the discrete pointing task presented to the participants is not realistic representation of the pointer use in a standard application. When an error was made it was recorded but didn't effect the other performance measurements. It was hard to determine how pointer speed and accuracy should be weighted to determine the device with the best overall performance.

In the second pointer study the three devices that showed good potential in the pilot study were evaluated in more detail. Eighteen subjects were presented with complex pointing tasks in a moving tractor. The task was designed to simulate the use of a pointer in a standard application and the participant was forced to perform corrective action when a wrong selection was made. The overall task time was recorded as prime performance measurement since it included the pointing time and the time for the corrective action if an error was made. Additionally to the objective performance measurements the participants in the study were asked to complete a questionnaire to evaluate the subjective usability of the tested devices. The analysis of the objective performance measurements showed consistent with the results of the pilot study there was again a significantly lower performance found in the moving tractor compared to the stationary tractor and for the small targets compared to the large targets. Additionally there was a significant difference found between the different age categories with the younger subjects outperforming the older ones. Concerning the tested devices the touch pad had the worst performance in all aspects of this study. The mouse and the touch screen had a significantly better performance than the touch pad with both of these devices performing almost identical. The mouse had a slightly better performance with the small targets in the moving tractor while the touch screen had a slight advantage with the bigger targets. In the subjective user evaluation the mouse was also rated best in all categories with the touch screen having the second best overall rating. This study showed that the mouse is overall the best suited device for the tested environment since it was one of the two devices that had the best objective performance and was also rated best in the subjective rating of the participants. The touch screen performed almost identical to the mouse in the objective performance measurements and was second in the overall subjective

rating. When looking at the results of this study it has to be considered that the mouse was the device for which the subjects had the most prior exposure too. Both mouse and the touch screen seem well suited to be used in the described environment and when choosing a device for a real world application other factors like ease of implementation and tasks performed should also be considered.

The goal of the third study was to determine the device that offered the best usability is best suited to perform alphanumeric input with a standard application in a moving off-road environment. A wide variety of commercially available alphanumeric input devices and techniques were presented to a focus group that decided which ones have the potential to be well received by the user and perform well in the testes environment. Five devices were identified to be included in this study. All devices selected were keyboards with different characteristics like key size, key spacing, key travel, or actuation force. For this study eighteen participants were asked to perform a series of typing task in a moving tractor utilizing the tested devices. Besides the objective performance measurements a questionnaire was administered to evaluate the subjective usability rating of the tested devices. The objective performance measurements showed that as in the two pointer studies the tractor movement had a significant effect on the usability of the tested devices. With all tested devices the participants had a significantly better performance in the stationary environment compared to the moving environment. Besides that as expected the participants with more typing experience had a higher performance than the once with less typing experience and the touch typists had an overall higher performance than the non touch typists. Between the tested devices the standard QWERTY and the mini QWERTY had the overall best performance with no significant difference between them. They were also the devices with that received the highest ratings in the subjective user evaluation. Both devices are identical in most of their characteristics and represent the keyboard characteristics like layout, key size, key spacing, and key activation force of the standard keyboard the majority of computer users are familiar with. Especially for the touch typists who rely on their motor memory to utilize a keyboard this was expected. Using an input device with different characteristics would most certainly mean a decrease in performance unless the same motor memory is

gained with the different device. But also for the non touch typists, who don't rely on motor memory and therefore are less biased towards the standard keyboard, the best performance was found with the same devices. Both of these devices seem to be well suited to be used in the tested environment. For a specific situation other factors like space constrains, the pointing device that is used, and the task that's performed determine which one is the better choice.

The goal of the fourth study was to identify which combination of pointing and alphanumeric input device provides the interface with the best usability and to determine the performance a user can expect when using a standard application in the tested environment. The pointing and typing devices identified in the prior studies to be well suited for the tested environment were combined into sensible user interfaces and installed in a tractor. Twelve tractor owner/operators were asked to use these interfaces with standard applications in a moving tractor and rate their subjective usability and the performance they can achieve compared to an office environment. The participants were also asked to perform a series of tasks typical for the use of a standard application in a moving tractor and in an office environment to evaluate the performance that could be achieved in both environments. For both, the absolute and the relative performance compared to the office environment, the mouse and the touch screen/stylus interface achieved the best results. The performance the participants could achieve with both interfaces was about equal with minor advantages for either one for specific tasks. With the touch screen/finger interface the participants had the lowest ratings in both relative and absolute performance. For the relative performance that can be achieved in the tested environment compared to an office environment average values between 60% and 90% were found depending on the interface used and the task performed. In the subjective usability evaluation of the participants the mouse and the touch screen/stylus also achieved the best ratings with only minor differences between them while the touch screen/finger interface receives lower ratings by all participants. Even though the majority of the participants rated their perceived performance in the moving tractor lower than it was measured most of them rated the overall usability and productivity they can achieve with the tested interfaces as satisfactory.

Overall this study was able identify the best commercially available pointing and alphanumeric input device to be used with a standard windows application in a moving off-road environment and determine the performance a user can expect with such an interface.

The devices that were determined to be best suited were the mouse and the touch screen/stylus for pointing tasks and the standard QWERTY and mini QWERTY keyboard for alphanumeric input. For the most part these are the same devices that are commonly used in an office environment and most users are familiar with. Besides the touch screen none of the alternative devices evaluated were able to provide better performance in the moving environment or even compete with the mouse and the standard keyboard in performance or user preference. The reason for this might lie in the fact that during the course of this study it was not possible to provide the participants the same amount of exposure and experience with the other tested devices that they had with the 'standard' devices. But this would also be expected to be true for the overall population since the vast majority of computer users are most familiar with the mouse and the standard keyboard. It would be hard to motivate a user to invest time and sacrifice performance to get familiar with an alternative device if an improvement in comfort and performance is not guaranteed.

The majority of the participants rated the performance they could achieve with a system like this as satisfactory. Especially the participants that were touch typists and had a lot of prior computer experience performed well and achieved performances that were close to those in an office environment. It is expected that especially younger owner/operators who are more likely to use computers to manage their farming operation would be interested in installing such a system in a tractor. For this kind of farmer it would provide a possibility for additional productivity and it would allow them to multitask and perform additional work while doing fieldwork.

6.2 RECOMMENDATIONS FOR FUTURE RESEARCH

Besides determining which devices are best suited there are many additional factors that have to be considered when implementing a computer workplace in a mobile off-road environment.

The productivity a user can achieve with such a system strongly depends on the task performed and riding conditions of the tractor. Future research has to determine how much different field conditions and operation speeds influence the performance and what difference ride comfort improvements like a suspended cab or rear axle can influence the productivity that can be achieved with such a system. Other issues that need to be investigated are the effects of an extended usage of such a system under rough riding conditions. While extended use under rough conditions will decrease the performance of the user due to fatigue, an additional complication might be motion sickness.

When implementing a user interface in the described environment other factors besides the performance and subjective user preference also have to be considered. Especially for tractor applications were dust and dirt is always present the durability and dependability of a device in this kind of environment has to be evaluated.

Besides that the overall safety of such a system has to be evaluated. When using a semi autonomous steering system like John Deere's Green-Star Auto-Trac System™ obstacle avoidance is still the responsibility of the operator. Future research has to determine if it is possible for the operator to split his attention and do productive work on the computer while still being able to provide a safe operation of the farming equipment. This includes awareness of the machines periphery to avoid accidents as well as awareness of the machine operation and the implement to avoid damaging the equipment. Also to be investigated are which kinds of sensors and safety devices are necessary to assist the operator with this task.

6.3 REFFERENCES

Afifi, A., Clark, V.A., & May, S., (2004) Computer-aided multivariate analysis, Chapman & Hall/CRC, Boca Raton.

AS2708-2001 (2001) Australian Standard: Keyboarding speed tests, Standards Australia International 2001

Bohan M., Solocum, J., Shaikh, A.D., Chaparro, B., (2003) Examining first-time usage of the RollerMouse, Usability News 5.2 2003

Cerney M.M., Mila, B.D.,Hill, L.C.,(2004) Comparison of mobile text entry methods, Proceedings of the Human Factors and Ergonomic Society 48th annual meeting 2004

Douglas, S.A., Kirkpatrick, A.E. and MacKenzie I.S., (1999) Testing pointing device performance and user assessment with the ISO 9241, Part 9 Standard. Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI'99 pp.215-222 ACM: New York.

