Enhancing the Performance of Eye and Head Mice: A Validated Assessment Method and an Investigation into the Performance of Eye and Head Based Assistive Technology Pointing Devices

by

Richard E. A. Bates

Thesis submitted in Partial Fulfilment of the Requirements of De Montfort University for the Degree of Doctor of Philosophy

School of Computing Sciences

May 2006

Enhancing the Performance of Eye and Head Mice: A Validated Assessment Method and an Investigation into the Performance of Eye and Head Based Assistive Technology Pointing Devices

by

Richard E. A. Bates

ABSTRACT

This work poses the question "Could eye and head based assistive technology device interaction performance approach that of basic hand mouse interaction?" To this aim, the work constructs, validates, and applies a detailed and comprehensive pointing device assessment method suitable for assistive technology direct pointing devices, it then uses this method to add enhancement to these devices, finally it then demonstrates that such enhanced eye or head based pointing can approach that of basic hand mouse interaction and be a viable and usable interaction method for people with high-level motor disabilities.

Eye and head based pointing devices, or eye and head mice, are often used by high-level motor disabled people to enable computer interaction in the place of a standard desktop hand mouse. The performance of these eye and head mice pointing devices when used for direct manipulation on a standard graphical user interface has generally been regarded as poor in comparison to that of a standard desktop hand mouse, thus putting users of head and eye mice at a disadvantage when interacting with computers.

The performance of eye and head based pointing devices during direct manipulation on a standard graphical user interface has not previously been investigated in depth, and the reasons why these devices seem to demonstrate poor performance have not been determined in detail. Few proven methods have been demonstrated and investigated that enhance the performance of these devices based on their performance during direct manipulation. Importantly, and key to this work is that, no validated assessment method has been constructed to allow such an investigation.

This work seeks to investigate the performance of eye and head based pointing devices during direct manipulation by constructing and verifying a test method suitable for the detailed performance assessment of eye and head based assistive technology pointing devices. It then uses this method to determine the factors influencing the performance of

eye and head mice during direct manipulation. Finally, after identifying these factors, this work hypothesises, and then demonstrates that applying suitable methods for addressing these factors can result in enhanced performance for eye and head mice. It shows that the performance of these enhanced devices can approach the performance of standard desktop hand mice with the use of highly experienced users, together with the enhancement of a supporting modality for object manipulation, and a supporting interface enhancement for object size magnification; thus demonstrating that these devices can approach and equal the performance of basic hand mouse interaction.

Contents

List of figures

List of tables

List of equations

List of experiments

Acknowledgements

I gratefully acknowledge the following people who have supported this work in many, often varied, ways. Howell for being a good friend and marvellous supervisor, Ian for being a good friend and patient employer, Ursula and Ray for their ever present support, love and encouragement over all this time, Geordie, Cindy and Chloe for being groovy, and all my chums who have had to listen to me complaining about PhD work, and to whom I now owe a few beers.

Declaration

The work described in this thesis is the original work of the author except where specific reference or acknowledgement is made to the work or contribution of others. Some of the material in this work has appeared previously in publications and presentations by the author or to which the author has contributed. In all cases, only the contribution made by the author has been included in this work unless specific reference to the contrary has been made.

Material has been drawn from the following sources in particular:

Bates, R. (1999) Multimodal Eye-based Interaction for Zoomed Target Selection on a Standard Graphical User Interface. Proceedings of Interact'99 vol II pages 7-8. British Computer Society.

Bates, R., Bierton, R. (2000) Experimental Determination of Quantifiers for Usability Questionnaire Design. Proceedings of Human Computer Interaction 2000 vol II pages 57-58. British Computer Society.

Bierton, R., Bates, R. (2000a) Experimental Determination of Optimal Scales for Usability Questionnaire Design. Proceedings of Human Computer Interaction 2000 vol II pages 55-56. British Computer Society.

Bierton, R., Bates, R. (2000b) A Method for Classifying Pointing Device Interaction Errors on Standard GUI Widgets. Proceedings of Human Computer Interaction 2000 vol II pages 61-62. British Computer Society.

Bates, R. (2000) Using Soft Devices as Enabling Translators - Standard GUI Interaction For All! Proceedings of Human Computer Interaction 2000 vol II pages 53-54. British Computer Society.

Bates, R. (2001) Have Patience With Your Eye Mouse! Proceedings of RAATE 2001 (Recent Advances in Assistive Technology and Engineering), 5th-6th November 2001, Birmingham, UK.

Bates, R. (2002a) A Computer Input Device Selection Methodology for Users with High-Level Spinal Cord Injuries - Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Trinity Hall, University of Cambridge 25th-27th March 2002.

Bates, R. (2002b) Have patience with your eye mouse! Eye-gaze interaction with computers can work- Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Trinity Hall, University of Cambridge 25th-27th March 2002.

Bates, R. Istance, H.O. (2002a) Why are eye mice unpopular? A detailed comparison of head and eye controlled assistive technology pointing devices- Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Trinity Hall, University of Cambridge 25th-27th March 2002.

Bates, R. Istance, H.O. (2002b) Zooming interfaces! Enhancing the performance of eye controlled pointing devices - Proceedings of ASSETS 2002, The Fifth International ACM SIGCAPH Conference on Assistive Technologies, 8th–10th July 2002, Edinburgh, Scotland.

Bates, R. (2002c) Eye-gaze interaction with computers can work - Proceedings of the Scientific meeting of Institute of Physics in Engineering and Medicine (IPEM) 9th-12th September 2002, Durham.

Bates, R. Istance, H.O. (2003) Why are eye mice unpopular? A detailed comparison of head and eye controlled assistive technology pointing devices – Journal of the Association for the Advancement of Assistive Technology in Europe (AAATE): Universal Access in the Information Society Volume 2, Number 3, October 2003 Pages: 280 - 290 Special Issue on "Countering Design Exclusion", guest-edited by Simeon Keates and John Clarkson, Springer-Verlag Heidelberg ISSN: 1615-5289.

Bates, R. Istance, H.O. (2004) Towards eye based virtual environment interaction for users with high-level motor disabilities. International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT), pages 275-282, 20th-22nd September 2004, New College, Oxford, UK, ISBN 0704911442.

Bates, R., Istance, H.O., Donegan, M., Oosthuizen, L. (2005) Fly Where You Look: Enhancing Gaze Based Interaction in 3D Environments. In Proceedings of HCI International 2005, Volume 7 - Universal Access in HCI: Exploring New Interaction Environments, Caesars Palace, Las Vegas, USA, July 22-27 2005.

Donegan, M., Oosthuizen, L., Bates, R., Daunys, G., Hansen, J. P., Signorile, M. I., & Majaranta, P. (2005) Providing eye control for those who need it most - a study on user requirements. In Proceedings of ECEM13 - 13th European Conference on Eye Movements, Bern, Switzerland, August 2005.

Chapter 1

Background and Introduction

1.1 Background

This research was conceived from a meeting with Howell Istance at De Montfort University some years ago where Howell demonstrated, and enthused about, eye tracking as a means of enabling computer interaction for people with high-level motor disabilities.

The demonstration was not particularly successful, and after perhaps half an hour I still had difficulty in using the system at all. However, the potential of the system was clear and I sought, under the supervision of Howell, to conduct research into methods of enhancing the performance and usability of eye tracking as a means of enabling interaction with computer interfaces. This thesis is a product of this work.

1.2 Scope of this work

Briefly experimenting with eye tracking equipment showed that eye based pointing was a difficult and challenging modality for computer interaction. Although it appeared to be a natural form of pointing, as the eye naturally 'points' at objects of interest, there were difficulties in pointing accuracy and object manipulation. Further investigation, from anecdotal evidence from rehabilitation centres and literature in the field, found that eye based direct interaction was unpopular and deemed as 'difficult', and that eye based pointing tended almost exclusively to be applied to very specialised interfaces designed specifically for the modality, or to standard graphical user interfaces but via some secondary on-screen device that only allowed indirect, and typically cumbersome, interaction with the interface. Both of these approaches appeared to lose the benefits of direct manipulation in their efforts to overcome the perceived problems of eye based pointing, and by doing so greatly reduced any potential performance advantages eye based pointing may have given users.

From this, there was a clear need to investigate why direct interaction with graphical user interfaces was not done, and how it might be possible to enable or enhance this direct interaction with eye based pointing. This would offer the possibility for people who use eye pointing to interact directly with standard graphical user interfaces, and hence use all of the applications and functionality everyone else enjoys without the encumbrance of indirect interaction or specialised interfaces. Hence such an investigation into the performance of eye based pointing during direct interaction on a standard graphical user interface became the focus of this work.

A method of investigating and assessing the performance of eye based pointing during 'real world' use was now required. A search for suitable 'real world' assessment methods did not reveal any suitable candidates, with diverse and specialised tasks being found that only reflected elements of using eye based pointing during limited specific tasks. The methods found also did not appear to give much insight into ways of enhancing the performance of eye based pointing to enable usable direct interaction. With no suitable existing assessment scheme found, a new method of investigating and assessing the performance of eye based pointing was required.

It was clear that constructing and applying a method to investigate eye based pointing in isolation would not place the performance of the device in any context. Finding some method of improving the performance of an eye mouse, by say 20%, would have little meaning if the overall performance were still very poor. A comparative approach was needed, hence the bounds of the work were expanded to encompass head pointing, the closest direct pointing alternative to eye pointing for people with high-level motor disabilities, and also to standard desktop hand mouse pointing to give a known and familiar baseline for the work.

1.3 Aim of this work

This work seeks to answer the question "Could eye and head based assistive technology device interaction performance approach that of basic hand mouse interaction?"

In order to answer this question, the main elements of this research were broken down into a set of aims and outcomes:

- To construct a structured test method suitable for the detailed objective and subjective performance assessment of eye and head based assistive technology pointing devices during direct interaction on a standard graphical user interface.

- To verify that the above structured test method is suitable for assessing inaccurate direct pointing devices such and eye and head based pointing devices.

- To verify that the above structured test method can reveal where performance improvements may be achieved.

- To verify the above structured test method against a known baseline both in terms of its range and its sensitivity.

- To use the results from the verified structured test method to examine the performance of eye based direct interaction with graphical user interfaces and to place this performance in context by comparison with a head based assistive technology device and a standard desktop hand mouse.

- To use the results from the verified structured test method to determine the limiting factors influencing the objective and subjective performance of eye and head based pointing during direct interaction on a standard graphical user interface.

- To use the results from the verified structured test method to hypothesise suitable enhancements to eye and head based pointing to address these performance limiting factors, and hence enhance the objective and subjective performance of head and eye mouse direct interaction on a standard graphical user interface.

- To use the verified structured test method to determine to what extent these enhancements to eye and head based pointing improve the objective and subjective performance of these devices, and to then determine if these enhanced devices can approach the performance of standard desktop hand mice and so offer a viable and usable interaction method for people with high-level motor disabilities.

- To use these aims and outcomes to answer the main research question "Could eye and head based assistive technology device interaction performance approach that of basic hand mouse interaction?"

1.4 Structure of the thesis

This work starts with a general overview to familiarise the reader with the principles of head and eye based interaction, *Chapter 2 "Eye and Head Based Interaction"*. The chapter shows the justification for this work, illustrating example user groups for head and eye based interaction and showing the need for these devices. The chapter then goes on to give a brief overview of the positive and negative issues surrounding the usability and performance of head and eye based pointing interaction with graphical user interfaces. This chapter then briefly examines the range of head and eye mice and associated on-screen keyboards and dwell click software that are available to enable head and eye based direct or indirect interaction with a graphical user interface.

The next step in this work was to determine a method for the evaluation of head and eye based interaction. *Chapter 3, "Assessment Methods"*, marks the start of the construction of such a suitable assessment method for the devices. It soon became clear, as shown in chapter 3, that simple evaluation tests, such as target acquisition, did not give a great insight into why eye, and head, based pointing tended to exhibit poor performance during direct interaction, nor did these tests give much insight into methods for enhancing the performance of these devices to enable usable direct interaction – the aim of this work. In addition, such 'abstract' tests did not give detailed measurements that would accurately reflect the performance, usability issues, and areas for possible performance enhancement of the devices when they were used for 'real world' interaction. A search for suitable 'real world' assessment schemes did not find any suitable candidates; hence a new 'real world' test method suitable for head and eye based direct interaction assessment was required.

To construct a suitable test method, *Chapter 4, "Constructing Real-World Tasks"*, details the construction of a set of 'real world' test tasks for the assessment of hand, head and eye mice. This is followed by *Chapter 5 "Measuring the Performance of Pointing Device Interaction"* that goes on to discuss and determine methods of assessing the objective and subjective performance of the devices when performing these test tasks.

A suitable subjective assessment questionnaire scheme for the devices was not available for assessing user reaction to the devices; hence *Chapter 6 "Assessment Scales"* and *Chapter 7 "Choice of Scales"* show the construction of a questionnaire scheme for the assessment method. Finally, in order to be sure the method would give valid and reliable results, *Chapter 8 "Validation of the Assessment Method"* completes the assessment method by performing a detailed validation of the method.

The next step was to construct suitable head and eye mice for the evaluation. *Chapter 9 "Constructing Eye and Head Mice"* gives a survey of typical head and eye tracking technologies before describing the construction of suitable head and eye mice, and associated text entry and interface object selection systems. These systems were specifically developed or selected to be suitable for assessing the performance of head and eye based interaction as they allowed detailed examination and modification and enhancement of their performance and operating characteristics.

Chapter 10 "The Performance of Hand, Head and Eye Mice" marks the start of the device assessment and shows the performance results of the devices on the assessment method. Here, the performances, both objective and subjective, of head and eye mice are assessed against the baseline performance of a standard desktop hand mouse. It was found that both head and eye mice showed poorer performance on the method than the baseline hand mouse, and that performance with eye mice, in general, was poorer than head mice.

These results are further examined in greater detail in *Chapter 11 "A Detailed Examination of Eye and Head Mouse Performance"*, which attempts to reveal why these devices showed poor performance. Here, typically, eye mice were found to exhibit rapid but inaccurate pointing, and head mice slower but slightly more accurate pointing. It was also found that eye mice typically exhibited very long learning times to become proficient with the device in comparison to head mice.

After identifying these factors, or operational characteristics, of head and eye mice, *Chapter 12 "Enhancing Eye and Head Mouse Performance"* proposes, and then shows, the construction of software enhancements for head and eye mice to reduce the effects of these limiting factors. Little could be done to address the pointing speed of the devices, as this was a product of the movement properties of the human head or eye, but the pointing accuracy of the devices could be enhanced by increasing interface target sizes, and the learning times of the eye mouse could be examined by isolating results by the hours of experience of users with the device.

Chapter 13 "The Effect of Enhanced Eye and Head Mice" marks the start of the enhanced device assessment and shows the performance results of the enhanced devices on the assessment method. Here, the performances, both objective and subjective, of the standard and enhanced head and eye mice are compared and contrasted both against each other and against the baseline performance of a standard desktop hand mouse. From these results, it was found that the enhancement benefited both the head and eye mice, lifting their performance above the non-enhanced standard devices, although neither enhanced eye or head mice achieved the performance of the hand mouse baseline.

These results are further examined in greater detail in *Chapter 14 "A Detailed Examination of Enhanced Eye and Head Mice"*. This chapter completes the examination of eye and head based pointing and examines the optimal performance that could be achieved by head and eye mice. It shows that eye mice, with enhancement, can exceed the performance of head mice and approach the performance of hand mice on the test method, and proposes that head and eye based pointing can become viable and usable interaction methods for people with high-level motor disabilities.

Finally, *Chapter 15 "Conclusions"* summarises the findings and contribution to knowledge of the work.

In summary, this work covers the construction and validation of a detailed 'real-world' assessment method suitable for eye and head based assistive technology direct pointing devices. The work then uses the method to assess in detail the performance of head and eye based pointing devices against a baseline of a standard desktop hand mouse. Finally, from this assessment, the method is used to reveal methods of enhancing the performance of the head and eye mouse devices. These methods are then implemented and evaluated with the method to determine the improvements made, with the aim of enabling usable direct interaction on an unmodified standard graphical user interface with head, and particularly, eye based pointing.

1.5 Navigating the thesis

This thesis may be read in a linear fashion, but the aim and design has been to enable, or perhaps spare, the reader from reading all Chapters in a linear fashion. Hence, Chapters or themes of little interest may be skipped whilst still preserving the main aims and structure of the work. For instance, after the introductory Chapters 1 and 2, if interested in assessment methods the reader may carry on to Chapter 3 and onward. If more interested in the performance of hand, head and eye mice the reader may jump to and start at Chapter 9, skipping the assessment method all together. In addition, these two main areas can be further divided, for instance Chapters 6 and 7 show the construction and choice of questionnaire scales that may be applied to other subjective assessment schemes, or Chapters 13 and 14 show the application and effect of enhancement to head and eye mice. To this end, some possible paths and shortcuts through the thesis are illustrated (Figure 1.1).

To give the reader a rapid overview of this work, and to aid understanding of the scope and accomplishments of this work, the reader is encouraged to read these introductory chapters (Chapters 1 and 2) and then briefly read the conclusions of this work (Chapter 15).

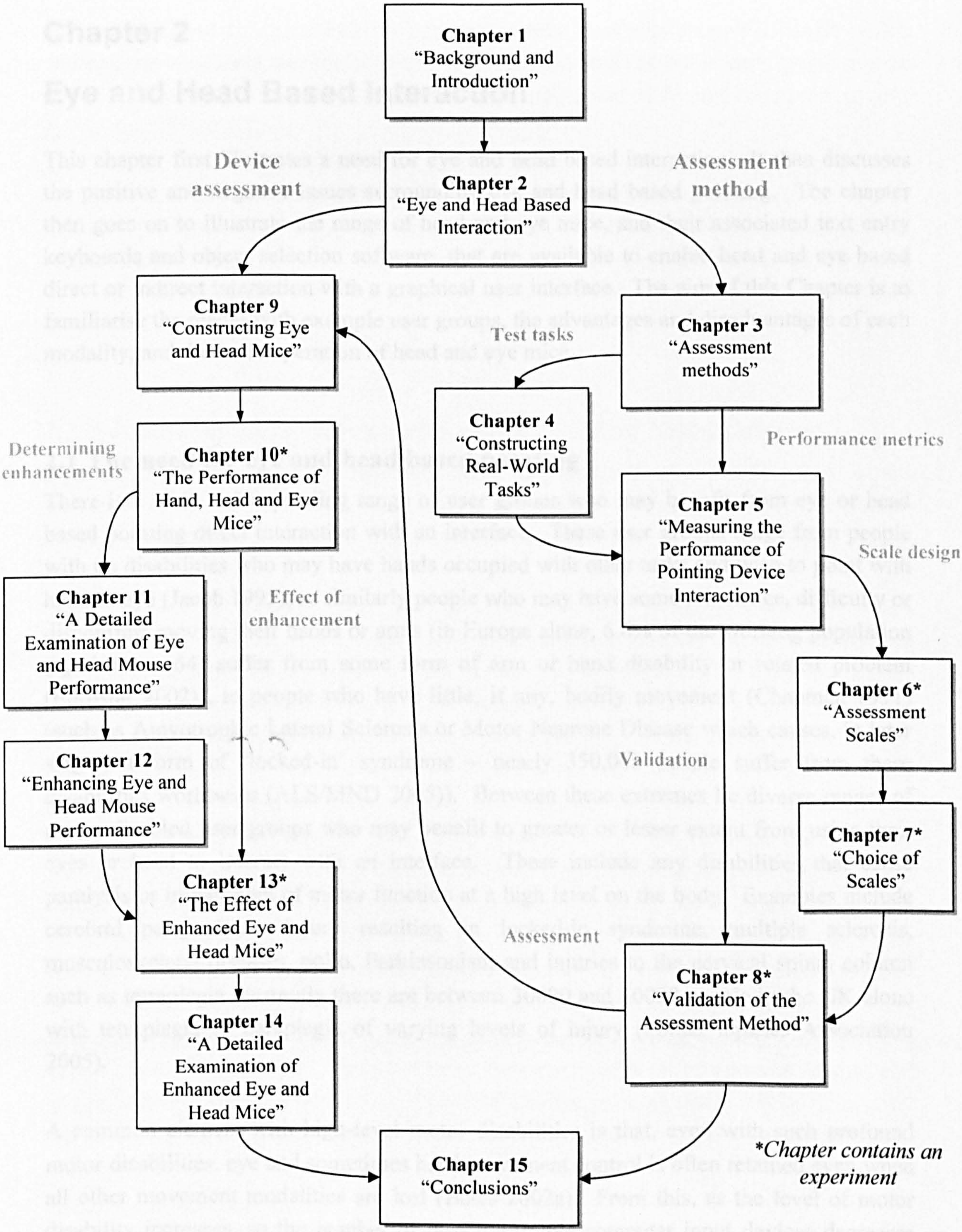


Figure 1.1 Navigating the thesis

Chapter 2

Eye and Head Based Interaction

This chapter first illustrates a need for eye and head based interaction. It then discusses the positive and negative issues surrounding eye and head based pointing. The chapter then goes on to illustrate the range of head and eye mice, and their associated text entry keyboards and object selection software, that are available to enable head and eye based direct or indirect interaction with a graphical user interface. The aim of this Chapter is to familiarise the reader with example user groups, the advantages and disadvantages of each modality, and the basic operation of head and eye mice.