Fitts, P. M., (1954) The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 47 (1954), 381-391

Hochberg, Y., Tamhane, A.C., (1987) Multiple comparison procedures. Wiley, New York

ISO 9241-4 , (1998) Ergonomic requirements for office work with visual display terminals, Keyboard requirements

ISO 9241-9, (1998) Ergonomic requirements for office work with visual display terminals, Non-keyboard input device requirements

ISO 9241-11, (1998) Ergonomic requirements for office work with visual display terminals, Guidance on usability

Jeffries, R., Miller, J.R., Wharton, C., Uyeda, K.M., (1991) User interface evaluation in the real world: a comparison of four techniques. Association for Computing Machinery, 1991

Rider, K.A., Park, W., & Prohaska, M.P., (2003) Analyzing the effect of ride motion on the performance in-vehicle reaching tasks. Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society, Denver, Colorado, 2003.

Lueder, R & Grant, C., (1997) Alternative Keyboard Designs written for Ergonomics Branch the National Institute of Occupational Safety & Health

MacKenzie, I. S., (1991) Fitts' law as a performance model in human-computer interaction. Doctoral dissertation. University of Toronto: Toronto, Ontario, Canada.

MacKenzie, I.S., (2005) Within-subject vs. between-subjects design :which to use?. York University: Toronto, Ontario, Canada.

MacKenzie, I.S., & Buxton, W., (1993) A Tool for rapid evaluation of input devices using Fitts' law models. SIGCHI Bulletin, 25 (3), 58-63.

MacKenzie, I.S., & Buxton, W., (1994) The prediction of pointing and dragging times in graphical user interfaces. Interaction with computers, 6, 213-227.

MacKenzie, I.S., Nonnecke, B. Riddersma, S., McQueeen, C. & Meltz, M.,(1994) Alphanumeric entry on pen-based computers. International Journal of Human-Computer Studies, 41, 775-792.

MacKenzie, I. S., Sellen, A., & Buxton, W., (1991) A comparison of input devices in elemental pointing and dragging tasks. Proceedings of the CHI `91 Conference on Human Factors in Computing Systems, pp. 161-166. New York: ACM.

MacKenzie, I.S., Sourkoreff, R.W., (2002) Text entry for mobile computing: Models and methods, theory and practice. Human-Computer Interaction, 17, 147-198.

MacKenzie, I.S., Sourkoreff, R.W., (2003) Phrase sets for evaluating text entry techniques. Extended Abstract of the ACM Conference on Human Factorsin Computing Systems- CHI 2003, pp. 754-755

MacKenzie, I.S., Zhang, S.X. & Sourkoreff, R.W., (1999) Text entry using soft keyboards. Behavior & Information Technology, 18, 235-244

MacKenzie, I.S.& Zhang, S.X., (1999) The design and evaluation of a high performance soft keyboard. Proceedings of the ACM Conference on Human Factors in Computing Systems- CHI 99, pp. 25-31

MacKenzie, I.S.& Zhang, S.X., (2001) An empirical investigation of the novice experience with soft keyboards. Behavior& Information Technology, 20,411-418

Ober, B.S., (1983) The difficulty level of typewritten copy in industry. Delta Pi Epsilon Journal, 1983, vol. 25, no 1, pp 1-8

Pavlovych A. and Stuerzlinger W., An analysis of novice text entry performance on large interactive wall surfaces, Dept. of Computer Science, York University Toronto

Pavlovych A. and Stuerzlinger W., Less Tap: A fast and easy to learn text input technique for phones. , Dept. of Computer Science, York University Toronto

Salvucci, D.D., (2001) Predicting the effects of in-car interface use on driver performance: an integrated model approach. International Journal of Human Computer Studies vol. 55 , pp 85-107

Sears, A., (1991) Improving touch screen keyboards: design issues and a comparison with other devices. Interacting with Computers vol. 3, 253-269

Sears A., (2003) Data Entry for mobile devices using soft keyboards: understanding the effects of keyboard size and user task. International journal of human-computer interaction, 16 (2), 163-184

Shneiderman, B., (1998) Designing the user interface: strategies for effective human-computer-interaction, 3rd ed. Addison Wesley Longman, Inc. 1998

Silverberg, M., MacKenzie, I.S., & Korhonen, P., (2000) Predicting text entry speed on mobile phones. Proceedings of the ACM Conference on Human Factors in Computing Systems – CHI 2000, pp. 9-16. New York: ACM.

Perry, D.J., (1968) An analytical comparison of the relative word-combination frequency of business correspondence with phrase frequency of selected shorthand textbooks, vols I and II, University of North Dakota 1968

Sears, A., Revis, D., Swatski, J., Crittenden, R., Schneiderman, B., (1993) Investigating touchscreen typing: the effect of keyboard size on typing speed. Behaviour & Information Technology, vol. 12, 17-22

Sears, A.& Ying, Z., (2003) Data entry for mobile devices using soft keyboards. International Journal of Human-Computer Interaction 16(2), 163-184

Soukoreff, W., & MacKenzie, I. S., (1995). Generalized Fitts' law model builder. Companion Proceedings of the ACM Conference on Human Factors in Computing Systems – CHI '95, pp. 113-114. New York: ACM

Soukoreff, W., & MacKenzie, I. S., (2004).Recent development in text-entry error rate measurement, Extended abstract of the ACM conference on Human Factors in Computing Systems- CHI 2004

Wikund, M.E., Dumas, J.S., &Hoffmann, L.R. (1987). Optimizing a portable terminal keyboard for combined one-handed and two-handed use. Proceedings of the Human Factor Society—31st Annual meeting—1987(pp. 585-589). Santa Monica, CA: Human Factor Society

West, L.J., (1968) The vocabulary of instructional materials for typing and stenographic training-Research findings and implications. Delta Pi Epsilon Journal, 1968, vol. 10, no 3, pp 13-25

Zhai, S., (2000) An exploration of quantitative techniques for virtual keyboard Design ACM Symposium on User Interface Software and Technology (UIST 2000) pp 119-128

ACKNOWLEDGMENTS

I wish to thank my advisor Patrick E Patterson and the other members of my committee Samuel K. Adams, Roger E. Baer, John K. Jackman, Mack C. Shelley and Adrian Sannier for their support and advice.

Many people at Iowa State University assisted and encouraged me in various ways during the course of my studies and I want to thank them for that and I would like to especially thank the faculty and staff of the IMSE department for all the help they provided over my time at Iowa State University.

I owe a special note of gratitude to Julia Apple-Smith who not only supported and helped me in her function as international program coordinator at the beginning of my journey at Iowa State but also became a good friend who always had an open ear for my problems.

I extend my thanks to everybody at the John Deere Product Engineering Center who supported my work in many different ways and without whom this wouldn't have been possible. I want to especially thank my supervisor Bruce Newendorp who always supported my efforts and provided me with the possibility to do this research. I would also like to thank Jim Stevenson who provided valuable advice on the statistical analysis and other aspects of my research.

I'd like to thank those who volunteered as test subjects for various studies and endured long hours of completing my experiments which was not always the most pleasant experience.

Writing a dissertation is a long and interesting experience and it is obviously not possible without the support of numerous people. Thus my sincere gratitude goes to my parents Rudolf and Walburga, all my friends and colleges and everybody else who supported me in completing this dissertation.

APPENDIX A:

ANALYSIS AND PROCEDURE DETAILS OF CHAPTER 2

Analysis Details Task Time

Univariate Tests of Significance for Time (Sheet1 in 1stdata.stw) Sigma-restricted parameterization
Effective hypothesis decomposition

	SS	Degr. of	MS	F	p
Intercept	2.634683E+10	1	2.634683E+10	15859.18	0.000000
Amplitude	1.880419E+08	1	1.880419E+08	113.19	0.000000
Width	5.356247E+08	1	5.356247E+08	322.41	0.000000
Device	2.691975E+09	4	6.729938E+08	405.10	0.000000
Motion	1.027195E+09	1	1.027195E+09	618.31	0.000000
Amplitude*Width	1.500621E+04	1	1.500621E+04	0.01	0.924285
Amplitude*Device	5.591974E+07	4	1.397993E+07	8.42	0.000001
Width*Device	7.784915E+07	4	1.946229E+07	11.72	0.000000
Amplitude*Motion	3.374605E+06	1	3.374605E+06	2.03	0.154135
Width*Motion	9.350416E+07	1	9.350416E+07	56.28	0.000000
Device*Motion	1.831864E+08	4	4.579660E+07	27.57	0.000000
Error	1.112572E+10	6697	1.661299E+06		

Raw Residuals Task Time

Histogram of Raw Residuals
Dependent variable: Time
(Analysis sample)
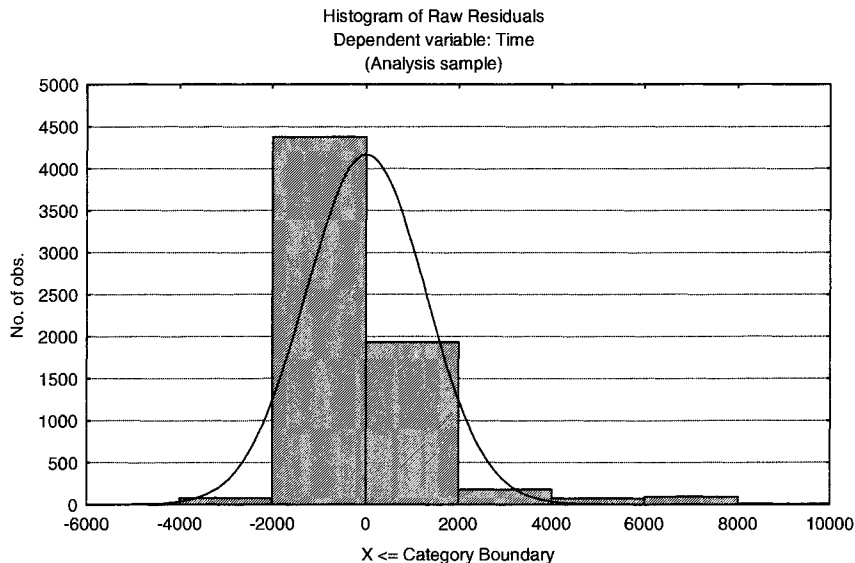
Analysis Details Throughput

	SS	Degr. of	MS	F	p

Univariate Tests of Significance for Throughput (Sheet1 in 1stdata.stw) Sigma-restricted parameterization Effective hypothesis decomposition

	SS	Degr. of	MS	F	p
Intercept	27846.49	1	27846.49	50900.66	0.000000
{1}Amplitude	170.58	1	170.58	311.81	0.000000
{2}Width	37.05	1	37.05	67.72	0.000000
{3}Device	4303.93	4	1075.98	1966.79	0.000000
{4}Motion	412.54	1	412.54	754.08	0.000000
Amplitude*Width	3.69	1	3.69	6.75	0.009404
Amplitude*Device	280.93	4	70.23	128.38	0.000000
Width*Device	13.04	4	3.26	5.96	0.000088
Amplitude*Motion	6.21	1	6.21	11.35	0.000758
Width*Motion	36.59	1	36.59	66.88	0.000000
Device*Motion	49.66	4	12.41	22.69	0.000000
Amplitude*Width*Device	11.09	4	2.77	5.07	0.000449
Amplitude*Width*Motion	0.10	1	0.10	0.17	0.675779
Amplitude*Device*Motion	15.35	4	3.84	7.01	0.000012
Width*Device*Motion	28.78	4	7.19	13.15	0.000000
1*2*3*4	0.70	4	0.18	0.32	0.864816
Error	3653.91	6679	0.55		

Raw Residuals Throughput

Histogram of Raw Residuals
Dependent variable: Throughput
(Analysis sample)

Analysis Details Error Rate

Univariate Tests of Significance for E (Sheet1 in 1stdata.stw) Sigma-restricted parameterization
Effective hypothesis decomposition

	SS	Degr. of	MS	F	p
Intercept	32.7323	1	32.73229	555.1297	0.000000
{1}Amplitude	0.0073	1	0.00729	0.1237	0.725105
{2}Width	5.7751	1	5.77515	97.9448	0.000000
{3}Device	10.0080	4	2.50201	42.4333	0.000000
{4}Motion	13.1263	1	13.12634	222.6187	0.000000
Amplitude*Width	0.0180	1	0.01801	0.3054	0.580550
Amplitude*Device	0.2628	4	0.06570	1.1142	0.347779
Width*Device	5.9503	4	1.48757	25.2288	0.000000
Amplitude*Motion	0.2037	1	0.20372	3.4550	0.063103
Width*Motion	2.1787	1	2.17872	36.9504	0.000000
Device*Motion	3.5217	4	0.88043	14.9319	0.000000
Amplitude*Width*Device	0.0765	4	0.01912	0.3243	0.861836
Amplitude*Width*Motion	0.0073	1	0.00729	0.1237	0.725105
Amplitude*Device*Motion	0.0455	4	0.01138	0.1931	0.942117
Width*Device*Motion	0.9842	4	0.24606	4.1730	0.002236
1*2*3*4	0.2271	4	0.05677	0.9628	0.426584
Error	393.8750	6680	0.05896		

Raw Residuals Error Rate

Histogram of Raw Residuals
(Analysis sample)
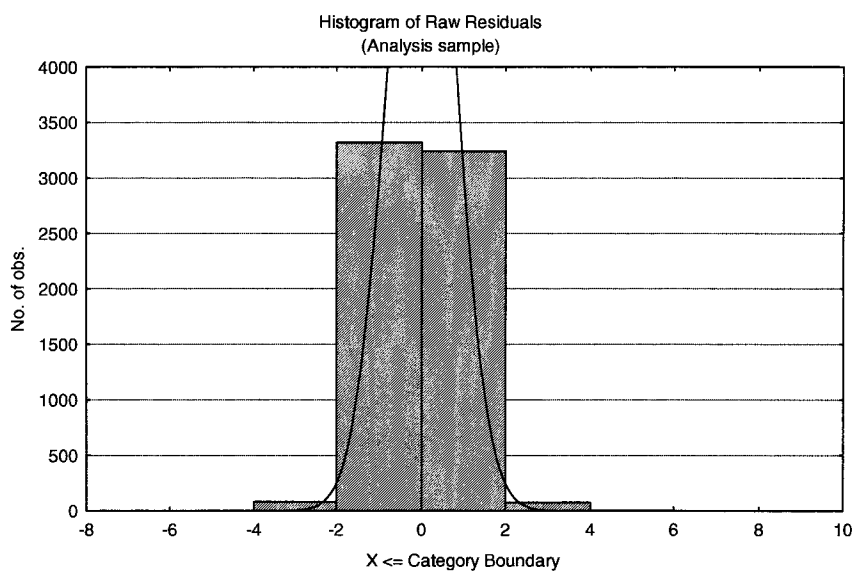
APPENDIX B:

ANALYSIS AND PROCEDURE DETAILS OF CHAPTER 3

Balanced test order design

subject	Session					
	1			2		
1	Mouse	Touch pad	touch screen	touch screen	Mouse	Touch pad
2	Mouse	touch screen	Touch pad	Touch pad	Mouse	touch screen
3	Touch pad	Mouse	touch screen	touch screen	Touch pad	Mouse
4	Touch pad	touch screen	Mouse	Mouse	Touch pad	touch screen
5	touch screen	Mouse	Touch pad	Touch pad	touch screen	Mouse
6	touch screen	Touch pad	Mouse	Mouse	touch screen	Touch pad
7	Mouse	Touch pad	touch screen	touch screen	Mouse	Touch pad
8	Mouse	touch screen	Touch pad	Touch pad	Mouse	touch screen
9	Touch pad	Mouse	touch screen	touch screen	Touch pad	Mouse
10	Touch pad	touch screen	Mouse	Mouse	Touch pad	touch screen
11	TS/ finger	Mouse	Touch pad	Touch pad	touch screen	Mouse
12	TS/ finger	Touch pad	Mouse	Mouse	touch screen	Touch pad
13	Mouse	Touch pad	touch screen	touch screen	Mouse	Touch pad
14	Mouse	touch screen	Touch pad	Touch pad	Mouse	touch screen
15	Touch pad	Mouse	touch screen	touch screen	Touch pad	Mouse
16	Touch pad	touch screen	Mouse	Mouse	Touch pad	touch screen
17	touch screen	Mouse	Touch pad	Touch pad	touch screen	Mouse
18	touch screen	Touch pad	Mouse	Mouse	touch screen	Touch pad