2.1 The need for eye and head based pointing

There is a wide and expanding range of user groups who may benefit from eye or head based pointing direct interaction with an interface. These user groups range from people with no disabilities who may have hands occupied with other tasks and wish to point with head or eye (Jacob 1995), or similarly people who may have some reluctance, difficulty or discomfort moving their hands or arms (in Europe alone, 6.6% of the working population (aged 16 to 64) suffer from some form of arm or hand disability or related problem (Eurostat 2002)), to people who have little, if any, bodily movement (Chapman 1991) (such as Amyotrophic Lateral Sclerosis or Motor Neurone Disease which causes, in later stages, a form of 'locked-in' syndrome – nearly 350,000 people suffer from these disabilities worldwide (ALS/MND 2005)). Between these extremes lie diverse ranges of motor disabled user groups who may benefit to greater or lesser extent from using their eyes or head to interact with an interface. These include any disabilities that cause paralysis or impairment of motor function at a high level on the body. Examples include cerebral palsy, brain injury resulting in locked-in syndrome, multiple sclerosis, musculoskeletal diseases, polio, Parkinsonism and injuries to the cervical spinal column such as tetraplegia (currently there are between 30000 and 40000 people in the UK alone with tetraplegia or paraplegia of varying levels of injury (Spinal Injuries Association 2005).

A common element with high-level motor disabilities is that, even with such profound motor disabilities, eye and sometimes head movement control is often retained even when all other movement modalities are lost (Bates 2002a). From this, as the level of motor disability increases, so the number of possible usable computer input devices decreases dramatically, with the majority of input devices becoming unusable once hand function is lost. As the level of motor disability approaches neck level only a range of single switch devices, some unusual and limited bandwidth[1] devices such as brain activity and muscle EMG, speech and head and eye movement were usable with sufficient bandwidth to give interaction (Bates 2002a).

Of these available modalities, only head and eye movement would give direct control over a pointing device. This is in contrast to, for example, switch and speech based devices that only give indirect control over pointing. Hence head and eye movement offer perhaps the only modalities that can be utilised for direct pointing interaction with a graphical computer interface.

2.2 Advantages and disadvantages of eye and head based pointing

Both head and eye pointing modalities have strengths and weaknesses based on the properties of human head and eye movement and control, the technological performance of head and eye tracking systems, and the interaction requirements of the interface.

Examining the advantages of eye gaze pointing over head pointing it has been stated (MacKenzie et al. 2001, Jacob 1995, Jacob 1991) that eye gaze has the potential to be a very natural and potentially efficient form of pointing, as people tend to naturally look at the object they wish to interact with. This property manifests itself by simply placing the pointing cursor on the interface at the gaze point of the user. As the user searches for, and then locates, an object on the interface so they find that the cursor has followed their eye gaze and is already located on that object. This has been stated as being a very intuitive means of pointing that requires little or no training of the user (Stampe and Reingold 1995). Unlike a cursor driven by the hand or head, placing the cursor at the gaze position eliminates the need for the user to make any further eye movements to locate the previous cursor position, and then to make further eye movements back and forth between cursor and object to steer the cursor onto the desired object. This repeated search is typical of head based pointing, which requires conscious movement and steering of the head to point at an object. The naturalness of using eye movements for pointing is further supported as eye pointing, with invisible cursor, has been shown to exhibit little detectable fatigue, and so eye pointing offers the possibility of near fatigue-free pointing (Saito 1992). This is not the case with head based pointing where continual head movement was found to be

[1] Where *bandwidth* may be defined as the amount of information communicated to the interface per unit time by the modality. For example, a switch generates low bandwidth binary information, a desktop mouse higher bandwidth *x ,y* positional information.

uncomfortable and fatiguing due to the neck muscles tiring (Evans and Blenkhorn 1999, LoPresti et al. 2000b).

In addition to being a natural and potentially sustainable form of pointing, the speed of eye-gaze to locate a target can be very rapid when compared to other pointing devices (Edwards 1998, Jacob 1995, Salvucci and Anderson 2000, MacKenzie 1992). In particular, eye pointing has been shown to be more rapid than traditional pointing devices such as desktop mice provided the target objects are large enough to be easily selected (Ware and Mikaelian 1987, Sibert and Jacob 2000). Human eye movement consists of two basic movements; fixations where the eye gaze position is fairly static and clear vision is possible, and saccades where the eye is moving at high speed ballistically between fixations. During saccades the eye is capable of very high angular velocities (400 to 700 degrees per second (Yarbus 1967), so if target objects can easily be located during fixations then rapid pointing is perhaps unsurprising, as the speed of cursor movement to the targets will be very high during the saccades between objects. This is in contrast with head pointing which can be comparatively slow for target acquisition tasks (Jagacinski and Monk 1985, MacKenzie 1992, Radwin et al. 1990) due to the high mass of the head restricting rapid movement. Also head pointing may be difficult, slow and inaccurate due to restrictions in the range of neck motion (LoPresti et al. 2000a, LoPresti et al. 2000b).

On the basis of the above, it appears that eye based pointing has considerable advantages over head pointing. However, eye gaze has some inherent disadvantages. The eye is not a highly accurate pointing device as it exhibits a positional tolerance (Carpenter 1991). The foveal area of the eye, which gives clear vision, covers a visual angle of approximately 0.5-1° arc of the retina, hence when fixating a target the eye only needs to be within approximately 0.5° visual angle of the target position to potentially clearly see the target[1]. This gives an unavoidable inaccuracy in measured gaze position and this problem is greatly compounded with the inaccuracy of eye gaze tracking devices. Typically eye gaze tracking devices may be quite accurate after calibration but then tend to drift in accuracy (Stampe 1993, Stampe and Reingold 1995), resulting in additional pointing inaccuracies. This drift is often due to head movement after calibration since eye tracking has the additional complexity that eye orientation to the screen is also affected by head orientation to the screen. Hence either the position of the head must also be tracked to the same degree of accuracy as the eye with this movement then compensated for in the tracked eye position, or the head must remain in a fixed position by using a head or chin rest (Figure

[1] Pointing accuracy is measured in degrees visual angle from the head or eye to allow simple calculation of on-screen pointing accuracy for any given seating distance from the screen, for example 0.5° at a distance of 60cm would give an accuracy of approximately 5mm on screen.

2.1), bite-bar, or by default with users who have very high level motor disability resulting in a loss of head movement.

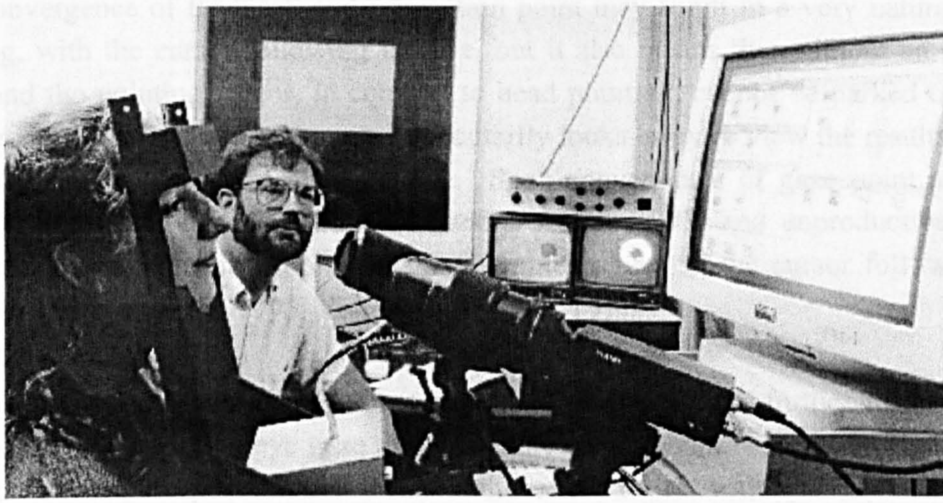


Figure 2.1 Reducing calibration drift with a chin rest[1]

Unlike head position that is under the full control of the user, eye gaze position cannot easily be consciously controlled or steered, as it tends to be driven by subconscious interest (Yarbus 1967). Hence the eye tends to fixate briefly on targets of interest before jumping to other points of interest. This lack of direct conscious pointing control requires effort by the user to point steadily at a target for any extended period of time and is found to be unnatural (Jacob 1991, Hansen et al. 2004). The difficulty of holding a steady gaze position on a target contrasts sharply with the deliberate, if slow, controlled conscious movement and positional accuracy of head based pointing (LoPresti et al. 2000a, LoPresti et al. 2000b, Jagacinski and Monk 1985, MacKenzie 1992).

The lack of easy conscious control over eye gaze position also means that any inaccuracies in tracked cursor position cannot easily be corrected by a slight eye gaze correction or offset. This is in contrast to head based pointing where, if users felt that the head pointing device did not accurately position the cursor where their head was pointing, they could easily compensate for cursor positioning inaccuracies by moving or offsetting their head position to reposition the cursor more accurately (Evans et al. 2000).

---

[1] From work by Howarth et al. 1992, www.lboro.ac.uk/departments/hu/groups/viserg/eyecon1.htm

In addition to exhibiting an inherent tracking inaccuracy and difficulty in conscious control, when used for pointing at objects the eye is being employed as both an input modality to the user, so the person can see the computer interface, and an output modality from the user to the interface, indicating the pointing intention of the user on the interface. This convergence of feedback and interaction point may result in a very natural form of pointing, with the cursor following the eye, but it also means that without an additional command the pointing cursor, in contrast to head pointing, cannot be parked or left at a position on the screen whilst the eye momentarily looks away to view the results of a user command or feedback from the interface. Such convergence of gaze point and cursor results in unwanted and potentially distracting (Jacob 1993) and unproductive pointing movements at the feedback point on the computer screen as the cursor follows the eye wherever is gazes (Jacob 1995, Velichkovsky et al. 1997).

The problem of convergence of interaction and feedback point is further compounded by the inherent inaccuracy of eye gaze tracking. The cursor could be 0.5° visual angle offset from the actual gaze position due to the width of the fovea, with this inaccuracy further compounded by any calibration drift from the gaze tracking equipment. Such unwanted cursor displacement gives a visual distraction on the interface that can cause a vicious circle of eye pointing and cursor feedback with the eye to attempting to follow the cursor, as the eye is subconsciously drawn to objects of interest, and the cursor being displaced by the eye. This results in the eye chasing the cursor on the screen. Simply making the cursor invisible to break this vicious circle may not be possible on interfaces where target objects are smaller than 0.5° visual angle, as the inherent inaccuracy of tracked gaze position may make target selection highly imprecise when the user cannot see if the cursor is on the desired target.

Clearly both head and eye pointing have specific advantages and disadvantages that would affect the performance and acceptability of these modalities when used for pointing. The next step is to examine the methods and properties of head and eye tracking systems that enable interaction with a standard unmodified graphical user interface, and that also attempt to overcome some of the problems inherent in head and eye based pointing.

2.3 Head mouse systems

Head mouse systems operate by simply tracking the head orientation to the screen, and placing the cursor at the point where the vector tracked from the head position intersects the plane of the screen. This pointing action is best described as 'nose following' (Evans and Blenkhorn 1999) where an imaginary line or arrow is projected from the nose toward the screen, with the cursor placed at the intersection of line and screen.

There are a range of freely available head mouse devices which use different technologies to track head position, each with their own characteristics, strengths and weaknesses (Table 2.1). Here the devices were categorised based on data from user and expert assessments[1] and manufacturers' own data and consider the ease of set up of the equipment, the pointing accuracy, and the sampling rate (and hence responsiveness), of the systems. Ratings of *low*, *medium* and *high* are used to indicate the performance of the devices, with high ratings showing perceived better performance or ease of use. The definitions of these ratings were based on the ranges of performance encountered during the search, with a high rating indicating the upper bounds found, and a low rating indicating the lower bounds found. The results are ordered, with more popular devices toward the top of the list (based on the availability of commercial systems, and anecdotal evidence from rehabilitation centres of the type of systems in use). Example manufacturers are given for each technology, with devices illustrated (Figure 2.2).

Typically head mouse devices were designed to be simple, low cost, easy to set up and use, and designed to be 'hands free' and hence have no direct link to the head (this is logical as un-encumbering or un-tethered devices tend to be more practical to use). These design considerations tended to produce devices that did not exhibit a high degree of accuracy, or responsiveness, but did have a moderately high degree of ease of use. However, no one system achieved a *high* rating overall.

There appears, to date, to be little work which evaluates the performance of head mice. However, work on infrared-based devices supported the finding that these devices were only moderately accurate in use due to variations in the irradiation patterns from the infrared emitters used to track the head (Evans and Blenkhorn 1999 Evans et al. 2000), and that ultrasonic devices are considerably (63%) slower in operation than a standard desktop hand mouse (Radwin et al. 1990).

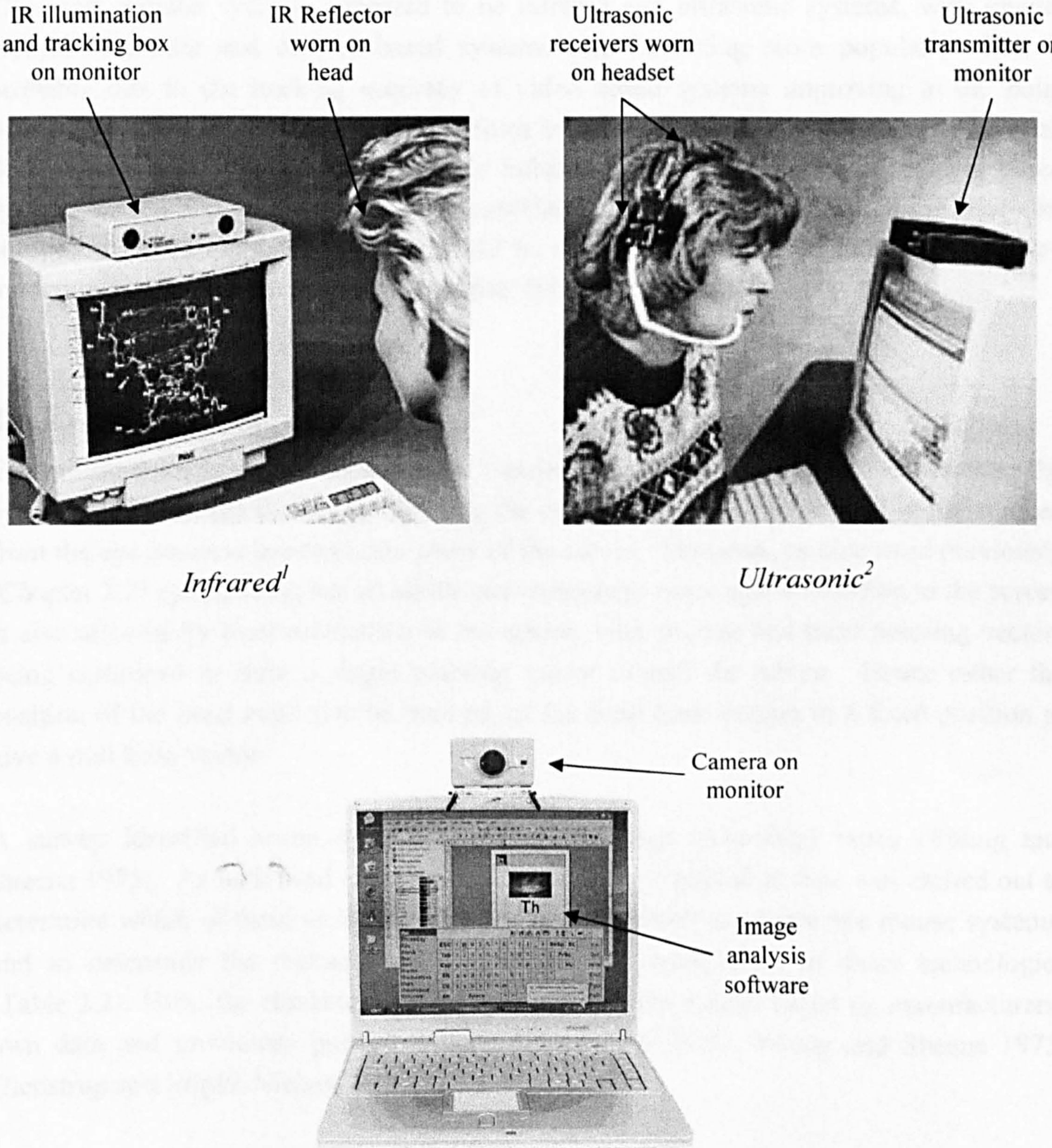
[1] ACE Centre (Aiding Communication in Education), www.ace-centre.org.uk

Categorisation of head pointing technologies						
Technology	*Method of tracking*	*Can be used for direct pointing on a standard GUI?*	*Requires user to wear objects on the head?*	*Ease of set up*	*Accuracy*	*Sampling rate*
Infra Red[1]	Track position of head and sticky reflective dot reference on head	Yes	Yes	High	Medium	High
Ultrasonic[2]	Track sound from ultrasonic transmitters worn on head	Yes	Yes	High	Medium	Medium
Software[3]	Software analysis of camera image of head	Yes	No	Medium	Low - medium	Medium
Gyroscopic[4]	Gyroscopes worn on head give position	Yes	Yes	Medium	High	Medium
Tilt switch[5]	Tilt switches worn on head	No	Yes	Medium	Low	Medium
Laser[6]	Laser pointer on head, light sensitive targets	No	Yes	Medium	High	Low
Optical[7]	Optical light sensor worn on head, light emitting targets	No	Yes	High	Low	Low

Key:			
Rating	Ease of set up (typically)	Accuracy (typically)	Sampling rate (typically)
Low	Requires skilled technical assistance	> 1.0°	< 25 Hz
Medium	Requires some skill and technical assistance	0.5° - 1.0°	25 – 60 Hz
High	Requires some skill but no technical assistance	< 0.5°	> 60 Hz

Table 2.1 Head tracking technologies

[1] HeadMouse from Origin Instruments, www.orin.com, [2] HeadMaster Plus from Prentke Romich, www.prentrom.com, [3] HeadMouse Extreme from Origin Instruments, www.orin.com, [4] Tracer HeadMouse from Boost Technology, www.boosttechnology.com, [5] Headway from Keytools, www.keytools.com, [6] Lucy from Shannon Electronics, [7] Optical Headpointer from Prentke Romich, www.prentrom.com.