Analysis Details Throughput

Univariate Tests of Significance for TP (moving+stationary.sta) Sigma-restricted parameterization
Effective hypothesis decomposition

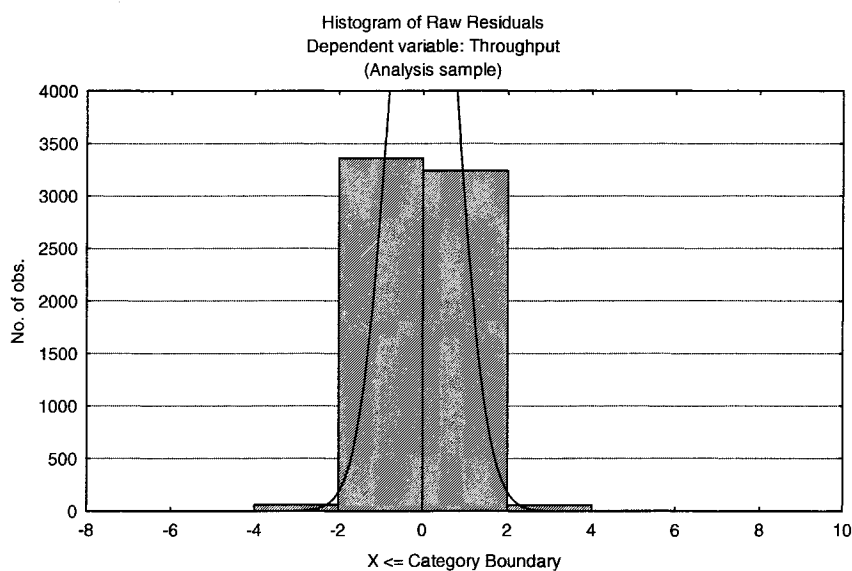|	SS	Degr. of	MS	F	p
Intercept	47129.77	1	47129.77	67437.42	0.000000
{1}pointer	2900.24	2	1450.12	2074.96	0.000000
{2}size	29.70	1	29.70	42.50	0.000000
{3}tractor state	931.18	1	931.18	1332.41	0.000000
{4}block	55.98	1	55.98	80.11	0.000000
{5}Age Cat	585.16	2	292.58	418.65	0.000000
pointer*size	5.93	2	2.96	4.24	0.014437
pointer*tractor state	64.62	2	32.31	46.23	0.000000
size*tractor state	1.04	1	1.04	1.49	0.222832
pointer*block	2.88	2	1.44	2.06	0.127648
size*block	2.37	1	2.37	3.39	0.065534
tractor state*block	19.83	1	19.83	28.38	0.000000
pointer*Age Cat	44.62	4	11.16	15.96	0.000000
size*Age Cat	0.09	2	0.04	0.06	0.938143
tractor state*Age Cat	4.09	2	2.04	2.93	0.053686
block*Age Cat	4.11	2	2.05	2.94	0.052983
pointer*size*tractor state	0.02	2	0.01	0.01	0.985408
pointer*size*block	0.28	2	0.14	0.20	0.815560
pointer*tractor state*block	3.57	2	1.78	2.55	0.078090
size*tractor state*block	0.00	1	0.00	0.01	0.941103
pointer*size*Age Cat	1.71	4	0.43	0.61	0.653737
pointer*tractor state*Age Cat	14.40	4	3.60	5.15	0.000384
size*tractor state*Age Cat	2.08	2	1.04	1.49	0.226111
pointer*block*Age Cat	3.35	4	0.84	1.20	0.308813
size*block*Age Cat	1.20	2	0.60	0.86	0.423002
tractor state*block*Age Cat	0.80	2	0.40	0.57	0.563630
1*2*3*4	0.20	2	0.10	0.14	0.869278
1*2*3*5	2.16	4	0.54	0.77	0.541765
1*2*4*5	2.37	4	0.59	0.85	0.495629
1*3*4*5	16.44	4	4.11	5.88	0.000101
2*3*4*5	0.68	2	0.34	0.48	0.616465
1*2*3*4*5	0.66	4	0.17	0.24	0.917752
Error	4780.25	6840	0.70		

Raw Residuals Throughput

Histogram of Raw Residuals
Dependent variable: TP
(Analysis sample)
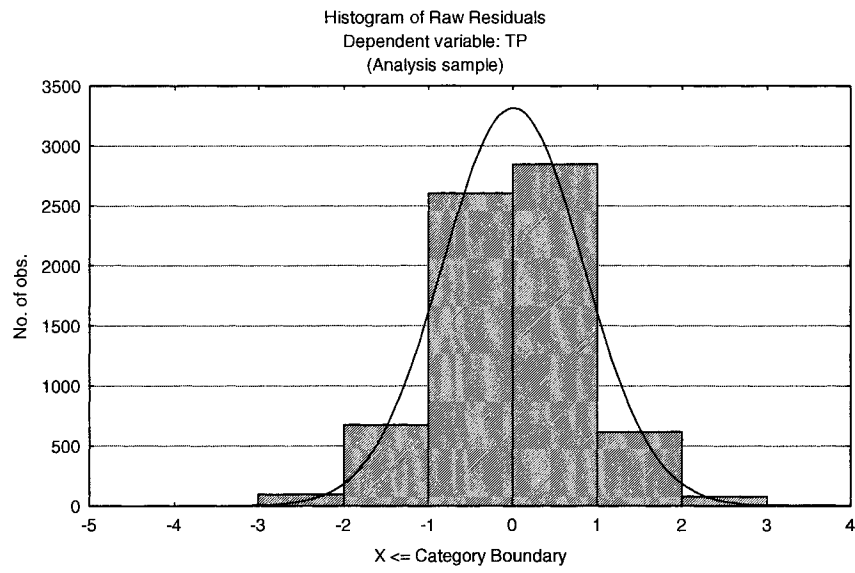
Analysis Details Tasks/sec

Univariate Tests of Significance for task/s (moving+stationary.sta) Sigma-restricted parameterization
Effective hypothesis decomposition

	SS	Degr. of	MS	F	p
Intercept	193.0380	1	193.0380	65578.79	0.000000
{1}pointer	6.4822	2	3.2411	1101.06	0.000000
{2}size	1.1633	1	1.1633	395.21	0.000000
{3}tractor state	5.3868	1	5.3868	1830.00	0.000000
{4}block	0.3493	1	0.3493	118.67	0.000000
{5}Age Cat	2.7892	2	1.3946	473.77	0.000000
pointer*size	0.1081	2	0.0540	18.36	0.000000
pointer*tractor state	0.2583	2	0.1291	43.87	0.000000
size*tractor state	0.0126	1	0.0126	4.26	0.038956
pointer*block	0.0394	2	0.0197	6.70	0.001244
size*block	0.0014	1	0.0014	0.47	0.493317
tractor state*block	0.1032	1	0.1032	35.07	0.000000
pointer*Age Cat	0.1596	4	0.0399	13.55	0.000000
size*Age Cat	0.0060	2	0.0030	1.02	0.361146
tractor state*Age Cat	0.0110	2	0.0055	1.87	0.154634
block*Age Cat	0.0183	2	0.0091	3.10	0.045023
pointer*size*tractor state	0.0005	2	0.0003	0.09	0.918286
pointer*size*block	0.0075	2	0.0038	1.28	0.277784
pointer*tractor state*block	0.0051	2	0.0026	0.87	0.419948
size*tractor state*block	0.0007	1	0.0007	0.23	0.634041
pointer*size*Age Cat	0.0024	4	0.0006	0.21	0.934332
pointer*tractor state*Age Cat	0.1586	4	0.0396	13.47	0.000000
size*tractor state*Age Cat	0.0052	2	0.0026	0.88	0.416574
pointer*block*Age Cat	0.0141	4	0.0035	1.20	0.308777
size*block*Age Cat	0.0001	2	0.0001	0.02	0.981624
tractor state*block*Age Cat	0.0079	2	0.0039	1.34	0.262282
1*2*3*4	0.0004	2	0.0002	0.07	0.934066
1*2*3*5	0.0116	4	0.0029	0.99	0.414125
1*2*4*5	0.0076	4	0.0019	0.64	0.632311
1*3*4*5	0.0139	4	0.0035	1.18	0.315616
2*3*4*5	0.0057	2	0.0028	0.96	0.382859
1*2*3*4*5	0.0049	4	0.0012	0.41	0.800060
Error	20.1343	6840	0.0029		