Infrared[1]

Ultrasonic[2]

Camera and software[3]

Figure 2.2 Example head mouse systems

[1]HeadMouse from Origin Instruments, www.orin.com, [2] HeadMaster Plus from Prentke Romich, www.prentrom.com, [3] HeadMouse Extreme from Origin Instruments, www.orin.com

The most popular systems appeared to be infrared and ultrasonic systems, with image-analysis software and camera based systems also becoming more popular[1]. This is probably due to the tracking accuracy of video based systems improving to the point where they now can rival or even outperform infrared and ultrasonic systems (Betke et al. 2002, Chen et al. 2003), and that unlike infrared and ultrasonic systems, camera based systems do not require the user to wear anything on the head and do not require specialist equipment as web or USB cameras could be used. Typically, head mouse systems are moderately expensive, with systems ranging from £1000 to £2000.

2.4 Eye mouse systems

Eye mouse systems operate in a similar manner to head tracking systems by tracking the eye orientation to the screen, and placing the cursor at the point where the vector tracked from the eye position intersects the plane of the screen. However, as discussed previously (Chapter 2.2) eye tracking has an additional complexity since eye orientation to the screen is also affected by head orientation to the screen, with the eye and head pointing vectors being combined to form a single pointing vector toward the screen. Hence either the position of the head must also be tracked, or the head must remain in a fixed position to give a null head vector.

A survey identified seven oculography (eye tracking) technology types (Young and Sheena 1975). As with head mouse systems, a search of published data was carried out to determine which of these technologies were used for freely available eye mouse systems, and to determine the characteristics, strengths and weaknesses of these technologies (Table 2.2). Here, the characteristics of the devices were judged based on manufacturers' own data and previously published work (Duchowski 2000, Young and Sheena 1975, Glenstrup and Engell-Nielsen 1995).

The search results categorise the ease of set up of the equipment, the pointing accuracy, and the sampling rate and hence responsiveness, of the systems and an additional factor of the invasiveness of the systems, i.e. do they require objects to be placed in contact with the eye. The inclusion of this factor was felt to be important due to the potential hazards of placing objects on the eye.

[1] Usage experienced at the ACE Centre (Aiding Communication in Education), www.ace-centre.org.uk and the ACT (Access to Communication and Technology), Regional Rehabilitation Centre, Oak Tree Lane Centre, Birmingham.

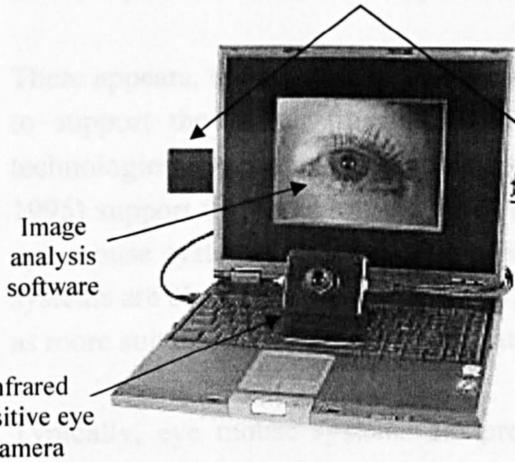
Categorisation of eye tracking technologies						
Technology	*Method of tracking*	*Used as an eye mouse system?*	*Invasive?*	*Ease of set up*	*Accuracy*	*Sampling rate*
Pupil and Corneal reflection[1]	Video tracking of light reflection from the cornea and dark pupil (Video-oculography)	Yes	No	Medium	Medium	Medium
Electro-potential[2]	Measurement of electro-potentials around eye (Electro-oculography)	Yes	No	Medium	Low	High
Pupil[3]	Video tracking of dark pupil (Video-oculography)	None currently available	No	High	Low	Medium
Scleral coil[4]	Electromagnetic tracking of coil inserted in eye	None known	Yes	Low	High	High
Dual Purkinje image[5]	Video tracking of light reflections from the cornea and lens boundary	None known	No	Medium	Medium	High
Limbus[6]	Video tracking of iris-sclera boundary	None known	No	Medium	Low	High
Contact lens[7]	Tracking of light reflected from contact lens inserted in eye	None known	Yes	Low	High	High

Key:			
Rating	Ease of set up (typically)	Accuracy (typically)	Sampling rate (typically)
Low	Requires skilled technical assistance	> 0.5°	< 50 Hz
Medium	Requires some skill and technical assistance	0.1° - 0.5°	50 – 100 Hz
High	Requires some skill but no technical assistance	< 0.1°	> 100 Hz

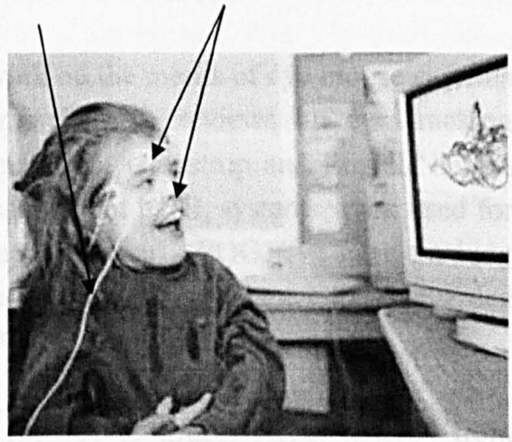
Table 2.2 Eye tracking technologies

[1]MON VOG from MetroVision Systems www.metrovision.fr, [1] Quick Glance from EyeTech Systems www.eyetechds.com, [1]SensoMotoric Instruments www.smi.de, there are numerous similar examples of pupil and corneal reflection, see http://ibs.derby.ac.uk/emed/ for a full list, [2]Eagle Eyes from www.bc.edu, [2] MON EOG from MetroVision Systems, www.metrovision.fr, [3]Vision Control Systems (no longer available), [4]Skalar Medical, www.skalar.nl, [5]Eyetracker 2000 from Forward Optical Technolgies, www.fourward.com, [6]MR Eyetracker from Cambridge Research Systems Ltd, www.crsltd.com, [7]None commercially available.

As with the head mouse systems, ratings of *low*, *medium* and *high* indicate the performance of the devices. As before, the results are ordered, with more popular devices toward the top of the list (based on the availability of commercial systems and the use of these systems as reported in previously published work). Example manufacturers are given for each technology, with example devices illustrated (Figure 2.3).



Video-oculography[1]



Electro-oculography[2]

Figure 2.3 Example eye mouse systems

From the search (Table 2.2), typically eye mouse devices were difficult to set up and use, sometimes invasive, but could offer greater tracking accuracy and higher responsiveness than head mouse systems. There was a trade-off between invasiveness and tracking accuracy, with non-invasive systems having lower accuracy. As with head mice, no one system achieved a *high* rating overall. By far the most popular system found in the search was video oculography using pupil and corneal reflection (VOG). 18 out of 44 eye tracking manufacturers used this system[3]. This popularity may be due to the non-invasive nature and simplicity of the system, with VOG systems using a simple camera and illumination system to determine gaze direction by comparing the pupil position with a

[1] Quick Glance from EyeTech Systems www.eyetechds.com
[2] Eagle Eyes from www.bc.edu
[3] Eye movement equipment manufacturers database, http://ibs.derby.ac.uk/emed

reflection of incident light reflected from the cornea of the eye (Duchowski 2000, Young and Sheena 1975, Glenstrup and Engell-Nielsen 1995). The remaining system found to be available and in use as an eye mouse was electro-potential oculography (EOG), with 3 out of 44 eye tracking manufacturers using this system[1]. EOG eye gaze tracking is simple and is based on electrical measurement of the potential difference between the cornea and the retina (about 1 mV). This potential creates an electrical field in the front of the head that changes orientation in sympathy with gaze direction and can be detected by electrodes placed around, but not in, the eyes (Duchowski 2000, Young and Sheena 1975, Gips et al. 1993, Gips et al. 1996, Glenstrup and Engell-Nielsen 1995).

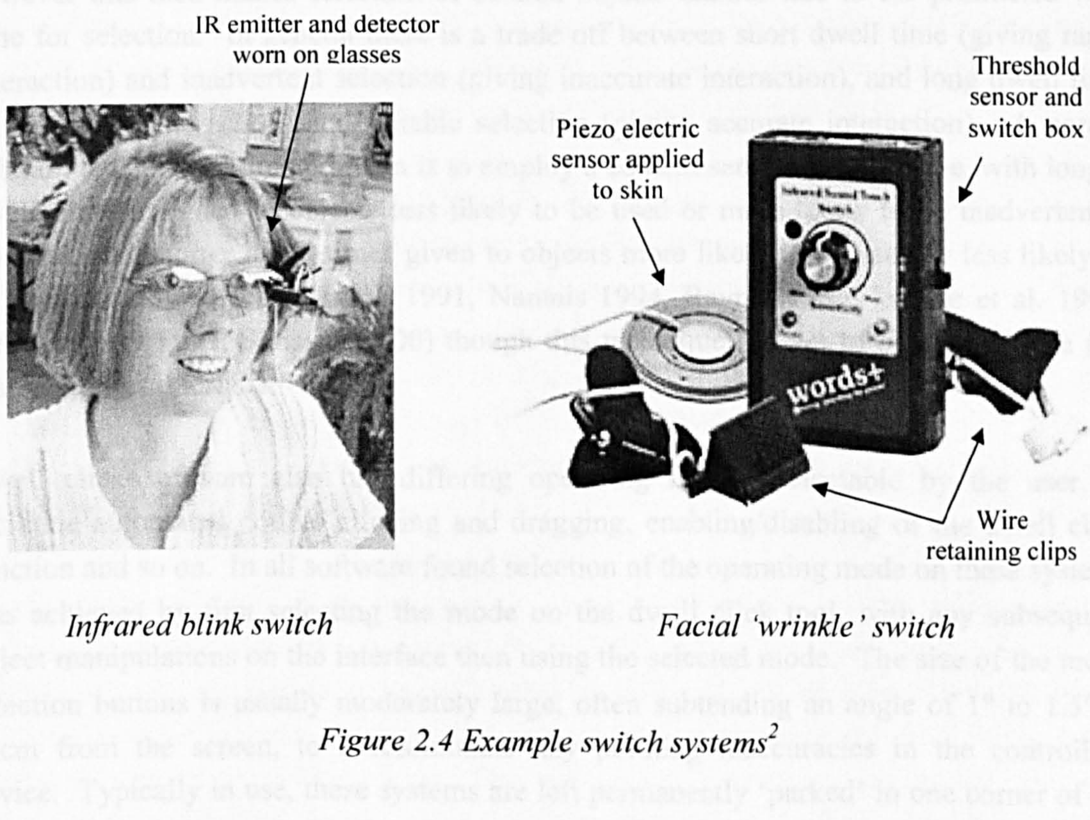
There appears, to date, to be a moderate body of work on the merits of eye mouse systems to support the search findings. The most comprehensive reviews on eye tracking technologies (Duchowski 2000, Young and Sheena 1975, Glenstrup and Engell-Nielsen 1995) support the popularity of VOG, and to a lesser extent EOG, systems when used for eye mouse systems. Here non-invasive VOG, and particularly EOG, eye gaze tracking systems are characterised as inaccurate in comparison to invasive systems but are regarded as more suitable for eye mouse applications than other technologies.

Typically, eye mouse systems are prohibitively expensive, with systems ranging from £10000 to £20000, although work is currently being undertaken to develop low cost systems based on pupil tracking using commodity cameras (Hansen et al. 2001). This is a simpler variant of video oculography, the most popular system, and uses an inexpensive camera such as a web or USB camera in conjunction with advanced video processing software to track the gaze direction of the pupil alone. These systems would be inexpensive, but to date are inaccurate, typically with an accuracy of 4° visual angle, though it is expected that the accuracy of these systems will improve perhaps to the point where they can be used for direct interaction on a standard graphical user interface. These simple systems would then rival the current cost of head mouse systems.

2.5 Object selection systems

In addition to controlling cursor movement, head and eye mouse systems both require methods of selecting objects on the interface once the cursor has been located on objects of interest. Here there are two basic groups of object selection operation; the first is *multimodal* operation where the user has sufficient physical ability to use a second modality to control a switch, such as a hand-held button or eyewink sensor, or by voice for example[1]. The second is *monomodal* operation where the user has no other available modality to support selection of objects and selection is typically achieved by using software to generate selection actions.

For multimodal operation selection of objects is simply achieved by moving the cursor onto the object on the interface by head or eye movement and then operating the switch or voice modality. When a secondary supporting modality is related closely to the head or eye, such as wink or blink (Rasmusson et al. 1999), facial wrinkle (Partala et al. 2001) or head or shoulder movement (Bates 1999, Beukelman and Mirenda 1992), then care must be taken not to influence the head or eye or position during selection. In these cases filtering of the cursor position data to stabilise the selection position can be applied (Jacob 1991, Stampe 1993, Stampe and Reingold 1995). Typical switch systems are shown (Figure 2.4).



Infrared blink switch *Facial 'wrinkle' switch*

Figure 2.4 Example switch systems[2]

[1] Voice as a selection modality is not within the scope of this work, and was rejected partly due to speech being interrupted during assisted respiration for high-level motor injuries (Bates 2002a), and also due to the lack of popularity of this modality in use, from conversation at the ACE Centre (Aiding Communication in Education), www.ace-centre.org.uk and the ACT (Access to Communication and Technology), Regional Rehabilitation Centre, Oak Tree Lane Centre, Birmingham).

[2] Both from Words+ www.words-plus.com

For *monomodal* operation selection is typically achieved by using 'dwell click' (Jacob 1991). In this case a software device or tool is used to continuously monitor the position of the cursor, with a selection generated at the cursor position when the cursor has remained static, or 'dwelling' on an object, for greater than a specified time, typically between 600ms to 1500ms, with a balance between an excessive dwell time causing fixation difficulties and long task times, and a short dwell time causing inadvertent selections (Istance et. al. 1996, Jacob 1990, Jacob 1991, Sibert and Jacob 2000, Stampe and Reingold 1995, Ware and Mikaelian 1987).

Dwell click has one major disadvantage in that inadvertent clicks may be generated simply by the user resting the cursor on an object for greater than the dwell click time. This is known as the 'Midas Touch' problem (or perhaps Midas Gaze) where objects on the interface are continually inadvertently selected (Velichkovsky 1997, Jacob 1991). One partial solution to this problem is to greatly extend the dwell time to several seconds; however this then makes selection of desired objects tedious due to the protracted wait time for selection. In general there is a trade off between short dwell time (giving rapid interaction) and inadvertent selection (giving inaccurate interaction), and long dwell time (giving slow interaction) and reliable selection (giving accurate interaction). A partial solution to the dwell time problem is to employ a context sensitive dwell time, with longer dwell times assigned to objects less likely to be used or more likely to be inadvertently selected, and shorter dwell times given to objects more likely to be used or less likely to be inadvertently selected (Jacob 1991, Nantais 1994, Rayner 1995, Istance et al. 1996, Velichkovsky 1997, Salvucci 2000) though this technique has yet to be applied to a full graphical user interface.

Dwell click software also has differing operating modes, selectable by the user, to facilitate automated double clicking and dragging, enabling/disabling of the dwell click function and so on. In all software found selection of the operating mode on these systems was achieved by first selecting the mode on the dwell click tool, with any subsequent object manipulations on the interface then using the selected mode. The size of the mode selection buttons is usually moderately large, often subtending an angle of 1° to 1.5° at 60cm from the screen, to accommodate any pointing inaccuracies in the controlling device. Typically in use, these systems are left permanently 'parked' in one corner of the screen, and are automatically placed on top of all other applications so that they are visible and available at all times. This approach does not usually produce visual conflicts with other on-screen applications since these systems have few functions, and hence require few buttons or selection options, and so occupy only a small fraction of the screen. Example dwell click software tools are shown (Figure 2.5).

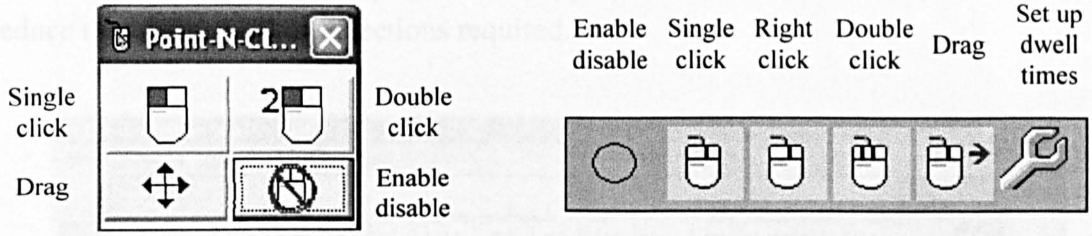


Figure 2.5 Example dwell click tools[1]

2.6 Text entry systems

In addition to object selection tools, head and eye mouse systems require methods of generating textual input to the interface. Typically this is achieved by using a virtual on-screen keyboard placed on the interface (Istance et al. 1996a, Istance et al. 1996b, Istance 1997, Shein et al. 1992, Shein et al. 1991, Leventhal 1991, Heuvelmans et al.1990, Frey et al. 1990, Stampe and Reingold 1995), although off-screen targets placed around the edge of the screen have been proposed to conserve screen area (Isokoski 2000).

Since virtual keyboards are not limited by the physical constraints of actual physical desktop keyboards, there is no need to limit a virtual keyboard to a simple 'qwerty' layout, although many virtual keyboards do mimic physical keyboards as a default layout (Figure 2.6). Instead, custom and even dynamically changing layouts are often used. These range from alphabetical layouts, which may be more rapid to learn, to 'Dvorak' layouts and to layouts based on the frequency of usage of letters (Leventhal 1991, MacKenzie 1999). In addition the layouts need not be rectangular, with linear, square and triangular layouts proposed (Leventhal 1991, MacKenzie 1999).

Virtual keyboards also allow dynamically changing keyboard content, where word or sentence prediction options are continually updated on the keyboard, thus reducing the number of keys required to compose a word or sentence. Word prediction also allows dynamic changes in dwell selection time with shorter dwell times for more likely letters and longer times for less likely letter. This approach can greatly reduce dwell times (20 – 65%) and hence increase typing speed (Nantais 1994, Salvucci 1999). Phrase expansion using bi and tri-grams (Mathy-Laikko et al. 1993) and sentence or paragraph expansion using macro-like expansions and 'sentence compansion' (where user types 'john apple

[1] Left: 'Point-n-Click' from Polital Enterprises www.polital.com, right: 'Dwell Clicker' from Sensory Software www.sensorysoftware.com

ate' and the computer types 'john ate the apple') (Demasco et al. 1992) are also used to reduce the number of key selections required.

Figure 2.6 On-screen keyboard default 'qwerty' layout[1]

Inadvertent selection of keyboard keys due to 'Midas Touch' (Velichkovsky 1997, Jacob 1991) has been addressed to some extent on virtual keyboards. Context sensitive dwell times, in a similar fashion to context sensitive dwell selection, can be used with longer dwell times assigned to less likely key selections and shorter dwell times assigned to more likely key selections, based on word prediction or frequency of key usage. Giving audible sound or speech feedback (as suggested by Brewster et al. 1996, implemented by Majaranta et al. 2004) or visible feedback (Istance et al. 1996b, Majaranta et al. 2004, Lankford 2000) of an impending key press has been used and reduced errors by giving the user time to move the cursor away from unintended key. In addition, implementing visual feedback with a shrinking symbol or character on each key to indicate elapsed dwell time was found to help centre, and keep, visual attention on desired targets (Majaranta et al. 2004).

Exploiting the possibilities of a non-static keyboard, a virtual keyboard can be self adapting to the state of the current application present on the interface and hence present only the required keys suitable for interaction at that time, thus completely different keyboard overlays may be dynamically interchanged on the same keyboard, such as alphabetical, numeric or command keypads. This has the effect of maximising the size of the keys in the available space of the keyboard whilst reducing the possibility of error by removing unnecessary keys (Istance et al. 1996b) (Figure 2.7).

[1] 'WiViK' on-screen full-function keyboard from Prentke Romich, www.prentrom.com

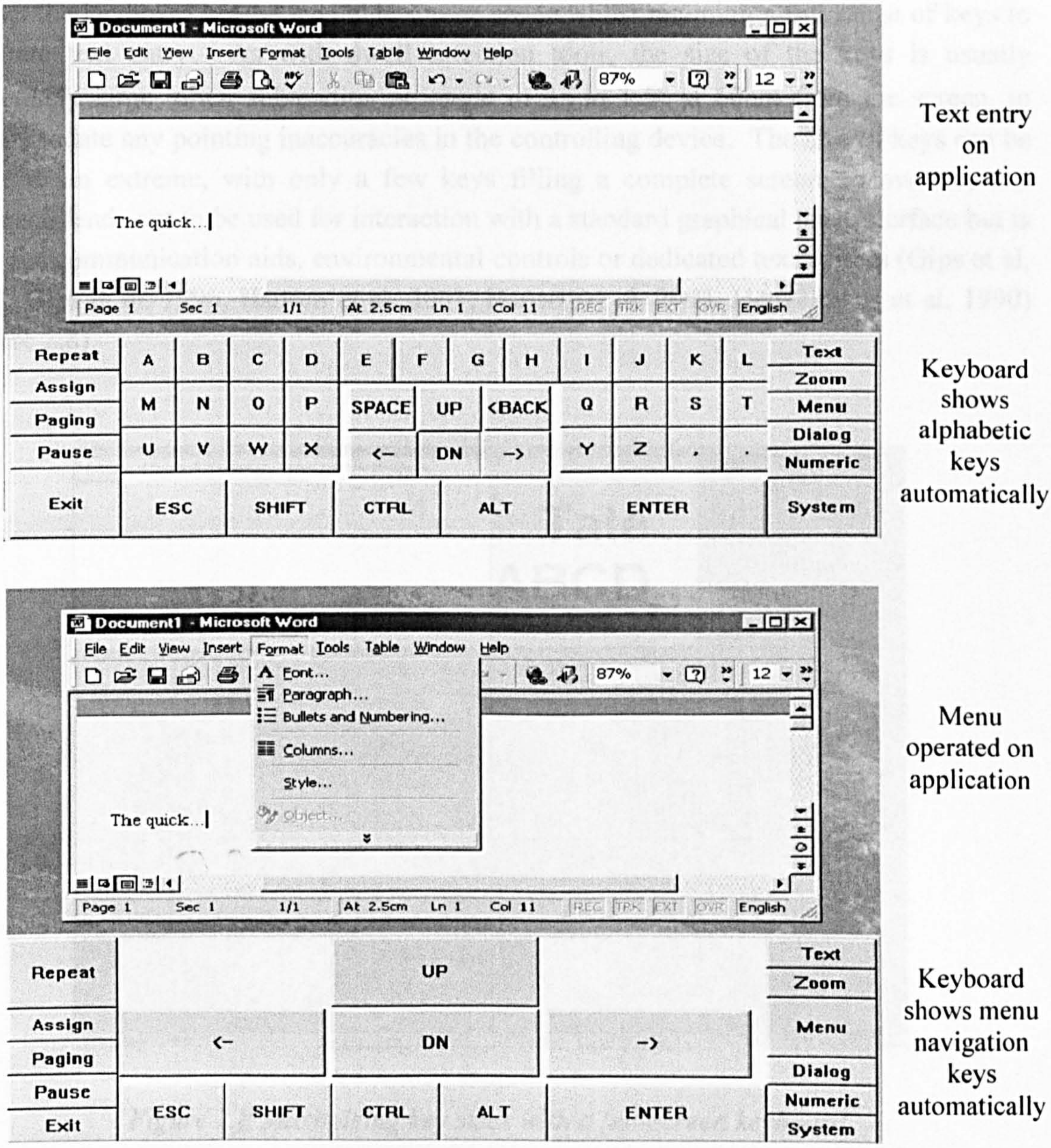


Figure 2.7 Dynamic keyboard self adapting to an application[1]

There is a trade-off between the number of keys on a keyboard, the size of these keys to aid ease of selection, and the amount of screen area occupied by the keyboard. A dynamic approach to keyboard content allows a large range of key commands to be available within the bounds of a single keyboard whilst maximising the size of the keys, thus reducing the

[1] 'EC Key' self-adapting keyboard, Istance et al. 1996b.

impact the keyboard has on available screen space whilst retaining a full range of keys to facilitate text entry. As with dwell selection tools, the size of the keys is usually moderately large, often subtending an angle of 1° to 1.5° at 60cm from the screen, to accommodate any pointing inaccuracies in the controlling device. The size of keys can be taken to an extreme, with only a few keys filling a complete screen. However, this approach tends not to be used for interaction with a standard graphical user interface but is used for communication aids, environmental controls or dedicated text editors (Gips et al. 1993, Gips et al. 1996, Hansen et al. 2001, Hansen et al. 2003, Heuvelmans et al. 1990) (Figure 2.8).

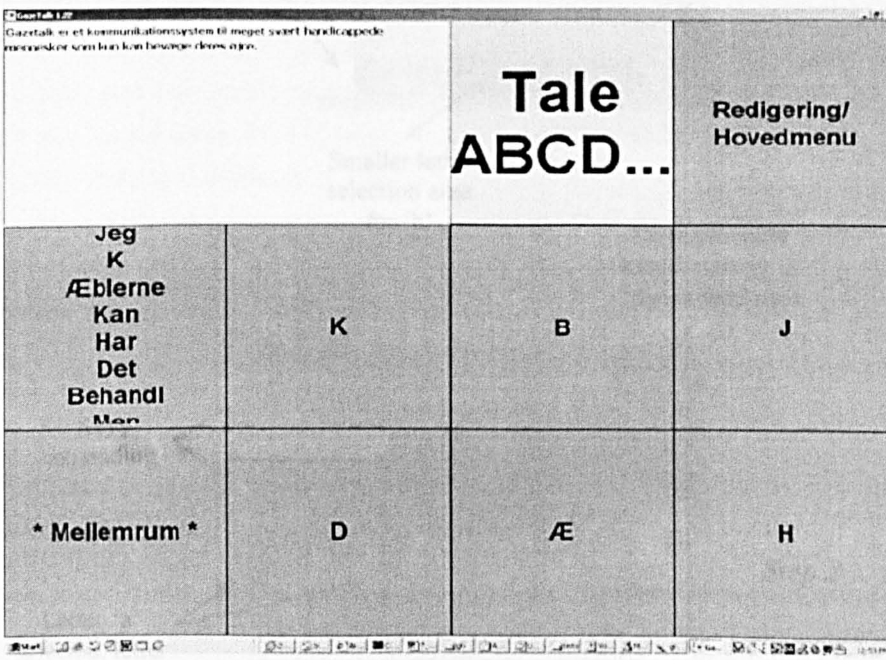


Figure 2.8 Maximising key sizes with a full-screen keyboard[1]

The dynamic qualities of virtual keyboards have been taken to an extreme with the 'Dasher' keyboard that has few static keys, with the keys effectively 'flying' toward the user based on word prediction (Figure 2.9). Here the user points close to a group of letters (on the right of the keyboard) thus invoking the keyboard to 'fly' these letters toward the user, hence making each letter larger and easier to select. Once the desired letter is large enough to be selected, the keyboard then starts flying the most likely subsequent letters toward the user, thus again enabling ease of selection.

[1] 'Gaze talk' full-screen large key keyboard with predictive text, Hansen et al. 2003

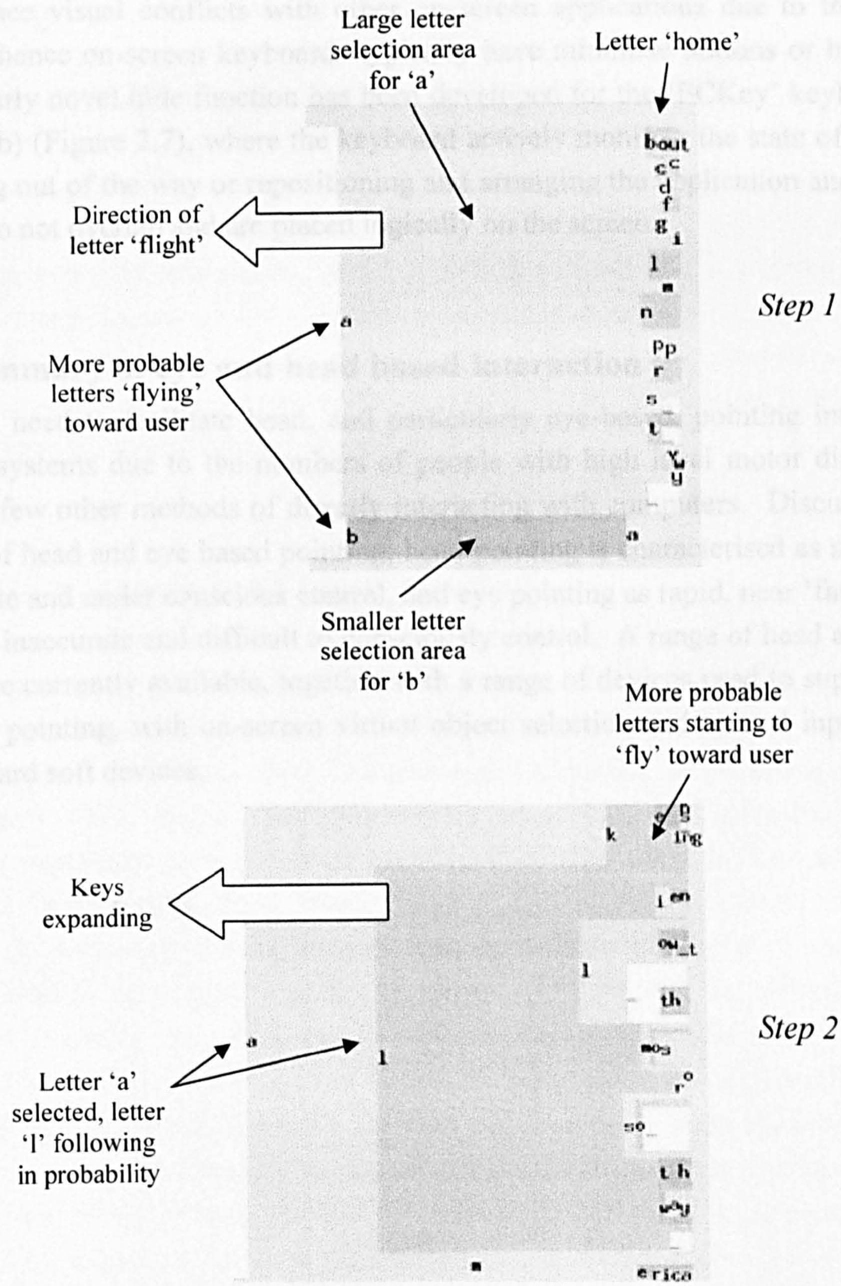


Figure 2.9 Flying letters on a dynamic keyboard[1]

Typically in use, virtual keyboards are left 'parked' filling the lower one third to one half of the screen, and are automatically placed on top of all other applications so that they are visible and available at all times. Unlike the smaller object selection tools, this approach

[1] 'Dasher' dynamic keyboard from University of Cambridge, www.inference.phy.cam.ac.uk/dasher

does produce visual conflicts with other on-screen applications due to the size of the keyboard, hence on-screen keyboards typically have minimise buttons or hide functions. A particularly novel hide function has been developed for the 'ECKey' keyboard (Istance et al. 1996b) (Figure 2.7), where the keyboard actively monitors the state of the interface, auto-hiding out of the way or repositioning and arranging the application and keyboard so that they do not overlap and are placed logically on the screen.

2.7 A summary of eye and head based interaction

There is a need to facilitate head, and particularly eye-based, pointing interaction with computer systems due to the numbers of people with high level motor disabilities who may have few other methods of directly interacting with computers. Discussing the pros and cons of head and eye based pointing; head pointing is characterised as slow, effortful, but accurate and under conscious control, and eye pointing as rapid, near 'fatigue-free' but inherently inaccurate and difficult to consciously control. A range of head and eye mouse systems are currently available, together with a range of devices used to support head and eye based pointing, with on-screen virtual object selection and textual input dwell click and keyboard soft devices.

Chapter 3

Assessment Methods

This chapter examines the diverse methods available for assessing the performance of the head and eye mouse systems introduced in Chapter 2. It gives a review of the performance measurement studies which focussed on head and eye based pointing, and shows that these studies are not suitable, or do not give enough detail, for assessing the 'real world' performance of head and eye mice. Finally this Chapter will justify the development of a new assessment method suitable for the performance evaluation of the head and eye mice systems from Chapter 2.

3.1 Assessing eye and head based pointing

Previous work on methods of assessing the performance of eye and head based pointing devices on graphical user interfaces fell into two areas; abstract target acquisition tests (for example, MacKenzie 1992, MacKenzie 1991, MacKenzie and Buxton 1992, Accot and Zhai 1997, Sibert and Jacob 2000, Douglas and Kirkpatrick 1999, Murata 1991, Istance and Howarth 1993, Bates 1999, Radwin et al. 1990), and simulated 'real world' interaction sequences on a graphical user interface (for example, Istance et al. 1996a, Jacob 1993, Hansen et al. 2004, Majaranta et al. 2004). Abstract target acquisition tests are based on presenting the test subject with a sequence of targets of varying size and spatial separation on an otherwise blank screen. The object of the assessment is for the test subject to simply use a head or eye mouse to select the targets as they appear on the interface (Figure 3.1). Selection can either be monomodal, using a dwell click tool, or multimodal with a switch. Typically the data collected from these experiments is sparse, with the time taken to select targets and the number of errors being recorded. In contrast, simulated 'real world' tests are typically based on the user performing a small set of tasks or interaction sequences on either a real graphical user interface, or a simulated and simplified version of a real interface. Often these tests only assess one type of interaction, such as typing on an on-screen keyboard, that occur on a real interface (Figure 3.2). The data from these experiments is usually determined by the nature of the assessment task, for example words per minute for a typing task, but other metrics such as cursor paths, eye scan paths or user subjective reaction are often recorded, giving a richer data set.

Both abstract target tests and 'real world' simulation tests have advantages and disadvantages. Typically, abstract test scenarios tend to be easier and more rapid to administer and evaluate (due to the high level of automation possible in data collection) but possibly offering less detail due to their simplified abstract nature. In contrast 'real world' test scenarios are often slower and more difficult to administer (due to the low level of automation in data collection) but possibly give more information as they are more complex but closer to genuine interaction with a device.

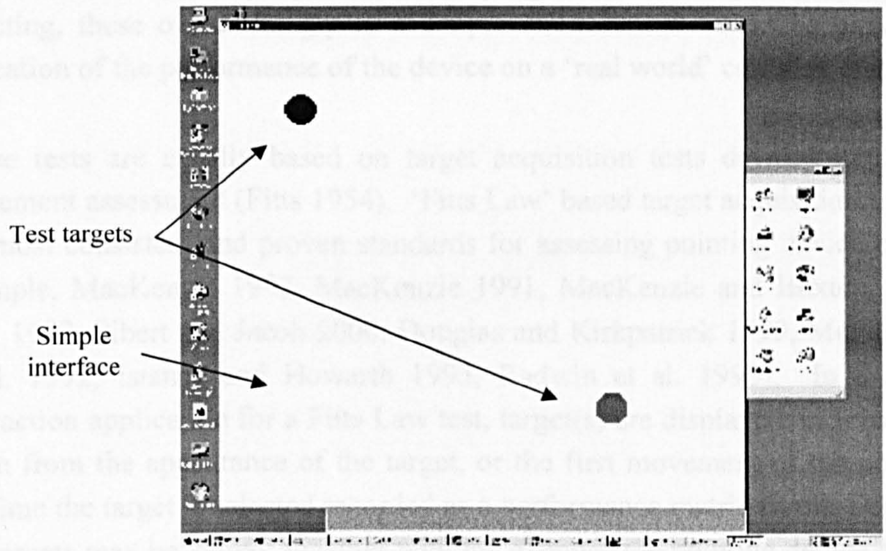


Figure 3.1 Example abstract target acquisition test[1]

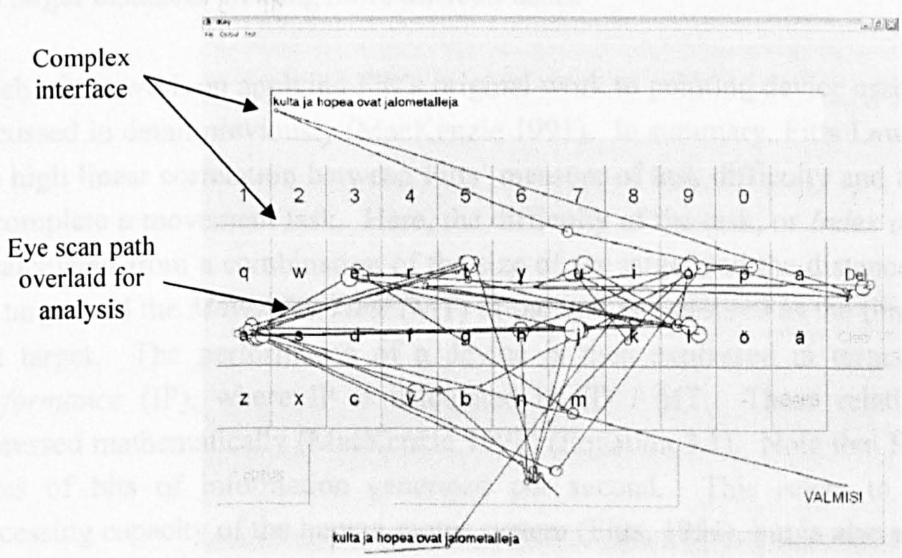


Figure 3.2 Example 'Real World' typing test[2]

[1] Target acquisition test from Zhai 1997
[2] Text editor and keyboard from Majaranta et al. 2004

3.2 Abstract tests on head and eye based pointing

Abstract target acquisition tests assess the performance of head and eye mouse systems when selecting a sequence of abstract targets. The rationale behind these simple tests is that, although often complex, a graphical user interface is simply a collection of discrete objects. Hence determining the performance of a pointing device when moving to, and selecting, these objects singly in a simple environment would be expected to give an indication of the performance of the device on a 'real world' complex environment.

These tests are usually based on target acquisition tests devised for hand and stylus movement assessment (Fitts 1954). 'Fitts Law' based target acquisition tests have become the most consistent and proven standards for assessing pointing device performance (For example, MacKenzie 1992, MacKenzie 1991, MacKenzie and Buxton 1992, Accot and Zhai 1997, Sibert and Jacob 2000, Douglas and Kirkpatrick 1999, Murata 1991, Howarth et al. 1992, Istance and Howarth 1993, Radwin et al. 1990). In a human-computer interaction application for a Fitts Law test, target(s) are displayed on screen, with the time taken from the appearance of the target, or the first movement of the pointing device, to the time the target is selected recorded as a performance metric for the device. The size of the targets may be varied; together with the distance the pointing device cursor is required to travel from its starting point to select the target. The permutation of target size and distance to the target allows the difficulty of test tasks to be varied, with smaller targets and larger distances creating more difficult tasks.