Raw Residuals Tasks/sec

Histogram of Raw Residuals
Dependent variable: task/s
(Analysis sample)
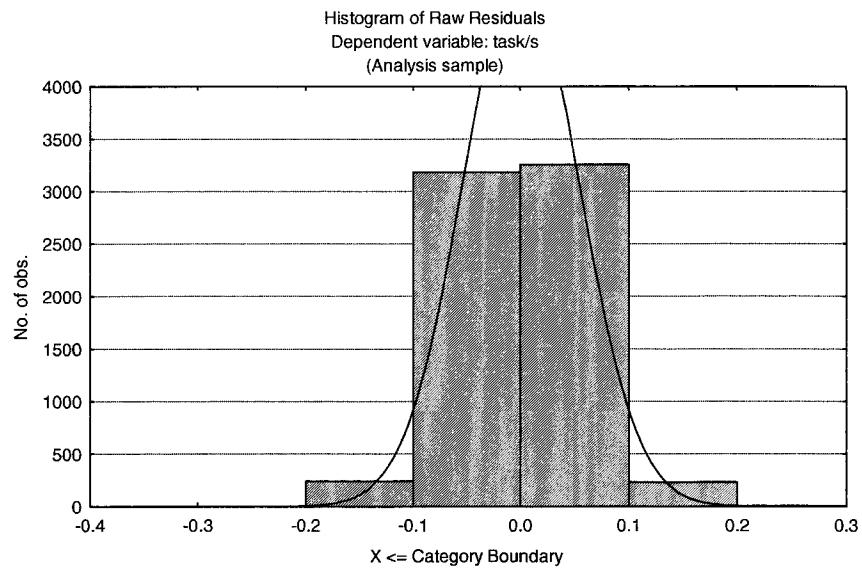
APPENDIX C:

ANALYSIS AND PROCEDURE DETAILS OF CHAPTER 4

Balanced test order design

	Session 1			Session 2		
subject	device 1	device 2	device 3	device 1	device 2	device 3
1	BIG QWERTY	mini QWERTY	STD QWERTY	indestructible	BIG ABC	TS mini
2	indestructible	BIG QWERTY	TS mini	mini QWERTY	STD QWERTY	BIG ABC
3	TS mini	STD QWERTY	BIG QWERTY	BIG ABC	mini QWERTY	indestructible
4	STD QWERTY	BIG QWERTY	mini QWERTY	TS mini	indestructible	BIG ABC
5	TS mini	indestructible	BIG QWERTY	BIG ABC	mini QWERTY	STD QWERTY
6	BIG QWERTY	TS mini	STD QWERTY	indestructible	BIG ABC	mini QWERTY
7	mini QWERTY	STD QWERTY	BIG QWERTY	BIG ABC	TS mini	indestructible
8	BIG QWERTY	TS mini	indestructible	STD QWERTY	BIG ABC	mini QWERTY
9	STD QWERTY	BIG QWERTY	TS mini	mini QWERTY	indestructible	BIG ABC
10	STD QWERTY	BIG ABC	mini QWERTY	TS mini	indestructible	BIG QWERTY
11	mini QWERTY	indestructible	BIG ABC	BIG QWERTY	TS mini	STD QWERTY
12	BIG ABC	TS mini	indestructible	STD QWERTY	BIG QWERTY	mini QWERTY
13	mini QWERTY	STD QWERTY	BIG ABC	BIG QWERTY	TS mini	indestructible
14	BIG ABC	mini QWERTY	indestructible	STD QWERTY	BIG QWERTY	TS mini
15	indestructible	BIG ABC	TS mini	mini QWERTY	STD QWERTY	BIG QWERTY
16	BIG ABC	mini QWERTY	STD QWERTY	indestructible	BIG QWERTY	TS mini
17	Indestructible	BIG ABC	mini QWERTY	TS mini	STD QWERTY	BIG QWERTY
18	TS mini	indestructible	BIG ABC	BIG QWERTY	mini QWERTY	STD QWERTY

Analysis Details Gross Typing Speed

	SS	Degr. of	MS	F	p
Univariate Tests of Significance for Gross Typing speed in WPM (data.sta) Sigma-restricted parameterization Effective hypothesis decomposition					
Intercept	2217901	1	2217901	19335.53	0.000000
{1}Age categorized	44433	2	22217	193.68	0.000000
{2}Typist	16072	1	16072	140.12	0.000000
{3}tractor	2723	1	2723	23.74	0.000001
{4}layout	303458	5	60692	529.11	0.000000
Age categorized*Typist	12234	2	6117	53.33	0.000000
Age categorized*tractor	380	2	190	1.66	0.191133
Typist*tractor	569	1	569	4.96	0.026031
Age categorized*layout	11843	10	1184	10.32	0.000000
Typist*layout	53920	5	10784	94.01	0.000000
tractor*layout	1365	5	273	2.38	0.036476
Age categorized*Typist*tractor	356	2	178	1.55	0.211980
Age categorized*Typist*layout	3376	10	338	2.94	0.001101
Age categorized*tractor*layout	1329	10	133	1.16	0.314124
Typist*tractor*layout	858	5	172	1.50	0.187434
1*2*3*4	423	10	42	0.37	0.960141
Error	363389	3168	115		

Raw Residuals Gross Typing Speed

Histogram of Raw Residuals
Dependent variable: Gross Typing speed in WPM
(Analysis sample)
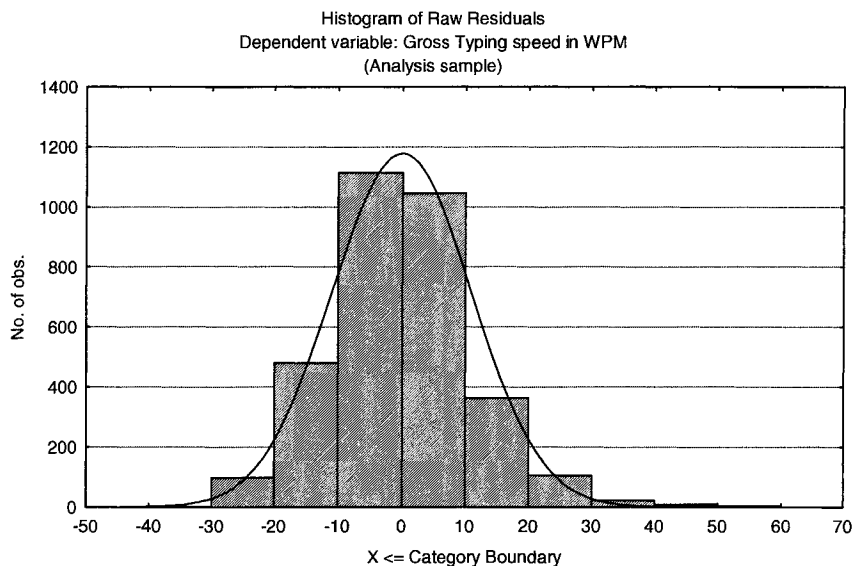
Analysis Details Accuracy