Much of the work on applying Fitt's original work to pointing device assessment has been discussed in detail previously (MacKenzie 1991). In summary, Fitts Law states that there is a high linear correlation between Fitts' measure of task difficulty and the time required to complete a movement task. Here, the difficulty of the task, or *Index of Difficulty* (ID), is calculated from a combination of the size of the target and the distance to be moved to the target and the *Movement Time* (MT) of the task is measured as the time taken to obtain that target. The performance of a device is then expressed in terms of an *Index of Performance* (IP), where IP is calculated as ID / MT. These relationships may be expressed mathematically (MacKenzie 1991) (Equation 3.1). Note that IP is expressed in terms of bits of information generated per second. This refers to the information processing capacity of the human motor system (Fitts, 1954), but is also appropriate to the human motor system with the addition of pointing devices. Simply, the higher the bit rate of the device, the higher its performance will be on a target acquisition task. Since IP should be constant, and the relationship between MT and ID is logarithmically proportional, then MT may be easily plotted against ID for any given device (Figure 3.3).

$$ID = \log_2(A / W + 1)$$

Index of Difficulty (ID) of the task (dimensionless)

$$MT = a + b\ ID$$

Movement Time (MT) for the task (seconds)

$$IP = ID / MT$$

Index of Performance (IP) in Bits (of information generated) / second

Where:
A = Amplitude of movement, or distance to target
W = Width of target, or size of target
a, b = intersection and slope constants, determined by linear regression

Equation 3.1 Fitts Law relationships[1]

The utility of testing a device with a Fitts law type abstract test is that once the IP of a device has been determined, Fitts Law states that the time taken (MT) for any target acquisition task on an interface may be predicted, and hence the time for a sequence of interactions could be determined without further testing. Further, and perhaps the most useful property of using Fitts Law, is that it gives a known and established metric (IP) that can be used to compare the performance of many differing pointing devices (For example MacKenzie 1992).

There are, however, problems associated with such simple metrics. Firstly, Fitts law does not directly deal with error conditions such as target misses during attempted selections, which are then followed by a correct selection. One way to address this problem is to use only error free trials in the calculation, and report errors separately (Sibert et al. 2001, Jagacinski and Monk 1985, Epps 1986). Another is to include all trials, including trials with errors, by allowing test subjects to continually make target misses until a correct selection is accomplished, and also report the error rate (Miniotas 2000, Istance and Howarth 1993, Ware and Mikaelian 1987, Card et al. 1978, Fitts 1954). However, neither

[1] Hand tapping data from Fitts 1954, plotted by MacKenzie 1991.

of these approaches is appropriate to predicting performance on a 'real world' interface, where the consequences of incorrect selections may be high.

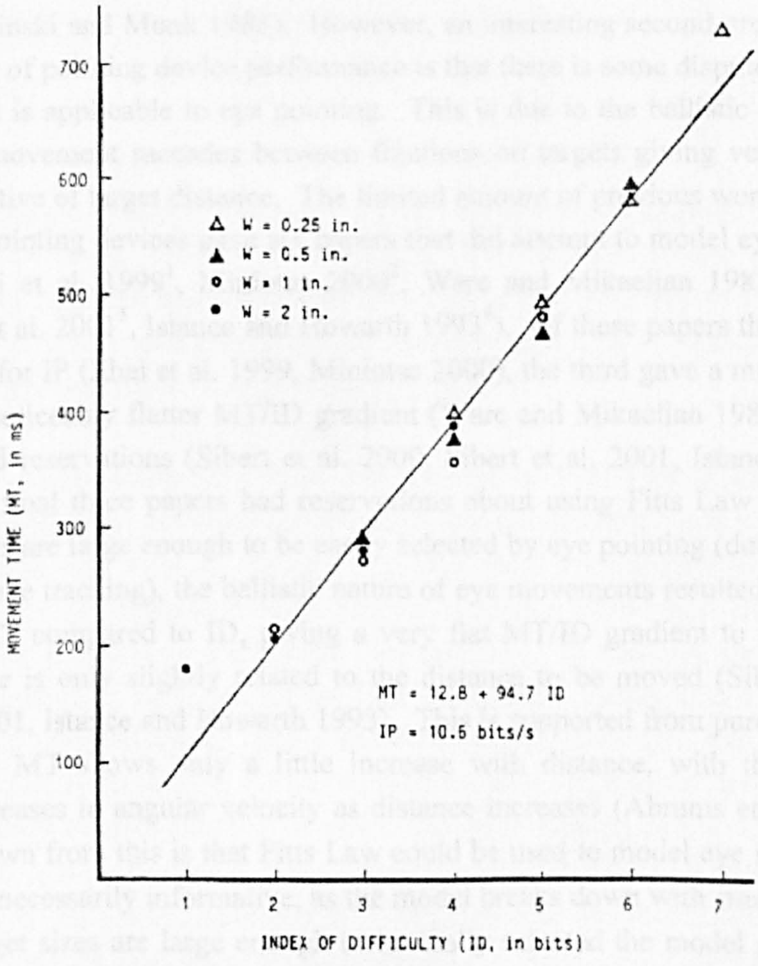


Figure 3.3 Fitts relationship between MT and ID[1]

The error-handling problem has been addressed by recording the exact position of target selections on, and around, the target. These selection positions form a distribution around the target, with the width of the distribution used to give an 'effective' target width for the target (MacKenzie 1991). For example, if a device produces erroneous selections outside the target boundary, then the target width used in the Fitts calculation of ID will be extended to encompass the width of these selections, thus the IP of the device is reduced

[1] Hand tapping data from Fitts 1954, plotted by MacKenzie 1991.

by generating errors. This approach is valid, but does complicate the calculation of IP as the distributions of every selection on every target must be calculated.

Fitts law has been shown to apply to head based pointing (LoPresti et al. 2000, Radwin et al. 1990, Jagacinski and Monk 1985). However, an interesting second problem with Fitts Law modelling of pointing device performance is that there is some dispute whether or not the Fitts model is applicable to eye pointing. This is due to the ballistic nature and high speed of eye movement saccades between fixations on targets giving very constant MT results irrespective of target distance. The limited amount of previous work that measured the IP of eye pointing devices gave six papers that did attempt to model eye pointing with Fitts law (Zhai et al. 1999[1], Miniotas 2000[2], Ware and Mikaelian 1987[3], Sibert et al. 2000[4], Sibert et al. 2001[5], Istance and Howarth 1993[6]). Of these papers the first two gave similar results for IP (Zhai et al. 1999, Miniotas 2000), the third gave a much lower result for IP with a noticeably flatter MT/ID gradient (Ware and Mikaelian 1987) and the final three expressed reservations (Sibert et al. 2000, Sibert et al. 2001, Istance and Howarth 1993). These final three papers had reservations about using Fitts Law and found that, provided targets are large enough to be easily selected by eye pointing (due to the inherent inaccuracy of eye tracking), the ballistic nature of eye movements resulted in a very small variation of MT compared to ID, giving a very flat MT/ID gradient to the point where movement time is only slightly related to the distance to be moved (Sibert et al. 2000, Sibert et al. 2001, Istance and Howarth 1993). This is supported from pure eye movement studies, where MT shows only a little increase with distance, with the eye showing noticeable increases in angular velocity as distance increases (Abrams et al. 1989). The conclusion drawn from this is that Fitts Law could be used to model eye pointing, but the results are not necessarily informative, as the model breaks down with smaller target sizes, and where target sizes are large enough to be easily selected the model gives a near flat MT irrespective of distance.

A final problem with using simple abstract target acquisition tests is that these tests cannot claim to give the *actual* performance of a device on a complex interface. Other factors not

---

[1] b= 220.0 (ms / bit)*
[2] b=176.0 (ms / bit)*
[3] b=73.0 (ms / bit)*
[4] b= 1.7 (ms / bit)*
[5] No actual figure given, statement made that eye mouse was not modelled by Fitts Law.
[6] b= 130 (ms / bit) however very poor fitting model R^2=.08 so authors declared eye mouse was not modelled by Fitts Law.

*Compare to typical standard desktop hand mouse results between b=392 (ms / bit) (Epps 1986), b= 126.0 (ms / bit) (Miniotas 2000), b= 117.0 (ms / bit) (Sibert et al 2001), b= 120.0 (ms / bit) (Istance and Howarth 1993).

present in simple abstract tests such as visual distractions and feedback from the graphical user interface (Velichkovsky et al. 1997, Jacob 1993), and the consequences of any errors generated during interaction with the interface will influence pointing device performance. Hence abstract target acquisition tests should be regarded as giving the *raw* pointing performance of a device and not the performance of the device in a 'real world' scenario.

3.3 'Real world' tests on head and eye based pointing

'Real world' tests assess the performance of head and eye mouse systems when interacting with simulated or actual 'real world' scenarios. The rationale behind these potentially complex tests is that, although often time consuming and laborious to conduct, the true performance of a device on a graphical user interface cannot be known unless that device is tested on such a 'real world' complex environment.

Unlike abstract target acquisition tests and the use of Fitts Law (Chapter 3.2), there appears to be no standard or commonly accepted test for assessing 'real world' interaction on an interface for any pointing device. Typically tasks are designed to test or assess a particular element of interaction with specific interest, rather than the full range of interaction that is possible on a graphical user interface. In addition, the factors that are assessed and quantified vary due to the task undertaken, rather than using a common method, making comparison of results between studies difficult. Examining previous work conducted on head and eye based pointing found a range of different test scenarios: A brief, with only a small number of tasks, but wide ranging assessment of eye-based interaction with text entry, text editing, application and menu manipulation and limited internet browsing was found (Istance et al. 1996), however this interaction was carried out indirectly with the interface, via a virtual keyboard ('ECKey', illustrated in Figure 2.6). In this work, performance metrics were the text entry rate in number of characters per minute, together with task times and task error rates. Another attempt at a range of assessment scenarios for eye based pointing involved typing on a full-screen keyboard, typing on an environmental control with full screen keys, and playing a simple game; with metrics of simple success or failure of the tasks (Chapman 1991). Here again, there was no direct interaction with the underlying standard Windows user interface, with interaction taking place only with specifically designed applications. Direct interaction has been assessed with a graphical user interface; however this eye-based interaction was with a non-standard graphical user interface, and involved selection and manipulation of ship representations on a military interface (Jacob 1993, Sibert 2000) (Figure 3.4). Here, performance metrics were task time and task error rate.

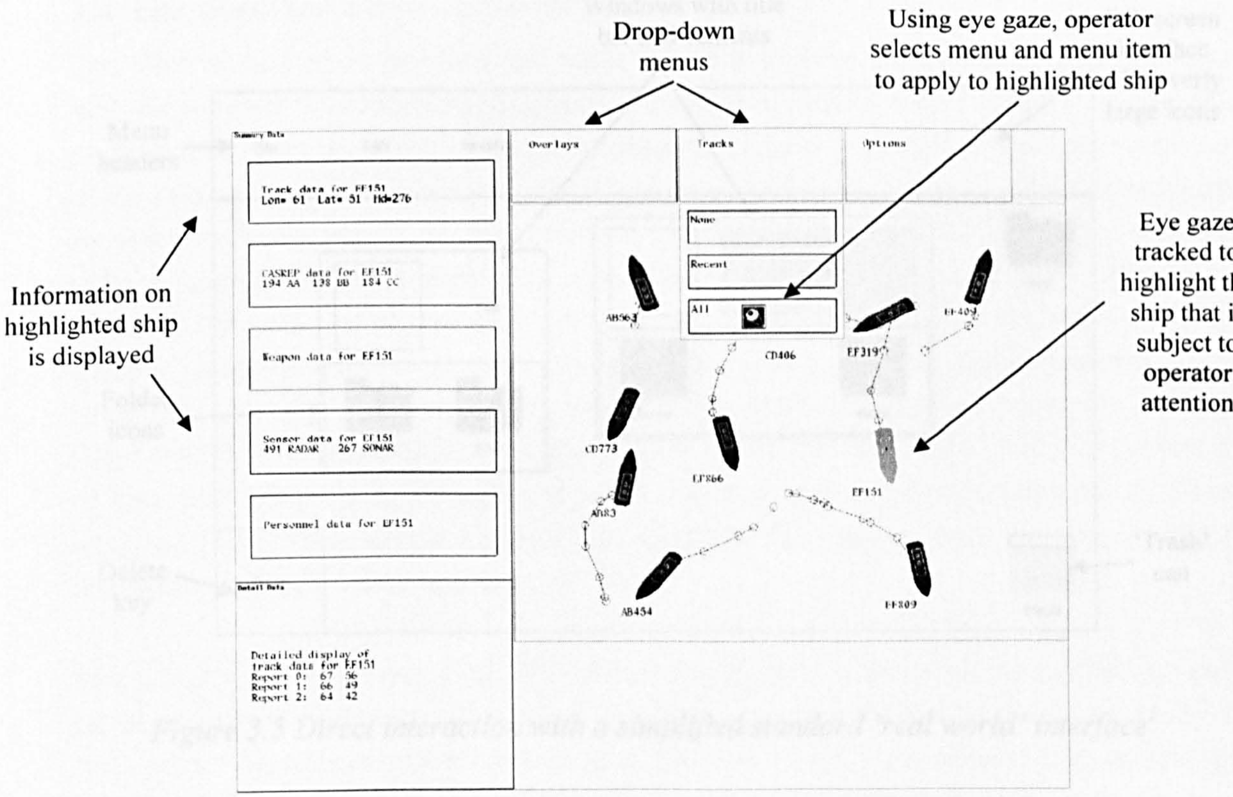


Figure 3.4 Direct interaction with a non-standard 'real-world' interface[1]

Other assessments of direct interaction with non-standard interfaces were gaze-based identification of objects of interest on a graphics displays (Starker and Bolt 1990, Goldberg and Schryver 1993) where gaze scan paths were recorded to determine the object magnification intent of the user and so zoom in the interface on objects of interest. In the latter of these, the success of determining zoom intent was measured.

Assessments were made on reduced or simulated 'real world' scenarios of a simplified interface, with a simulated drop-down menu created (Byrne et al. 1999) and an enlarged and simplified icon and menu based interface used to test an enhanced dwell click tool to highlight objects of interest before selection (Salvucci and Anderson 2000) (Figure 3.5). Here metrics were task time and error rate. However, in these examples, the interface was a highly abstracted and simplified version of a standard graphical user interface rather than direct interaction on a standard graphical interface.

[1] Ship tracking and information non-command interface, Jacob 1993.

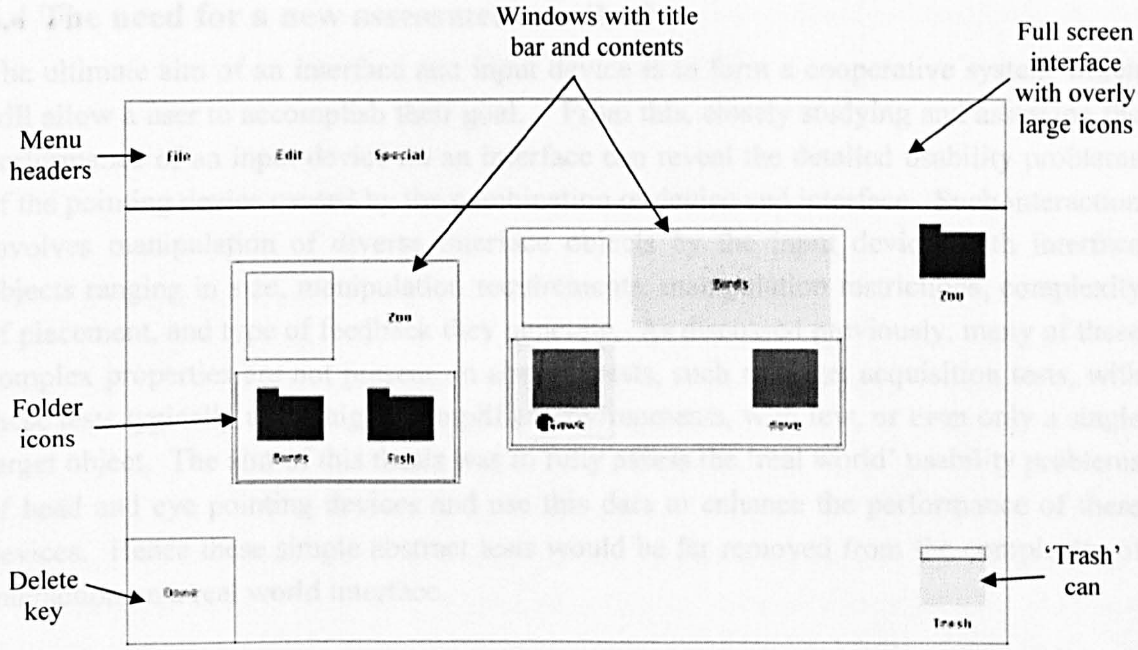


Figure 3.5 Direct interaction with a simplified standard 'real world' interface[1]

Remaining 'real world' assessments were based around text entry, typically with full-screen sized keyboards without direct interaction and manipulation of the underlying interface. Metrics for these studies were typing rate and subjective like or dislike of the overall system (Stampe and Reingold 1995), typing rate, error count, task time, gaze scan paths of the eye on the interface and subjective like or dislike of the system (Majaranta et al. 2004), and typing rate and user subjective qualification of typing efficiency and satisfaction with the system (Hansen et al. 2004).

From this, it was evident that, unlike abstract target acquisition tests, there was no previous commonly used assessment of head or eye based pointing *direct* interaction on a standard graphical user interface. It was also notable that the subjective user reaction to using a device was commonly assessed in 'real world' based assessments. Typing rate was common to papers assessing keyboards, and this could be used to compare differing head and eye based pointing devices if the same keyboards and text entry tasks were used, or if the same device was used and differing keyboards assessed for their efficacy. However, this is limited to assessing a single task type, not interaction with a full interface.

[1] Simplified eye-gaze based 'real world' interface, Salvucci and Anderson 2000.

3.4 The need for a new assessment method

The ultimate aim of an interface and input device is to form a cooperative system which will allow a user to accomplish their goal. From this, closely studying and assessing the performance of an input device on an interface can reveal the detailed usability problems of the pointing device caused by the combination of device and interface. Such interaction involves manipulation of diverse interface objects by the input device, with interface objects ranging in size, manipulation requirements, manipulation restrictions, complexity of placement, and type of feedback they generate. As discussed previously, many of these complex properties are not present on abstract tests, such as target acquisition tests, with these tests typically using highly simplified environments, with few, or even only a single target object. The aim of this thesis was to fully assess the 'real world' usability problems of head and eye pointing devices and use this data to enhance the performance of these devices. Hence these simple abstract tests would be far removed from the complexity of interaction on a real world interface.

Since this work was concerned with attempting to gain a thorough understanding and insight on the performance of head and eye based pointing devices in the 'real world', a 'real world' based assessment scenario is an obvious choice. It is necessary to accurately mimic typical 'real world' direct interaction sessions with these devices on a standard graphical interface and then measure, in a detailed, validated and repeatable method, their performance on such interaction sessions. It would also be desirable to retain the strengths of abstract target acquisition tests, such as ease of application, and perhaps more importantly their known and established metrics that allow comparison of the performance of many differing pointing devices. To achieve this, the aim of an assessment method suitable for this work would be to take the factors that describe repeatable and standardised 'real world' interaction, to assess the interaction partnership of head and eye based pointing devices in this interaction in a standardised way, and to enable the assessment method to be applicable to other devices to allow comparison. Finally, as discussed in Chapter 1, although eye-based pointing is the main focus of this work, a head mouse is included in this work to give a similar assistive technology pointing device comparison, with a standard desktop hand mouse also used to give a known and probably high performance baseline for the work. An assessment method must encompass the performance bounds of all of these devices, with performance metrics that are both sensitive enough to measure small changes in the performance of a device, and also broad enough to measure the potentially high performance differences between devices. In addition, the method must be able to accommodate both monomodal interaction, aided by a dwell click tool as discussed in Chapter 2, and multimodal interaction with switch selection, together with textual entry via on-screen virtual keyboards. As discussed previously, to date no such suitable 'real world' assessment method has been found in use; hence it was decided to develop a suitable assessment method.