	SS	Degr. of	MS	F	p
Univariate Tests of Significance for Accuracy (data.sta) Sigma-restricted parameterization Effective hypothesis decomposition					
Intercept	1478.605	1	1478.605	53536.39	0.000000
{1}Age categorized	0.344	2	0.172	6.23	0.001989
{2}Typist	0.015	1	0.015	0.55	0.458860
{3}tractor	0.633	1	0.633	22.93	0.000002
{4}layout	6.572	5	1.314	47.59	0.000000
Age categorized*Typist	0.587	2	0.293	10.62	0.000025
Age categorized*tractor	0.129	2	0.064	2.33	0.097214
Typist*tractor	0.119	1	0.119	4.32	0.037751
Age categorized*layout	0.182	10	0.018	0.66	0.762041
Typist*layout	0.711	5	0.142	5.15	0.000103
tractor*layout	0.679	5	0.136	4.92	0.000173
Age categorized*Typist*tractor	0.039	2	0.019	0.70	0.497301
Age categorized*Typist*layout	0.756	10	0.076	2.74	0.002348
Age categorized*tractor*layout	0.299	10	0.030	1.08	0.371153
Typist*tractor*layout	0.297	5	0.059	2.15	0.056884
1*2*3*4	0.637	10	0.064	2.31	0.010732
Error	87.496	3168	0.028		

Raw Residuals Accuracy

Histogram of Raw Residuals
Dependent variable: Accuracy
(Analysis sample)
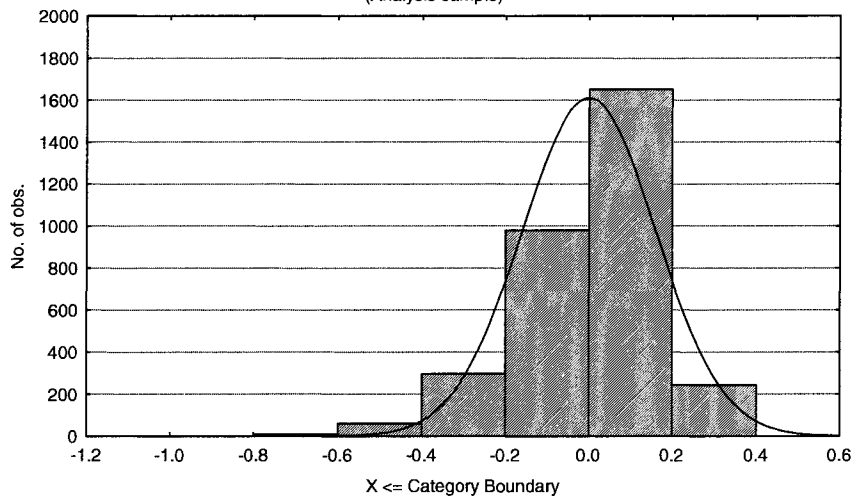
Analysis Details Net Typing Speed

	SS	Degr. of	MS	F	p
Univariate Tests of Significance for Net typing speed WPM (data.sta) Sigma-restricted parameterization Effective hypothesis decomposition					
Intercept	1671861	1	1671861	11091.28	0.000000
{1}Age categorized	29384	2	14692	97.47	0.000000
{2}Typist	12994	1	12994	86.20	0.000000
{3}tractor	4247	1	4247	28.18	0.000000
{4}layout	264829	5	52966	351.38	0.000000
Age categorized*Typist	10083	2	5041	33.44	0.000000
Age categorized*tractor	792	2	396	2.63	0.072419
Typist*tractor	849	1	849	5.63	0.017687
Age categorized*layout	8608	10	861	5.71	0.000000
Typist*layout	55301	5	11060	73.37	0.000000
tractor*layout	2536	5	507	3.36	0.004925
Age categorized*Typist*tractor	92	2	46	0.30	0.737877
Age categorized*Typist*layout	4011	10	401	2.66	0.003084
Age categorized*tractor*layout	2323	10	232	1.54	0.118447
Typist*tractor*layout	1463	5	293	1.94	0.084397
1*2*3*4	1592	10	159	1.06	0.393141
Error	477533	3168	151		

Raw Residuals Net Typing Speed

Histogram of Raw Residuals
Dependent variable: Net typing speed WPM
(Analysis sample)
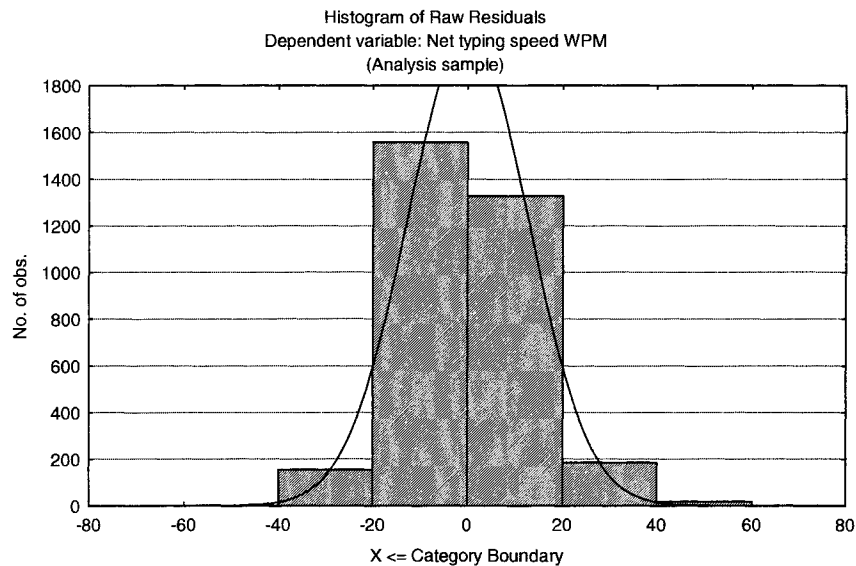
APPENDIX D:

ANALYSIS AND PROCEDURE DETAILS OF CHAPTER 5

Balanced test order design

subject	Session 1			Session 2			Session 3		
1	Mouse	TS stylus	TS finger	TS finger	Mouse	TS stylus	TS stylus	TS finger	Mouse
2	Mouse	TS finger	TS stylus	TS stylus	Mouse	TS finger	TS finger	TS stylus	Mouse
3	TS stylus	Mouse	TS finger	TS finger	TS stylus	Mouse	Mouse	TS finger	TS stylus
4	TS stylus	TS finger	Mouse	Mouse	TS stylus	TS finger	TS finger	Mouse	TS stylus
5	TS finger	Mouse	TS stylus	TS stylus	TS finger	Mouse	Mouse	TS stylus	TS finger
6	TS finger	TS stylus	Mouse	Mouse	TS finger	TS stylus	TS stylus	Mouse	TS finger
7	Mouse	TS stylus	TS finger	TS finger	Mouse	TS stylus	TS stylus	TS finger	Mouse
8	Mouse	TS finger	TS stylus	TS stylus	Mouse	TS finger	TS finger	TS stylus	Mouse
9	TS stylus	Mouse	TS finger	TS finger	TS stylus	Mouse	Mouse	TS finger	TS stylus
10	TS/ stylus	TS finger	Mouse	Mouse	TS stylus	TS finger	TS finger	Mouse	TS stylus
11	TS/ finger	Mouse	TS stylus	TS stylus	TS finger	Mouse	Mouse	TS stylus	TS finger
12	TS/ finger	TS stylus	Mouse	Mouse	TS finger	TS stylus	TS stylus	Mouse	TS finger

Analysis Details Selection time

	SS	Degr. of	MS	F	p
Intercept	1.507649E+09	1	1.507649E+09	574.5181	0.000000
pointer	1.418194E+07	2	7.090968E+06	2.7021	0.067369
age cat	3.733828E+07	2	1.866914E+07	7.1142	0.000840
hand	4.127564E+07	1	4.127564E+07	15.7289	0.000076
Computer usage	4.684503E+07	4	1.171126E+07	4.4628	0.001373
TS exp.	6.087301E+07	2	3.043651E+07	11.5984	0.000010
pointer*age cat	1.836007E+07	4	4.590019E+06	1.7491	0.136657
pointer*hand	9.007165E+06	2	4.503583E+06	1.7162	0.180085
pointer*Computer usage	1.883371E+07	8	2.354213E+06	0.8971	0.517950
pointer*TS exp.	3.119733E+06	4	7.799332E+05	0.2972	0.879884
Error	4.172474E+09	1590	2.624198E+06		