3.5 The requirements for the assessment method

As discussed previously, one of the principal downfalls of 'real world' testing was the lack of any standardised sequence of tests tasks that could be performed by a range of devices, and that would allow comparison of results between those devices. To overcome this limitation, the proposed assessment method would need to clearly contain a set of 'standard' test tasks that would be applied to any device performing the method. It would be conceivable to allow random interaction with an interface with the hope that such random interaction would eventually form an 'average' interaction, however such sessions would need to be lengthy and this concept introduces bias from the user as they have free-will during interaction. Therefore, any set of test tasks that would be repeatable and allow task for task comparison between devices would need to be clearly prescribed. In addition, these test tasks should be representative as closely as possible with 'real world' interaction. Hence the assessment method should contain clearly described 'real world' test tasks. The previous discussion on assessment methods found that task times and error rates were important, together with task success or failure, with these *objective* metrics often supported by *subjective* metrics of user reaction to the device and test. Hence objective and subjective user data should be included in an assessment method.

The objective success of any pointing device on performing an assessment method must be measured to determine the objective *performance* of that device. Typically objective performance has been measured by the time taken to complete a task and a basic task quality metric of the number of errors generated during the completion of the task. Although adequate, with a device that has a shorter task completion time and a lower error rate (higher quality of interaction) during the task almost certainly being more suitable for the task than a device with a longer task time and higher error rate, these metrics are quite crude and do not offer great insight into the detailed performance of a device. Perhaps a device has a shorter task time but higher error rate than another device with a longer task time but lower error rate – which device is most suitable for the task? To resolve this problem, task times are typically used as the main comparator between pointing devices, with the error rates being reported separately (For example: MacKenzie 1992, Douglas and Kirkpatrick 1999) and the reader left to decide which metric is most important for their application of the results. A measurement scheme was required that would overcome this difficulty by taking into account both task times and error rates, or the quality of interaction, together with task success or failure, to form a *composite* objective metric of device *performance* on the test tasks. Hence the aim of the assessment method would be to measure a range of objective metrics to allow examination of differing factors, together with presenting a single composite objective result of performance.

To better gain a full understanding of the performance of a device it is regarded as not adequate to simply measure the objective performance of a device without also assessing the subjective *reaction* of the user when using the device (Bevan et al. 1991 and 1995). Perhaps a device performed well objectively, with low task times and error rates, but the user *worked* hard to control the device, or the device was *uncomfortable* to use. Would this device be more suitable to the task than a device that objectively performed less well but required less work from the user, or was more comfortable to use? This problem has been partially addressed previously, with a multitude of differing questionnaires (For example: Douglas and Kirkpatrick 1999, Smith 1996), but there appeared to be no standard or common questionnaire schemes that were applied and that offered a full insight into the subjective reaction of users to devices. Typically, some schemes addressed only user 'workload' (Bates 1999, Brewster 1994); some also addressed user 'comfort' or 'ease of use' (Douglas and Kirkpatrick 1999, Murata 1991, Fernstrom 1997). In a similar manner to the composite objective metric for the method, some form of subjective measurement was required that would encompass the elements of these assessment areas in a *composite* subjective metric of user *reaction* to a device. Together, these objective and subjective metrics would give an overall balanced assessment of a device, or in effect how 'usable' a device is.

3.6 A summary of assessment methods

Target acquisition or Fitts law based approaches are simple to administer, offer a 'standard' test that can be used to assess devices and results between studies, but offer few metrics for analysis and insight into the performance of devices. In addition assessing eye based pointing with these type of tests may not give great insight into the performance of eye based pointing devices. Typically 'real world' based tests can offer a richer set of data and hence greater insight into the performance of pointing devices, but these tests rarely offer compatibility in assessment techniques or results between studies. Thus no currently found assessment method would be suitable for the assessment of head and eye mice when undertaking direct interaction with a standard graphical user interface. This showed the need for a new assessment method based on objective and subjective measurement of performance and user reaction in a new assessment method based on 'real world' interaction.

Chapter 4

Constructing Real World Test Tasks

This chapter shows the construction of a set of 'real world' test tasks suitable for monomodal and multimodal head and eye based interaction, as well as standard desktop hand mouse interaction, that can be performed on a standard graphical user interface as outlined in the requirements of Chapter 3. These test tasks are based around using a word processor and internet browser, perhaps the two most commonly used applications, for a range of typical tasks. Since users of head and eye mice would not normally use a standard desktop keyboard, as shown in Chapter 2, the tasks also include manipulation of an on-screen keyboard to allow textual input.

The chapter first discusses how test tasks are defined, in terms of the level of detail required and how the tasks are described. It then defines a taxonomy of common graphical user interface objects to quantify and qualify the properties of an interface. The chapter then uses an analysis of typical user interaction sessions to determine usage profiles for each of these interface objects on the taxonomy. Finally, based on the taxonomy and usage profiles, the chapter shows the construction of detailed typical real world test tasks, one for word processing and one for web browsing.

Throughout this chapter the aim was to minimise the number of tasks in the test, so the test would not be excessively long and would be as efficient as possible, while keeping the test as representative as possible of complex 'real world' interaction. In addition, the test tasks were to test all valid interaction types on the interface, and also allow detailed analysis of interaction by differing factors of interest, such as target size and interaction technique. By adopting this approach it was envisioned that the test tasks would produce a standardised test model representative of 'real world' interaction, that would also allow detailed analysis, and that would be more efficient than a random 'real world' interaction session.

4.1 Choosing a real world interface

There were many possible graphical user interfaces that could have been chosen for the basis of the 'real world' assessment scenario. However, it was logical to choose the most popular interface currently in use, and also the most popular interface used and supported by the devices in this work. By observation of usage in the field, and apparent availability of software drivers and specialist assistive technology applications, the interface of choice at the time of this work was Microsoft Windows (98, NT, 2000, XP). Hence the real world test tasks used in this work are based on interaction with the Windows range of interfaces. However, the work is mostly generic in nature and so could be applied to other similar graphical user interfaces.

4.2 The need for highly defined tasks

Having chosen an interface for the test tasks, the next step was to determine how the test tasks should be defined. As discussed previously in Chapter 3, for a real world set of test tasks it would be possible to simply allow a subject to 'play' randomly on an interface with the head and eye pointing devices, and then devise a method of assessing the performance of the devices. However, such wholly random interaction is not repeatable between devices or sessions, making comparisons difficult. Hence a more defined form of interaction was required - this meant that a real world test scenario was required that consisted of a series of specifically defined tasks.

Specifically defining tasks produces a problem that the tasks could be performed with differing interaction sequences, and using differing interface interaction techniques to reach the final goal of completing any given task in the test. For example, when asked to 'open a file', one subject may use a menu item to open the file, whereas another may use a toolbar button for the same task, resulting in the same outcome but by a *different interaction technique.* Alternatively, a subject may select 'italic' before typing italicised text, whereas another may type the text and then select 'italic', resulting in the same outcome, but by a *different sequence.* Hence from both of these scenarios it was clear that defining a series of tasks to perform on an interface without defining *how* these tasks should be performed, and in *what sequence* they should occur, could result in differing interaction sequences and differing interaction techniques between subjects and devices - making comparison between the individual performances of the subjects and devices difficult.

A compromise must be made away from truly natural 'random' real world interaction by setting out the test tasks in great detail to remove the possibility that different interaction techniques, and interaction sequences, are used to achieve the same goal. Adding the properties of *when* and *how* to the individual steps that made up the tasks would achieve this. Hence each task in the test would need to be given a predefined order in the test sequence, and a predefined method and set of objects that were to be used for the task. For example, to 'open a file' the test tasks would need to define that the menu must be used for this task, or for example to 'italicise text', the test tasks would define that 'italic' is selected by the toolbar button before the text is typed. This approach thus fixes *how* the tasks were to be performed and also fixes the sequence of *when* the tasks were to be performed, allowing simple and direct comparison between subjects and devices for any given task in the test sequence.

4.3 Using objects to describe tasks

To construct a series of test tasks in detail it was important to fully understand the properties and interaction requirements of the interface that facilitated the tasks. Typically, a graphical user interface may be regarded as a collection of interface components or objects that together give the interface a generalised 'look and feel' for the interface, and make up that interface. This interface-object based approach to describing an interface and the operations on the interface objects lent itself well to describing detailed test task sequences, having been used previously (Hartson et al. 1990 with User Action Notation 'UAN' and Casali and Chase 1995 also with 'UAN').

When attempting to define a task sequence in great detail it would be necessary to break down interaction to the lowest atomic level of specific object manipulations on the interface, with the test tasks constructed from sequences of these manipulations. For example, using this 'interaction grammar' to describe the actions and objects required to open a file, the task sequence could define that the 'File' menu object must be manipulated followed by the 'Open' sub-menu object. This approach clearly defined the objects (*how* the task is done) and order (*when* the task is done) of the task (Hartson et al. 1990).

4.4 Objects on a standard graphical user interface

Having based the description of the real world test tasks on detailed sequences of manipulations of interface objects; the next step was to determine the range of objects on the interface that could be manipulated. A survey of the chosen graphical user interfaces for this work (Windows 98, NT, 2000, XP, including an commonly used on-screen keyboard[1], as discussed in Chapter 2) produced a list of 33 common basic components found on these interfaces (Table 4.1). These are fully illustrated in the Appendices (Table A4.1).

[1] 'WiViK' on-screen full-function keyboard from Prentke Romich, www.prentrom.com

Typical Windows interface objects		
Static Text	Scrollbar button	Task bar button
Picture / animation	Scrollbar slider	Icon
Group Box	Scrollbar channel	Graphic hypertext link
Progress Indicator	Standard toolbar button	Large toolbar button
Text characters	Edit Box	Scroll Bar
Spin button	Window size control	Spin Control
Drop down list button	Menu	List Box
List box item	Textual hypertext link	Drop Down List Box
Window control button	Command Button	Tab Control
Check Box	Window title bar	Window
Radio Button	Start menu entry	Soft keyboard key

Table 4.1 Typical Windows interface objects

4.5 Object type – reduction of test tasks

One of the main aims of this chapter was to minimise the number of tasks in the test. Having determined the objects on the interface, the next step was to eliminate objects that the user does not need to interact with, other than to look at them, and to eliminate objects that are simply comprised of collections of other objects that would already be tested separately as part of the test tasks.

To achieve this, some knowledge of the manipulation properties of the objects on the interface was required. One method of achieving this was to classify the objects into distinct *active* and *passive* types. Here, objects that only displayed information may be classed as passive since they do not respond to manipulation and do not generate a command to the interface when manipulated (Philips et. al. 1991), whereas objects that respond to manipulation and do generate a command back to the interface when manipulated may be classed as active. In addition to active and passive, objects that may not be decomposed into simpler existing interface objects may be classed as *fundamental* whereas those that are constructed from collections or groups of interdependent fundamental objects may be classed as *composite* (Bierton and Bates 2003b). Note that collections or groupings of fundamental objects that may be manipulated independently from each other are not regarded as forming a composite object. Thus a graphical user interface and the relationships between the differing elements may be illustrated by a hierarchical structure of object types (Figure 4.1). Note that there are no passive composite objects on the Windows graphical user interface.

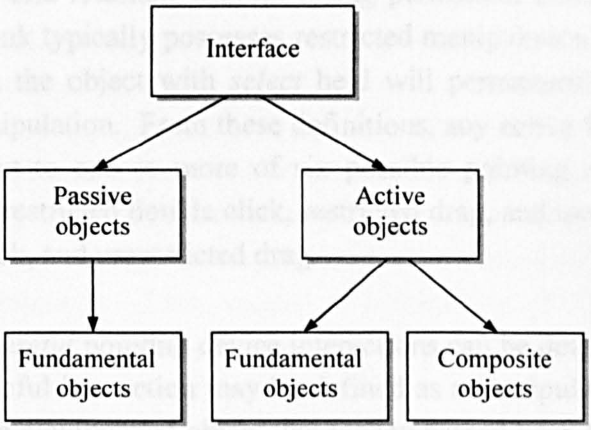


Figure 4.1 Windows object types[1]

By adopting this form of classification it became clear that test tasks would only need to test interaction with *active fundamental objects*, as interaction with composite objects can grouped from the interaction metrics of their fundamental parts and, by definition, manipulation was not required on passive objects. This approach was then applied to the list of typical object types found on the Windows interface (Table 4.1). This analysis eliminated 4 passive objects (static text, picture / animation, group box and progress indicator, all of which only display information) and 6 composite objects (scroll bar, spin control, list box, drop down list box, tab control and window, all of which are comprised of other objects) from the set of 33 candidate objects to be tested (Table 4.1). This gave a 30% reduction in the number of objects to be tested.

4.6 Object manipulation – reduction of test tasks

There are three basic forms of pointing device object manipulation on a Windows interface. These are a single click on an object, a double click on an object, and a drag of an object from one location to another on the interface (Philips et. al. 1991). Here, a click is defined as a *select* (button down) followed by a *release* (button up), and a drag is defined as a *select* followed by a cursor movement and a *release*. These actions can be either *restricted*, confined within a limited area defined by the active bounds of the object, or *unrestricted*, where the interaction area is limited only by the boundaries of the graphical user interface. For example, a button object on the Windows interface typically possesses unrestricted manipulation as, provided *select* is held, the cursor may be moved

[1] From Bierton and Bates 2003b.

away from the button and returned without losing permanent control of the button. In contrast, a hypertext link typically possesses restricted manipulation, as any movement of the cursor away from the object with *select* held will permanently lose control of the object during that manipulation. From these definitions, any active fundamental object on the interface may react to one or more of six possible pointing device manipulations: restricted single click, restricted double click, restricted drag, and unrestricted single click, unrestricted double click, and unrestricted drag.

The range of possible *useful* pointing device interactions can be determined for any object on the interface. A useful interaction may be defined as a manipulation that would result in a *command* from the manipulated object that is then passed back to the interface and so alters the state of the interface (Philips et. al. 1991). For example, an examination of the Windows interfaces showed that unrestricted double clicks were not valid interactions on these interfaces, as no objects on the interfaces responded *usefully* to this interaction. Hence this interaction type need not be tested. In a similar manner, all of the valid interaction types for the objects on the interface may be determined. These properties form a hierarchy of possible object manipulation types (Figure 4.2).

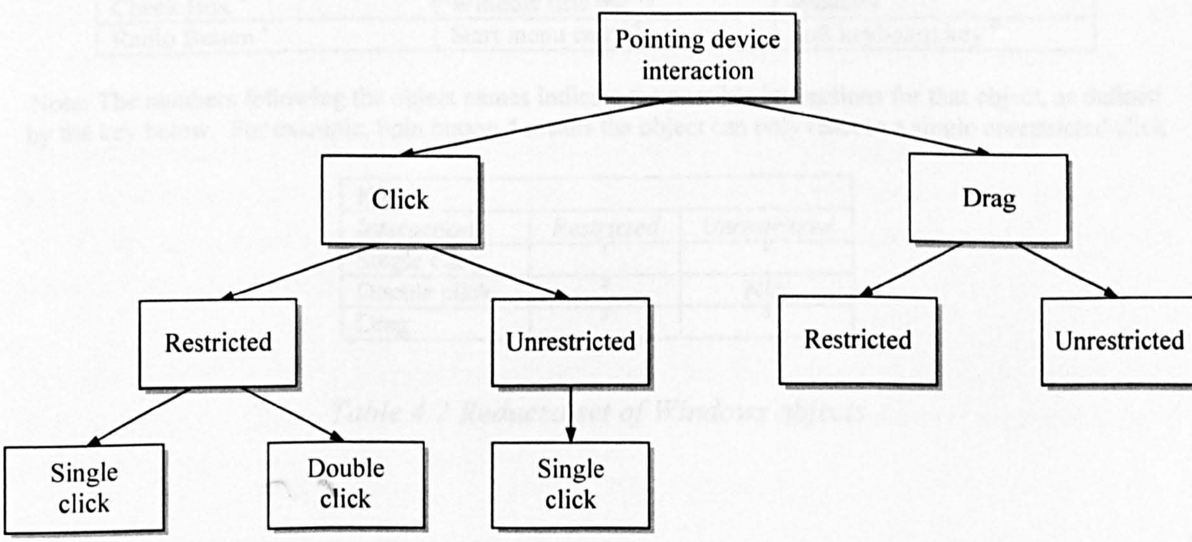


Figure 4.2 Windows pointing device interaction types[1]

[1] From Bierton and Bates 2003b.

By adopting this approach it was shown that the real world test only needed to include pointing device interaction techniques that would produce a useful command from any given object. Hence, each of the 23 remaining test objects on the interface were analysed for the response to restricted single clicks, restricted double clicks, unrestricted single clicks and both restricted and unrestricted drags. Out of a possible total of 5 interaction types with 23 objects giving 115 permutations of object type and manipulation type, this analysis removed 87 invalid permutations, leaving only 28 valid object and manipulation type permutations (Table 4.2). This resulted in a reduction of 76% in the number of permutations that needed to be tested.

Reduced set of typical Windows interface objects		
~~Static Text~~	Scrollbar button [4]	Task bar button [4]
~~Picture / animation~~	Scrollbar slider [3]	Icon [2,5]
~~Group Box~~	Scrollbar channel [4]	Graphic hypertext link [1]
~~Progress Indicator~~	Standard toolbar button [4]	Large toolbar button [4]
Text characters [2,4,5]	Edit Box [4]	~~Scroll Bar~~
Spin button [4]	Window size control [5]	~~Spin Control~~
Drop down list button [4]	Menu [4]	~~List Box~~
List box item [2,4]	Textual hypertext link [1]	~~Drop Down List Box~~
Window control button [4]	Command Button [4]	~~Tab Control~~
Check Box [4]	Window title bar [2,5]	~~Window~~
Radio Button [4]	Start menu entry [4]	Soft keyboard key [4]

Note: The numbers following the object names indicate the possible interactions for that object, as defined by the key below. For example, Spin button 4 means the object can only react to a single unrestricted click

Key:		
Interaction	*Restricted*	*Unrestricted*
Single click	1	4
Double click	2	N/A
Drag	3	5

Table 4.2 Reduced set of Windows objects

4.7 Object size – reduction of test tasks

All fundamental active objects have interaction areas or sizes that respond usefully to pointing device interaction. The size of an object may be defined in screen pixels, millimetres or, for example, with respect to head or eye-based devices, the angle subtended by the object at any given distance from the screen. (Note that the pixel sizes of objects do not change with screen resolution changes under Windows, hence pixels can be used to consistently describe object sizes. Object sizes in mm or visual angle will change due to screen resolution or user to screen distance).

By classifying fundamental active objects by their size in pixels on the interface, any objects that share the same object type and manipulation requirement, and have a similar size may be grouped into an interaction area size category. This has the effect of reducing the total number of object sizes that may need to be tested. For example, several different objects may share the same type and manipulation requirement, but have very slightly different sizes but all fall within an interaction area size category. Instead of testing all of the individual sizes, grouping the objects into a common size category would reduce the number of test permutations since testing one of these objects will, by definition, be equivalent to testing all objects with the same interaction area size category.

Typical object interaction areas were determined for the fundamental active objects remaining in object list (Table 4.2). Examining the interaction area sizes of the objects allowed the objects to be divided into interaction area categories. (Note that although the interaction area of an object is a product of both the x and y screen dimensions, when objects are highly rectangular and have one dimension much larger than the other, the interaction area size in terms of ease of selection can reasonably be described as the smaller of the two dimensions. This approach was valid for highly rectangular objects, as the smaller of the two dimensions would present the most difficulty during interaction).

This analysis gave a total of 17 object interaction area sizes ranging from 8 to 32 pixels. Many of the sizes were similar and a simple agglomerative hierarchical cluster analysis[1] of all of the object sizes suggested 4 equally spaced groups of representative interaction area sizes defined as: S1=8, S2=16, S3=24 and S4=32 pixels (Table 4.3).

Out of a possible total of 14 possible object sizes, this analysis condensed the object sizes to 4 categories. This resulted in a reduction of 71% in the number of permutations of object sizes that needed to be tested.

[1] The agglomerative hierarchical cluster analysis used the Euclidean distance between object sizes as a dissimilarity metric. This method starts with each object size forming its own separate cluster, the distance between individual clusters is then calculated and the two closest clusters are joined to form a single cluster. This process continues until all object sizes are contained within a suitable number of clusters. Four clusters of object size were chosen as the analysis quickly settled at, and remained stable for, this number of clusters.

Size groupings for Windows objects		
Object	*Size (x, y pixels)*	*Size category*
Text characters	8, 10	S1
Spin button	16, 8	S1
Drop down list button	16, 16	S2
Scrollbar button	16, 16	S2
Window control button	16, 16	S2
Scrollbar slider	16, >16	S2
Scrollbar channel	16, >16	S2
Check Box	>60, 16	S2
Radio Button	>60, 16	S2
Textual hypertext link	>16, >16	S2
List box item	30, 16	S2
Edit Box	>60, 18	S2
Standard toolbar button	22, 22	S3
Window size control	22, 22	S3
Tab	>30, 22	S3
Menu	>30, 22	S3
Command Button	>80, 24	S3
Task bar button	>80, 24	S3
Start menu entry	150, 24	S3
Window title bar	>200, 24	S3
Graphic hypertext link	≈32, ≈32	S4
Icon	32, 32	S4
Soft keyboard key	40 x 32	S4
Large toolbar button	52, 32	S4

Table 4.3 Interaction areas for Windows objects

4.8 A taxonomy of Windows interaction objects

The active fundamental elements of the interface could now be fully described by their representations, functions, interaction area categories and useful interaction techniques in a taxonomy of Windows interaction objects. This is illustrated here (Table 4.4) and fully described in the Appendices (Table A4.1).

The taxonomy marks the valid permutations of interaction area size category and useful interaction technique for the active fundamental objects present on the Windows interface with a '✓'. This clearly shows the large number of permutations that did not need to be tested to fully assess a pointing device when using the Windows interface.

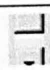
Taxonomy of Windows active fundamental objects			*Interaction area*		*Pointing device interaction*				
					Restricted			*Un – restricted*	
Object name	*Representation*	*Function*	*Description*	*Interaction area size category*	*Single click*	*Double click*	*Drag*	*Single click*	*Drag*
Typed Text	**Hello**	Display text, allow editing of text	Text character	S1		✓		✓	✓
Spin button		Change spin value	Button area	S1				✓	
Scrollbar slider		Scroll slider	Slider area	S2			✓		
Scrollbar channel		Scroll slider	Channel area	S2				✓	

Table 4.4 Taxonomy of Windows objects

The usefulness of the previous analysis and the resulting taxonomy can be illustrated with an example. Here, that the entries for both the 'Typed Text' and 'Spin button' objects overlap in their size categories and interaction requirements, with the 'Text' entry fulfilling all of the 'Spin button' requirements. Hence, unless there was some specific reason to test interaction with a spin control, the 'Spin button' entry could be removed from the real world test without removing any of the required test permutations of object size and interaction type.

The taxonomy shows the effect of the analysis. From 120 possible interaction permutations of object and interaction technique on the remaining active fundamental objects, these were reduced to a minimum of 14 required test permutations of size category and interaction technique. A reduction of 88% in the potential number of object sizes and interaction techniques that could be tested.

4.9 Summary of Windows interface interaction

A summary of the possible permutations of object sizes and interaction techniques on these objects was constructed by disregarding the nature of the individual objects. This approach was useful when simply examining the performance of a pointing device in terms of object sizes and interaction techniques. By using the taxonomy of Windows active fundamental interaction objects (Table 4.3) the active fundamental objects on the interface were grouped by size category and interaction technique so that all possible permutations of object sizes and interaction techniques were determined for the Windows interface (Table 4.5, with valid permutations marked with a '✓').

Summary of Windows object interaction					
Interaction area size category	*Pointing device interaction*				
	Restricted			*Un-restricted*	
	Single click	*Double click*	*Drag*	*Single click*	*Drag*
S1		✓		✓	✓
S2	✓	✓	✓	✓	
S3		✓		✓	✓
S4	✓	✓		✓	✓

Table 4.5 Summary of Windows object interaction

These permutations were the final product of the analysis of interaction with the Windows interface and clearly showed the essence of interaction with the interface. It is notable that a significant number of permutations of interaction type and object size do not occur and do not need to be included within the standardised test tasks, thus dramatically reducing the test lengths. The next step was to use the objects and interaction techniques in the taxonomy to form a logical sequence of real world tasks that would include all of the entries in the taxonomy.

4.10 Typical real world tasks

In constructing a set of 'typical' real world test tasks, it is difficult to determine what constitutes such a 'typical' set of real world tasks, as the nature and type of tasks performed on an interface are dominated by the final goals of the user. Perhaps the least biased approach is to simply monitor users over a period of time and record the actions they perform whilst accomplishing varied goals on the interface.

Such an analysis has been previously determined for interaction on the Windows interface in the two domains of interest to this work: word processing (using Microsoft Word) and web browsing (using Internet Explorer) (Bierton, 1999). In this work the proportions of interface object usage were measured using video recordings of the screen during interaction sessions using an on-screen keyboard for text entry and a standard desktop hand mouse for cursor manipulation. These proportions (Tables 4.8 and 4.9, column 2) give a clear insight into the activities during interaction in these domains and can be used to form the basis of a 'typical' real world set of tasks. For example, text entry using the on-screen keyboard used 50% of the total interaction time in the Word processing domain, with text editing using 20%, manipulating window objects such as toolbars and icons using 15%, manipulating menus 10%, and finally general interaction with dialog boxes using 5%.

These proportions of object usage (Tables 4.8 and 4.9), the summary of Windows interface interaction (Table 4.5) and the taxonomy of Windows objects (Table 4.4) containing the permutations of object sizes and interaction techniques can now used to construct efficient real world test tasks for both word processing and web browsing task domains.

For example:

- *Typing* takes 50% of the typical word processing task (Table 4.8) and uses on-screen keyboard keys. From the taxonomy (Table 4.4) the on-screen keyboard keys[1] used for typing required a single unrestricted click on a S4 object size. Hence 50% of the test should be taken with typing that employs single unrestricted clicks on S4 objects located in the same area of the screen. This was representative of typical real world interaction and also tested the given interaction technique and object size pair, fulfilling one entry on the object sizes and interaction techniques summary table (Table 4.5).

- *Window object* manipulation by *window sizing* would take only a small proportion of the word processing task. This required an unrestricted drag of a S3 object size. Since only a very small proportion of the test should be occupied with this activity, it could be removed from the test unless the test length was sufficiently long.

[1] 'WiViK' on-screen full-function keyboard from Prentke Romich, www.prentrom.com

However it must be included to fulfil the entry on the object sizes and interaction techniques summary table. This results in a slight distortion of the test away from an idealised real world interaction but must be done for completeness of the object size and interaction technique element of the test.

- *Dialog boxes* are a final example, using a total of 12 different fundamental objects (5 fundamental plus 2 composite containing 7 fundamental objects). All forms of dialog box interaction take a total of only 5% of typical word processing interaction so the proportion of the total test taken interacting with these objects individually will be very small. Including interaction with all of these objects in the test tasks would result in a large distortion away from an idealised real world interaction, however from the taxonomy it was clear that each of the objects shared the same interaction technique and interaction area. Thus it was possible to remove the duplicate items and, if required, only present a small number of these in the test tasks. This approach compressed the test size, whilst fulfilling the required entries on the object sizes and interaction techniques summary table and generally preserved the real world basis of the test.

The end product of this analysis was a set of 150 real world tasks, 82 in the word processing domain and 68 in the web browsing domain, based on 'typical' interaction, and fully fulfilling all valid object interaction types and sizes. This is illustrated here (Table 4.6, with the fulfilled permutations marked with a '✓') and fully described in the Appendices (Appendices Table A4.6, Figure A4.1)[1].

At the highest level the tests were simply representative of typical real world interaction so that an overall test performance could be measured. However, by including all valid permutations of object size and interaction technique on the interface a more detailed analysis of interaction may be obtained by measurement of individual object size and interaction technique performances, allowing a detailed analysis at the level of individual object sizes and interaction techniques. In addition, a very detailed step-by-step task analysis would give individual performance metrics for each of the different tasks within the test, highlighting the performance of single tasks. Using this approach gave a very powerful but simple set of tests that would enable detailed and multi-level analysis of interaction within a single test regime.

[1] The screen and test applications must be prepared before the test tasks can be commenced. Suitable pre-prepared Microsoft Word (Test File.doc) and Internet Explorer (Test1.htm) documents are available for download: www.cse.dmu.ac.uk/~rbates/test and are illustrated in Appendices Figure A4.1.

Task element	Task description	Object used	Interaction object size category	Pointing device interaction				
				Restricted			Un-restricted	
				Single click	Double click	Drag	Single click	Drag
1	Click the [Start] button on the task bar	Task bar button	S3				✓	
2	Open the Programs menu by clicking the [Programs] icon on the start menu	Start menu entry	S3				✓	
3	Start Word by clicking the [Microsoft Word] icon from the start menu	Start menu entry	S3				✓	
4	Click the [Soft Keyboard] button on the task bar	Task bar button	S3				✓	
5	Resize Word by double clicking the window title bar	Window title bar	S3		✓			
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 4.6 Test tasks for pointing device interaction

4.11 Verifying the test tasks

A check was now made to confirm that the test tasks contained all required permutations of object sizes and interaction techniques, and also that the differing proportions of tasks types were representative of typical real world interaction. The word processing and Internet browsing test tasks were analysed for the frequency and usage of object size and interaction technique combinations. Comparison of these results with the possible permutations of object sizes and interaction techniques on the interface verified that the test tasks contained entries for all of the possible interaction permutations (Compare the entries in Table 4.7 to the requirements in Table 4.5).

The word processing and web browsing test tasks were then analysed for the frequency of usage of object types to determine how representative the devised test tasks were to the typical real world object usage profiles previously determined (Bierton 1999). Comparison of the proportions of activities within each domain (comparing column 2 with column 3 of Tables 4.8 and 4.9) showed that the test tasks were representative of real world interaction. Where there were deviations from the expected proportions, these were essentially due to distortion of the test structure to include rarely used, but logically important, objects and interactions that would otherwise not be included within the test.

These interactions must be included to fulfil the requirements of the permutations of active fundamental object sizes and interaction techniques for the interface (Table 4.5). This analysis concluded the construction of the test tasks for the assessment method.

Interaction area size category	**Word processing Pointing device interaction**					**Web browsing pointing device interaction**				
	Restricted			Un – restricted		Restricted			Un –restricted	
	Single click	Double click	Drag	Single click	Drag	Single click	Double click	Drag	Single click	Drag
S1		1		3					1	1
S2		1	1	7		4		2	11	
S3		1		15	2		1		15	
S4		1		49	1	2			30	1

Table 4.7 Test tasks verification by Windows object interaction

Proportions of object usage for word processing tasks			
Word processing tasks	*Proportion of interaction*[1]	*Proportion of test tasks*	*Interface objects (*denotes composite)*
Typing	50%	60%	Keyboard key
Editing	20%	5%	Keyboard key Toolbar button
Window Objects	15%	15%	Toolbar button Scrollbar* Icon Start menu entry Task bar button Window*
Menus	10%	7%	Menu
Dialog Boxes	5%	13%	Check box Radio button Edit box Tab Spin control* List box* Command button

Table 4.8 Comparison of word processing test tasks to real world proportions

[1] From Bierton 1999.

Proportions of object usage for web browsing tasks			
Web browsing tasks	*Proportion of interaction*[1]	*Proportion of test*	*Interface objects (* denotes composite)*
Web Navigation	45%	12%	Textual hypertext link Graphic hypertext link
Window Objects	30%	30%	Large Toolbar button Scrollbar* Icon Start menu entry Task bar button Window*
Typing	10%	38%	Keyboard key
Menus	5%	3%	Menu
Dialog Boxes	5%	7%	Check box Radio button Edit box List box* Command button
Editing	5%	10%	Keyboard key Menu

Table 4.9 Comparison of web browsing test tasks to real world proportions

4.12 A summary of real world test tasks

'Real world' tasks are presented to fulfil the requirements of a 'real world' set of tasks as outlined in Chapter 3. The tasks go further than this requirement and contain both typical interaction and interaction with all permutations of object size and pointing device interaction technique present within the given test task domains. Each step in the test tasks is based on a single fundamental object size type and interaction technique so that analysis of the test results can be undertaken at the most fundamental level of interaction with the interface. The proportions of object usage within the test tasks are as close as possible to those previously determined (Bierton 1999) so that the test tasks closely mimic real world interaction. Each set of test tasks contains the object sizes and interaction techniques associated with objects that would be used during typical interaction within the relevant task domain. The tests can be used separately to determine pointing device performance within a single task domain or together to determine performance over both domains in the tasks to determine performance over the complete set of object sizes and interaction techniques.

[1] From Bierton 1999.

Chapter 5

Measuring the Performance of Pointing Device Interaction

This chapter shows the construction of a pointing device performance assessment scheme, introduced in Chapter 3, to enable detailed examination of the performance of hand, head and eye mice on the test tasks constructed in Chapter 4. It uses the diverse metrics available for assessing pointing device performance, and from these constructs a novel detailed performance assessment scheme based on complex metrics of objective multi-factor time and error compositions, and subjective user responses. Throughout this chapter, the aim is to construct metrics that will give a detailed insight not only into the performance of the pointing devices in this work, but also to indicate how the performance of the head and eye mice devices might be enhanced.

5.1 Objective and subjective metrics

As discussed in Chapter 3, the assessment of a device can be expressed as two components: the objective performance of the device, and the subjective user reaction to that device. Together these enable the *usability* (Bevan 1991) of the device to be assessed. The objective performance may be measured as a composite of task times and task quality or errors during interaction, and subjective reaction in terms of the evaluation from the test subject when using that device to perform the tasks. Ideally, to gain maximum insight into the devices, these metrics should be multi-factor, detailed and hence complex to fully assess the performance of the device, but also composite to present the results in a simple manner, and validated to show that they measure what they claim to measure.

Looking for suitable methods of expression for these metrics leads to the definitions of device *efficiency* and *satisfaction* as stated in the European ESPRIT MUSiC (Metrics for Usability Standards in Computing) performance metrics method (Bevan et al. 1991 and 1995, MacLeod et al.1997) and the recommendations outlined in the ISO 9241 Part 11 'Guidance on Usability' International Standard (Smith 1996). These metrics were defined as follows:

- *Efficiency*: the objective performance of the pointing device, expressed in terms of the amount and quality of interaction with the device and the time taken to perform that interaction.

- *Satisfaction*: the subjective acceptability of the pointing device, expressed in terms of the user workload and comfort when using the device and the ease of use of the device.

5.2 Measuring efficiency

Efficiency was defined as a composite of the amount of a task accomplished, the quality of the interaction during that task, and the time taken for the task. Examining the MUSiC performance definitions in detail (Bevan et al. 1991 and 1995, MacLeod et al.1997) gave the following relationships (Figure 5.1):

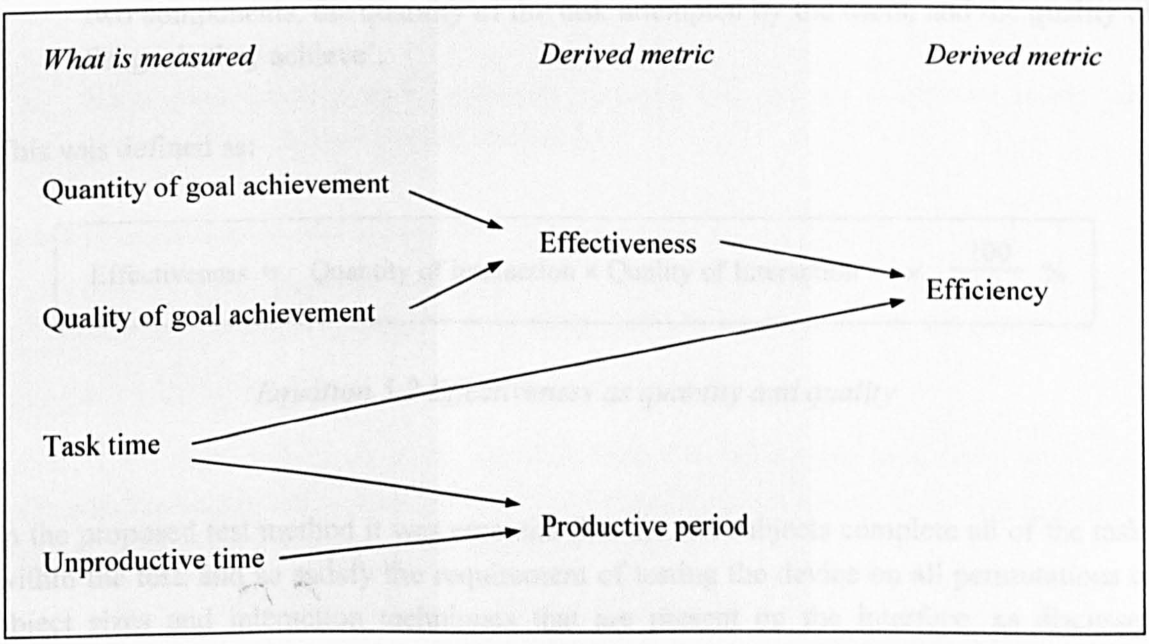


Figure 5.1 Relationship between measures and metrics[1]

Here *Efficiency* was described as follows:

- 'The Efficiency with which users use a […] product is defined as the ratio between their Effectiveness in carrying out their task, and the time it takes them to complete the task'.

[1] From Bevan et al. 1995.

Where efficiency was defined as:

$$\text{Efficiency} = \frac{\text{Effectiveness}}{\text{Time taken for interaction}} \times \frac{100}{1} \%$$

Equation 5.1 Efficiency as effectiveness and time

And effectiveness was described as:

- 'The Effectiveness with which users [...] carry out a task is defined as comprising two components, the quantity of the task attempted by the users, and the quality of the goals they achieve'.

This was defined as:

$$\text{Effectiveness} = \text{Quantity of interaction} \times \text{Quality of Interaction} \times \frac{100}{1} \%$$

Equation 5.2 Effectiveness as quantity and quality

In the proposed test method it was essential that the test subjects complete all of the tasks within the test, and so satisfy the requirement of testing the device on all permutations of object sizes and interaction techniques that are present on the interface, as discussed previously in Chapter 3. Hence the *quantity* of interaction will be 100% for this test method and can be removed from the model.

Substituting Equation 2 with Quantity equal to 100% into Equation 5.3 gives the final calculation for efficiency:

$$\text{Efficiency} = \frac{\text{Quality of interaction}}{\text{Time taken for interaction}} \times \frac{100}{1} \%$$

Equation 5.3 Efficiency as quality and time

5.3 Measuring task time

Task time was simple to quantify (Bevan et al. 1991 and 1995, MacLeod et al.1997), and was defined as:

- 'The time a user spends using a system to perform the evaluation task'.

And was further described as:

- 'Task Time begins when the user starts to interact with the product [...] and ends when the user indicates he or she has finished'.

With unproductive time defined as:

- 'How long the user took performing actions that did not contribute to the task output' (those defined later in section 5.4)

Hence productive time was defined as:

- 'The proportion of time the user spent performing actions that contributed to the task output'.

These were clear definitions, with the task time defined as the total time for a task, including any unproductive time, with the additional division of task time into productive and non-productive elements giving additional detail.