Univariate Tests of Significance for click time (newdata 4th_02_13.sta) Sigma-restricted parameterization Effective hypothesis decomposition Exclude condition: v3 ='office'

Raw Residuals Selection Time

Histogram of Raw Residuals
Dependent variable: click time
(Analysis sample)
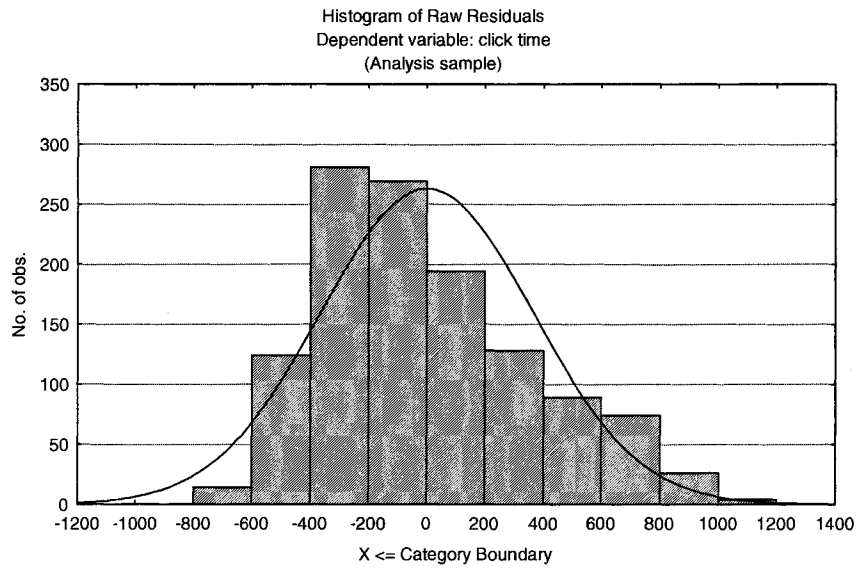
Analysis Details Net Typing Speed

Univariate Tests of Significance for NTS (newdata 4th_02_13.sta) Sigma-restricted parameterization
Effective hypothesis decomposition Exclude condition: v3 ='office'

	SS	Degr. of	MS	F	p
Intercept	940021.5	1	940021.5	14452.49	0.000000
session	715.0	2	357.5	5.50	0.004179
Typing Style	8365.8	1	8365.8	128.62	0.000000
Computer usage	19458.8	4	4864.7	74.79	0.000000
typing exp.	5901.9	2	2950.9	45.37	0.000000
session*Typing Style	216.5	2	108.3	1.66	0.189630
session*Computer usage	664.6	8	83.1	1.28	0.250987
session*typing exp.	259.6	4	64.9	1.00	0.407515
Error	103807.3	1596	65.0		

Raw Residuals Net Typing Speed

Histogram of Raw Residuals
Dependent variable: NTS
(Analysis sample)
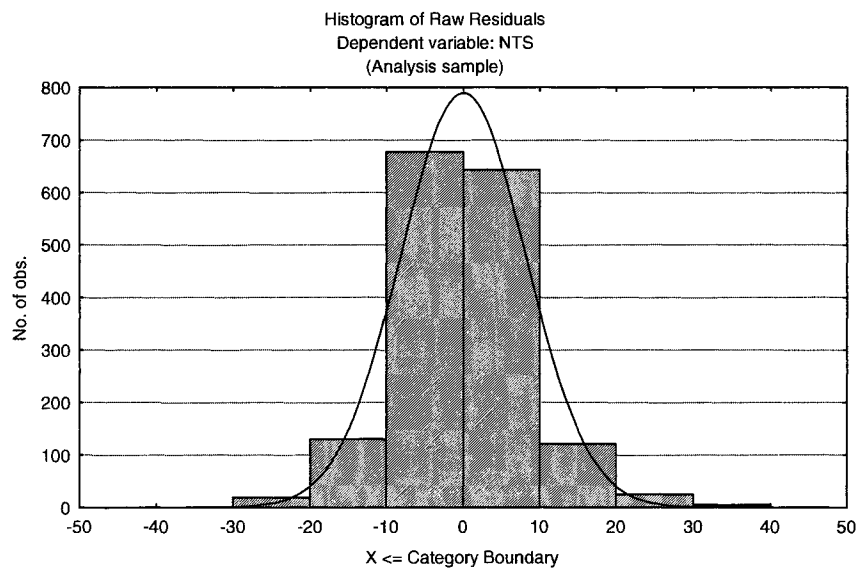
Analysis Details Change1

Univariate Tests of Significance for change1 (4th study questionaire categorized.sta) Sigma-restricted parameterization Effective hypothesis decomposition Exclude condition: v3 ='office' OR v20>6000

	SS	Degr. of	MS	F	p
Intercept	4.704240E+09	1	4.704240E+09	8563.771	0.000000
age cat	2.887109E+07	2	1.443555E+07	26.279	0.000000
hand	7.932296E+07	1	7.932296E+07	144.402	0.000000
Typing Style	1.330503E+07	1	1.330503E+07	24.221	0.000001
Computer usage	6.305892E+07	4	1.576473E+07	28.699	0.000000
Error	8.679236E+08	1580	5.493187E+05		

Raw Residuals Change 1

Histogram of Raw Residuals
Dependent variable: change1
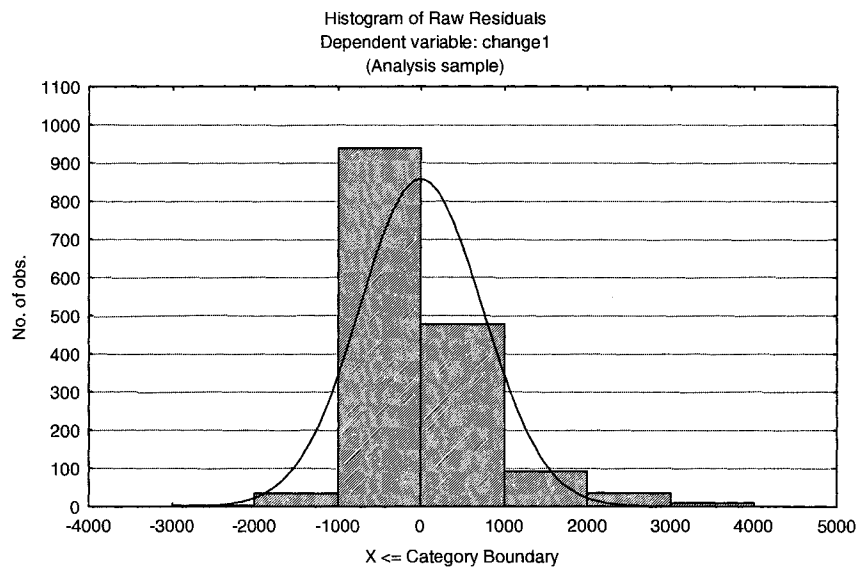(Analysis sample)

X <= Category Boundary

Analysis Details Change2

Univariate Tests of Significance for click/change2 (4th study questionaire categorized.sta) Sigma-restricted parameterization Effective hypothesis decomposition Exclude condition: v3 ='office' OR v22>10000

	SS	Degr. of	MS	F	p
Intercept	2.229802E+09	1	2.229802E+09	3033.120	0.000000
session	2.031644E+07	2	1.015822E+07	13.818	0.000001
pointer	1.296401E+08	2	6.482005E+07	88.172	0.000000
age cat	1.020860E+06	2	5.104299E+05	0.694	0.499564
hand	2.639690E+06	1	2.639690E+06	3.591	0.058284
Typing Style	6.931315E+06	1	6.931315E+06	9.428	0.002172
Computer usage	8.064686E+07	4	2.016171E+07	27.425	0.000000
typing exp.	4.604203E+06	2	2.302101E+06	3.131	0.043921
Error	1.177712E+09	1602	7.351512E+05		

Raw Residuals Change2

Histogram of Raw Residuals
Dependent variable: click/change2
(Analysis sample)
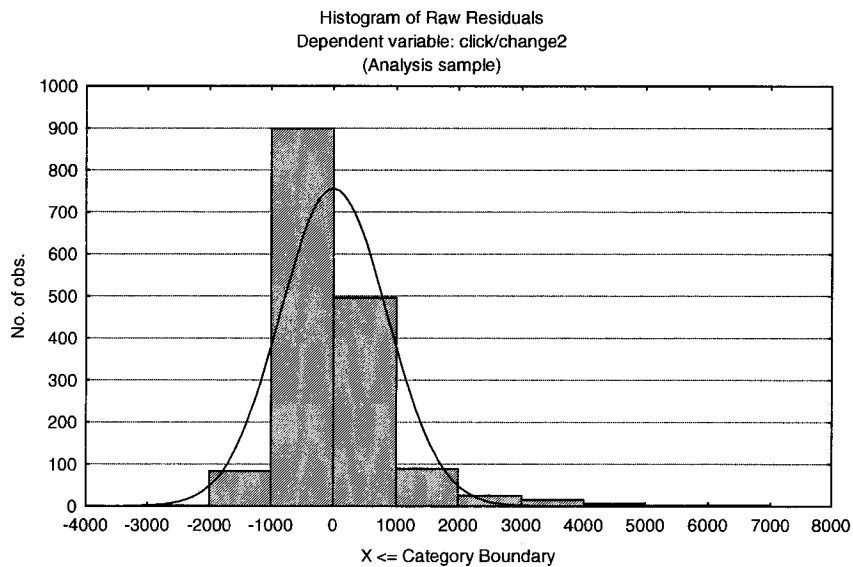
Analysis Details Tasks Time 2

Univariate Tests of Significance for time T2 (newdata 4th_02_13.sta) Sigma-restricted parameterization Effective hypothesis decomposition Exclude condition: v3 ='office'

	SS	Degr. of	MS	F	p
Intercept	6.956229E+09	1	6.956229E+09	1323.426	0.000000
session	8.693579E+06	2	4.346790E+06	0.827	0.437556
pointer	6.164140E+07	2	3.082070E+07	5.864	0.002903
pulldown	3.943276E+07	4	9.858189E+06	1.876	0.112187
cascade	1.403632E+07	4	3.509080E+06	0.668	0.614500
button	1.167797E+08	4	2.919494E+07	5.554	0.000193
age cat	1.323356E+08	2	6.616779E+07	12.588	0.000004
hand	1.851906E+08	1	1.851906E+08	35.233	0.000000
Computer usage	1.323420E+08	4	3.308551E+07	6.295	0.000050
mouse exp.	7.543189E+05	1	7.543189E+05	0.144	0.704867
TS exp.	3.272361E+08	2	1.636181E+08	31.128	0.000000
Error	8.373171E+09	1593	5.256228E+06		

Raw Residuals Task Time 2

Histogram of Raw Residuals
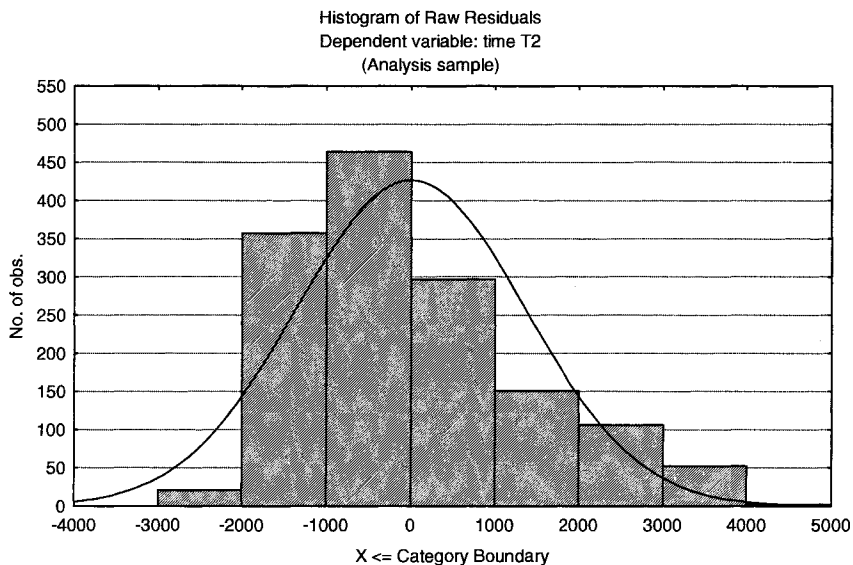Dependent variable: time T2
(Analysis sample)

Analysis Details Tasks Time 3

Univariate Tests of Significance for time T3 (newdata 4th_02_13.sta) Sigma-restricted parameterization Effective hypothesis decomposition Exclude condition: v3 ='office'

	SS	Degr. of	MS	F	p
Intercept	5.901759E+09	1	5.901759E+09	1435.072	0.000000
session	5.831934E+06	2	2.915967E+06	0.709	0.492272
pointer	2.494257E+08	2	1.247128E+08	30.325	0.000000
number	4.174359E+08	47	8.881614E+06	2.160	0.000012
age cat	3.237254E+08	2	1.618627E+08	39.359	0.000000
compu	2.730616E+08	4	6.826540E+07	16.599	0.000000
hand	3.290177E+08	1	3.290177E+08	80.004	0.000000
mouse exp.	3.909851E+05	1	3.909851E+05	0.095	0.757867
TS exp.	2.141731E+08	2	1.070865E+08	26.039	0.000000
Error	6.407302E+09	1558	4.112517E+06		

Raw Residuals Task Time 3

Histogram of Raw Residuals
Dependent variable: time T3
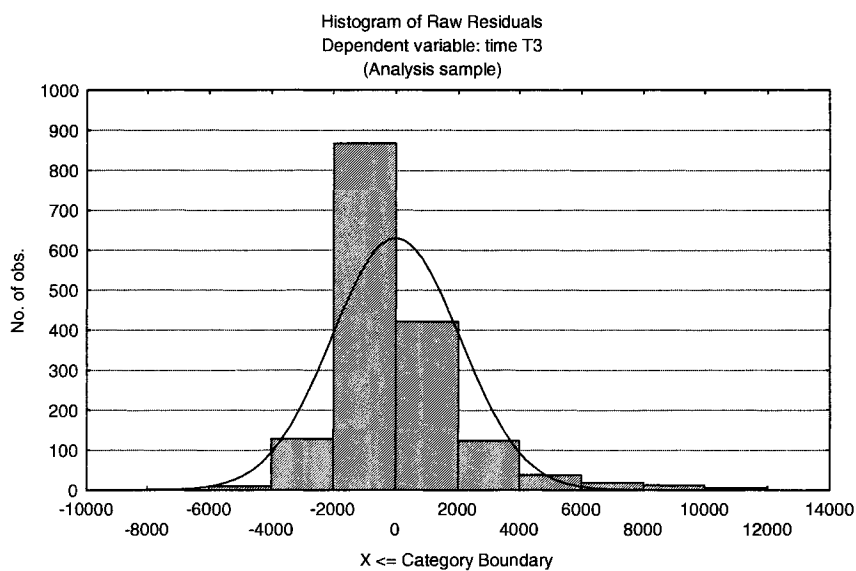(Analysis sample)

X <= Category Boundary

Analysis Details Tasks Time T4

Univariate Tests of Significance for time T4 (newdata 4th_02_13.sta) Sigma-restricted parameterization Effective hypothesis decomposition Exclude condition: v3 ='office'

	SS	Degr. of	MS	F	p
Intercept	3.753747E+09	1	3.753747E+09	1349.623	0.000000
session	5.007957E+04	2	2.503979E+04	0.009	0.991038
pointer	3.891503E+08	2	1.945752E+08	69.958	0.000000
box	1.456284E+08	7	2.080406E+07	7.480	0.000000
age cat	2.279947E+08	2	1.139974E+08	40.987	0.000000
compu	1.804702E+08	4	4.511756E+07	16.222	0.000000
hand	1.931021E+07	1	1.931021E+07	6.943	0.008497
mouse exp.	4.393911E+07	1	4.393911E+07	15.798	0.000074
TS exp.	1.072172E+08	2	5.360860E+07	19.274	0.000000
Error	4.444567E+09	1598	2.781331E+06		

Raw Residuals Task Time 4

Histogram of Raw Residuals
Dependent variable: time T4
(Analysis sample)