5.4 Measuring quality

As discussed in Chapter 3, this could typically be a count of errors generated during the task. However this 'pass/fail' approach was regarded as crude and it was expected that it would not give any great insight into what factors caused any errors that were counted. A more subtle approach was needed.

Hence quality of interaction (Bevan et al. 1991 and 1995, MacLeod et al.1997) was defined as:

- 'How good the attempt is'.

And was further described as:

- 'Quality is a measure of how good the task goals represented in the output are compared to their ideal representation. It is defined as the degree to which the task goals represented in the output have been achieved'.

A method was suggested for specifying quality (Bevan et al. 1991 and 1995, MacLeod et al.1997):

- '1. Decide what constitutes an ideal output of each goal'.
- '2. Specify a scoring procedure for measuring how good the output of each goal is compared to its ideal, that also takes into account any output that was not asked for. If the task goals vary in importance, a weighting can be applied'.

From this a definition and scoring procedure was required for quality. It was logical to state that an ideal output for the goal would be a 'perfect' cursor movement onto the task target object followed by 'perfect' manipulation of that target. Here 'perfect' cursor movement could be defined as no cursor movement deviation from a straight path from start point to end point on the target. This cursor movement was well defined (MacKenzie 2001) by measuring path deviations or pauses in cursor movement. These can be termed cursor *control corrections* (Figure 5.2).

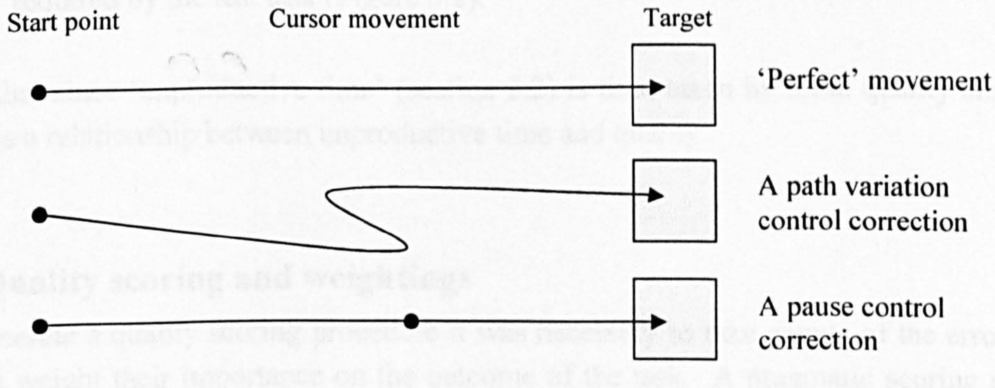


Figure 5.2 Measurement of cursor control corrections[1]

[1] From McKenzie 2001.

The number of control corrections can be measured by counting the number of path variations or pauses of cursor movement during the task (MacKenzie 2001). These variations and pauses clearly indicated a lack of control when compared to an idealised 'perfect' cursor movement as they generated output that was not asked for, and hence gave a reduction in task quality.

Secondly 'perfect' manipulation of the target would entail the correct selection of the correct target. Any deviation from this ideal can result in either a complete *miss* of the target with the selection action or the selection of the *incorrect* target.

Hence we have three elements that may be used to form a scoring scheme for quality that is more detailed than a simple count of errors. These were defined as follows:

- The number of *incorrect commands*, where an incorrect command is generated by the accidental selection of an active (Chapter 4) part of the interface, such as a button or menu item, that was not required by the test task.

- The number of *target misses*, where a miss is generated by the accidental selection of a passive (Chapter 4) part of the interface, such as a window background or the desktop, which was not required by the test task.

- The number of *control corrections*, where a control correction is caused by additional unwanted cursor movements required to move the cursor onto the target required by the test task (Figure 5.2).

Note that since 'unproductive time' (section 5.2) is time taken by these quality elements; there is a relationship between unproductive time and quality.

5.5 Quality scoring and weightings

To generate a quality scoring procedure it was necessary to take counts of the error types and to weight their importance on the outcome of the task. A pragmatic scoring scheme was used by (Szczur 1994) where interaction quality was rated on a 1 'failed' to 5 'near perfect interaction' scale. In this work tasks were given an initial score of 5 'perfect' with any subsequent degradation in performance reducing the score. When the score was reduced to 1, the task was deemed failed. This scheme was adopted for the assessment method.

The next task was to apply weightings to the error types. Logically, a *control correction* would have less impact on task completion than a *target miss*, and again, a *target miss* would have less impact than an *incorrect command* generated. So, assuming that generating incorrect commands severely degraded interaction quality (as these commands required further interaction to undo the mistake) they should be weighted heavily. Target misses had less impact on interaction (by requiring only a second target manipulation attempt) and so were rated less heavily. Finally control corrections were weighted the least heavily (requiring only a correction in cursor position) as these had the smallest effect on the quality of interaction.

It was difficult to determine any form of *exact* weighting, except that by adopting a 1-5 scoring scale, any error that had a major impact on quality should be weighted to occupy a large part of that scale, but not so highly that such an error would immediately fail the task without the subject having the opportunity to overcome the error. Hence, incorrect commands were weighted as three times their count, the majority of the scale but still allowing some additional opportunity to incur lesser errors whilst recovering from an incorrect command. Being of the least impact on quality, control corrections were weighted at the least part of the scale, with a value of one times their count. This left target misses, which lay in importance between the two and were hence weighted at twice their count. (It should be noted that the value of these weightings is subject to verification later in this work in Chapter 8).

Thus the quality scoring weightings were defined (Figure 5.3):

- 3 × count of incorrect commands
- 2 × count of misses
- 1 × count of control corrections

Figure 5.3. Error count weightings

Each task within the test would initially given a quality rating of 5. As the quality of interaction during the task was degraded by the weighted counts of the error types, so the rating was reduced until either the task is completed or the quality rating is reduced to 1, at which point the current task was regarded as failed.

This use of task failure leading to task abandonment and the start of the next test task avoided scenarios where a test subject would spend excessive time attempting a difficult task that might otherwise be abandoned or achieved in a different manner under non-test conditions and allowed a 'natural' flow of interaction through the test. To achieve this, quality should be roughly measured in real time during tests, with the task abandoned when quality was reduced to 1 or below. Combinations of more than one quality element within a test task were cumulative and the final interaction quality was hence calculated by a simple formula:

$$\text{Quality} = 5 - \left[\begin{array}{l} (3 \times \text{count of incorrect commands}) + (2 \times \text{count} \\ \text{of misses}) + (1 \times \text{count of control corrections}) \end{array} \right]$$

Equation 5.4 Calculation of task quality

Taking these definitions and weightings and the scoring scheme descriptions previously described (Szczur 1994) enabled a table of definitions of quality to be constructed to enable defined and quantitive quality measurement to be conducted (Table 5.1).

Quality rating scheme			
Rating			*Description*
Fail	1	Very low	Very low quality of interaction due to excessive incorrect commands, misses or control corrections
Pass	2	Low	Low quality of interaction with 1 incorrect command or 1 miss and 1 control correction, or 3 control corrections
	3	Medium	Medium quality of interaction with no incorrect commands, 1 miss or 2 control corrections
	4	High	High quality of interaction with no incorrect commands, no misses and only 1 control correction
	5	Very High	Near ideal interaction with no incorrect commands, no target misses and no control corrections

Table 5.1 Interaction quality rating scheme

5.6 Efficiency as a percentage

A simple and easy to understand method of expressing efficiency was desirable. Combining the definition of efficiency (Equation 5.3) with the definition of quality (Equation 5.4) allowed efficiency to be defined as follows:

$$\text{Efficiency} = \frac{\text{Quality of interaction (1-5)}}{5 + \text{Time taken for interaction (secs)}} \times \frac{100}{1} \%$$

Equation 5.5 Calculation of efficiency

Adding a constant of '5' to the divisor matched the dividend score range of 1 to 5, enabling a percentage result to be calculated. Thus a 'perfect' task that had the highest level of quality and took no time would give an efficiency of 100%, with any reduction in quality or increase in time degrading the measured efficiency. This simple efficiency metric should be calculated at an individual task level, and is scalable and may be applied at a task-by-task level, or aggregated at a group of tasks level or for the whole test to give an overall efficiency for a device on the test.

5.7 Measuring satisfaction

In order to measure the subjective response of the user to the device, it was necessary to know *by what factor* and *by what amount* the user was influenced by the device:

- 'Measuring user satisfaction, or the acceptability of a system, requires knowledge of the internal state of the user'. (Bevan 1991).

A survey of literature (see Chapter 3) found that there were a multitude of differing questionnaires being applied to device evaluation, all assessing some aspect of the subjective reaction of the user to a device (For example: ISO 1998, Smith 1996, Douglas 1999). Parameters such as 'actuation force', 'movement smoothness', 'accuracy of pointing', 'operation speed' and 'difficulty of use' together with workload parameters such as 'mental effort', 'physical effort', 'body fatigue' and 'body comfort' were used. However, none of these sources offered a comprehensive set of questionnaire factors that fully addressed the expected assessment needs of the assistive technology pointing devices in this work. For example, none assessed factors such as eye comfort, pointing speed, frustration etc. The most appropriate course of action was to take the most suitable assessment factors for the hand, head and eye mice from a range of questionnaires and assemble a new questionnaire assessment scheme suitable for the devices to be tested. This approach was not novel, with customised questionnaires being used previously for device assessment (Brewster 1994, Douglas and Kirkpatrick 1999).

Subjective 'satisfaction' has been defined at the start of this Chapter as a composite of the amount of user workload exerted when using the device, the level of comfort experienced when using the device, and the ease of use of the device. Hence three areas need to be addressed: workload, comfort and ease of use.

5.8 Measuring workload

Searching for suitable workload factors, the MUSiC method (Bevan et al. 1991 and 1995, MacLeod et al.1997) gave the following definition for workload:

- 'Measures of cognitive workload are provided by the SMEQ (Subjective Mental Effort Questionnaire), and TLX (Task load Index) questionnaires, and by heart rate variability measures'.

Of these three measures of workload, none were commonly used for pointing device assessment, but of these, the NASA Task load index (Hart et al, 1988) was more commonly used (Bates 1999, Brewster 1994), and was perhaps the most simple, and non-invasive, to apply.

The NASA Task Load Index was based upon a multi-dimensional rating procedure that provided an overall workload score based on an average of ratings on six workload subscales: Mental, Physical, Temporal, Performance, Effort, and Frustration (Hart et al, 1988). In normal application the TLX requires two passes to apply paired comparisons and hence weightings to the ratings. However this appears to be unnecessary and a 'raw' form may be used, where the workload topics are treated as simple questionnaires with the result averaged and no second pass required (Byers et al. 1989), thus simplifying the application of the rating procedure. One workload factor was found that tended to exhibit duplication, with evidence showing that Effort is an effective amalgamation of the Mental and Physical factors when used in Human Computer Interaction assessment (results from Bates 1999, Brewster 1994). There was a desire to make the questionnaire concise and without confusing duplication, hence the Effort factor was dropped, leaving the remaining five factors. These gave the following workload factors for the questionnaire (Figure 5.4):

- Physical effort
- Mental effort
- Time pressure
- Frustration
- Performance

Figure 5.4 Workload factors

5.9 Measuring comfort

Searching for suitable user comfort factors found suggestions in the ISO 9241 Part 9 'Non-keyboard Input Device Requirements' International Standard (ISO 1998, Smith 1996) and 'Testing Pointing Device Performance and use Assessment with the ISO9241, Part 9 Standard' (Douglas et al. 1999). In these, specific body areas were defined to suit the requirements of the test and subjects asked to rate their level of comfort (or discomfort) for these areas. Typical examples included 'headache', 'wrist ache' and 'finger ache' for a desktop hand mouse. With this precedence, and evaluating which areas the devices in this work were likely to influence and the abilities of the expected user groups, the following areas were selected as factors for the questionnaire (Figure 5.5):

- Headache
- Eye discomfort
- Facial discomfort
- Mouth discomfort
- Neck discomfort

Figure 5.5 Comfort factors

Note that the facial and mouth factors were included to allow the questionnaire to be used to assess the performance of facial and mouth operated selection devices, such as eyebrow switches, eye blink switches, and sip-puff switches, that are often used with the devices in this work. The aim of this work was to include these devices in a later assessment of the outcomes of this work with disabled users, and so inclusion of these factors was necessary within the questionnaire assessment scheme. Also note that trade-offs between a user accepting higher facial discomfort against another accepting greater eye discomfort are difficult, however an overall rating would give an indication of overall discomfort.

5.10 Measuring ease of use

Finally, searching for suitable device ease of use factors again found suggestions in the ISO 9241 Part 9 Standard (ISO 1998, Smith 1996, Douglas et al. 1999). In a similar manner to comfort (Section 1.5.12) specific device properties were defined to suit the requirements of the test and subjects asked to rate their perceived level of ease of use of the device for each property. Typical examples included 'speed of pointing' and 'ease of system control'. Again with this precedence, and evaluating which property the devices in this work are likely to exhibit, the following properties were selected as factors for the questionnaire (Figure 5.6):

- Accuracy of pointing
- Speed of pointing
- Accuracy of selection
- Speed of selection
- Ease of system control

Figure 5.6 Ease of use factors

5.11 A summary of questionnaire factors

The previous sections of this Chapter found appropriate factors for the subjective satisfaction questionnaire, giving a questionnaire of three sections each comprised of five individual factors (Table 5.2).

At the most fundamental level each factor may be reported individually, in addition factors may be amalgamated (in the same manner as the NASA-tlx, by simple averaging) within their sections to give ratings for workload, comfort and ease of use. Note that aggregating all sections to form a single satisfaction result would be invalid, as each section assesses a different aspect of the subjective response of the subjects to the device. The next step in construction of the subjective assessment questionnaire was to find suitable questionnaire scales for the assessment.

Satisfaction assessment areas and factors			
Area	*Workload*	*Comfort*	*Ease of use*
Factors	Physical effort	Headache	Accuracy of pointing
	Mental effort	Eye discomfort	Speed of pointing
	Time pressure	Facial discomfort	Accuracy of selection
	Frustration	Mouth discomfort	Speed of selection
	Performance	Neck discomfort	Ease of system control

Table 5.2 Satisfaction assessment areas and factors

5.12 A summary of measuring pointing device interaction performance

Objective and subjective metrics were constructed to assess the performance of pointing devices when undertaking the test tasks outlined in Chapter 4. The objective metrics were based on detailed task time and interaction quality measurements to form an overall objective measurement of device efficiency, with the subjective metrics based on detailed multi-factor questions to form a comprehensive subjective assessment of user reaction to the devices.

The metrics allow insight not only into the performance of the pointing devices in this work, but also to indicate how the performance of the head and eye mice devices might be enhanced. Based on the very diverse ranges of current questionnaire scales, the chapter concluded that a suitable questionnaire scale was required for accurate subjective measurement.

Chapter 6

Assessment Scales

This chapter first discusses the lack of suitable assessment scales in previous work, and then shows the construction of a new assessment questionnaire scale suitable for the subjective part of the device performance metrics defined in Chapter 5. It discusses how assessment questionnaire scales are constructed, and then selects a range of candidate quantifiers suitable for pointing device assessment scales. The chapter then conducts an experiment on those derived quantifiers to determine their range and distributions. Finally this chapter uses the experimental results to show the derivation of a range of possible scales potentially suitable for the assessment of hand, head and eye mice.

6.1 Choice of questionnaire assessment scales

A literature search of the types of subjective usability and workload questionnaire assessment scales commonly used in Human Computer Interaction research was conducted and found that assessment scales for pointing device assessment were rarely used, and that there appeared to be variation in fundamental questionnaire scale design, with no one consistent design employed for assessment. Scales were found ranging from four, five and twenty intervals with a variety of differing labelling schemes:

- Four intervals with full labelling of all points assessing "slowest / least accurate / liked the least", to "quickest / most accurate / liked the most" has been used for assessing touch pad preferences (MacKenzie and Oniszczak 1998).

- Five intervals with end labels only assessing "easy / too fast / uncomfortable", to "difficult / too slow / comfortable" has been used to assess joystick and touch pad performance (Douglas and Kirkpatrick 1999).

- At the other extreme, twenty interval end only labelled scales have been used for assessing "low to high" workload factors when using a sound enabled interface (Brewster 1994), or an eye tracker for target selection (Bates 1999).

These few examples of questionnaire usage used scales that seemed to be chosen and used with no sound scientific justification given, and none showed any validation of the chosen scales. This lack of a common standard design hinders the comparison of usability results between studies, resulting in a wide range of isolated and often dissociated results. In addition, this variation gave no indication as to what would be a good design for the test method. Clearly it was necessary to address this problem by generating suitable validated questionnaire scale designs for the test method.

6.2 Designing scales

Having made the decision to construct a new assessment scale, the next step was to address the fundamental design problems for such a scale: number of intervals and labelling names and layout scheme. For instance, should a scale with 5 choices (number of intervals) named with 'low' and 'high' (names of labels) with just the end intervals labelled (layout scheme) (Figure 6.1), or perhaps it should have 11 intervals with label naming 'easy', 'medium' and 'difficult' with the end intervals and the middle interval labelled?

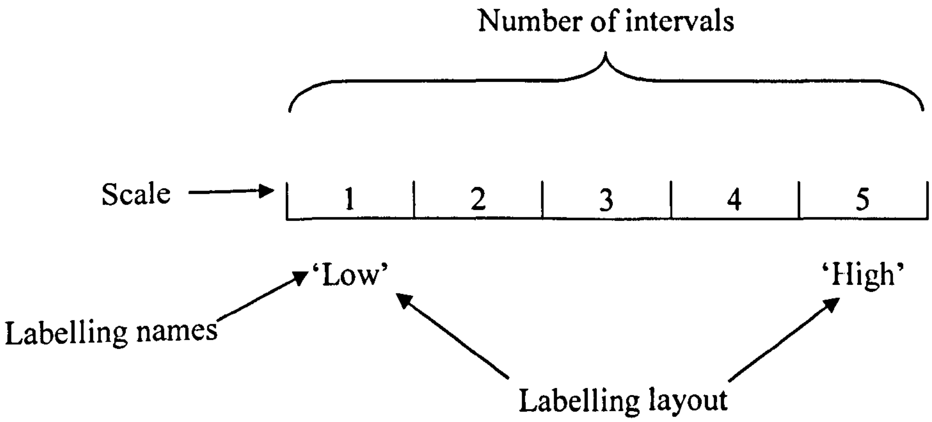


Figure 6.1 Example questionnaire scale

6.3 Generating suitable questionnaire quantifiers

The first step in designing a questionnaire scale was to determine the names of the labels, or *quantifiers*, used to annotate the questionnaire scale. Quantifier sets for questionnaire design have been produced previously, notably by (Bass 1974) and also (Spector 1976) and (Schriesheim 1974).

These sets were, however, not entirely suitable for the type of questions found in usability and workload questionnaires. For example, the set produced by Bass had typical recommended quantifiers for 7 interval scales that comprised: 'All', 'An extraordinary amount of', 'A great amount of', 'Quite a bit of', 'A moderate amount of', 'Somewhat', 'None'. Examining the range of questionnaire topics we wish to use, such as 'Mental effort' or 'Speed of pointing' it was clear that the Bass quantifiers were not suitable for such usability and workload questions. For example, questions such as 'Rate the speed of pointing?' cannot reasonably have the choices 'None' or 'All'. Modification of these end-point terms to allow the quantifiers to be used was inadvisable; as such modification would unbalance the original anchor points of the scale. In addition, examining the questionnaire topics to be used (Chapter 5) showed that two types of questionnaire would be needed, one for rating values that were bipolar, (ranging from a negative value, through a null point, to a positive value) and another for rating values that are unipolar, (ranging from nothing to a higher positive value) (Figure 6.2).

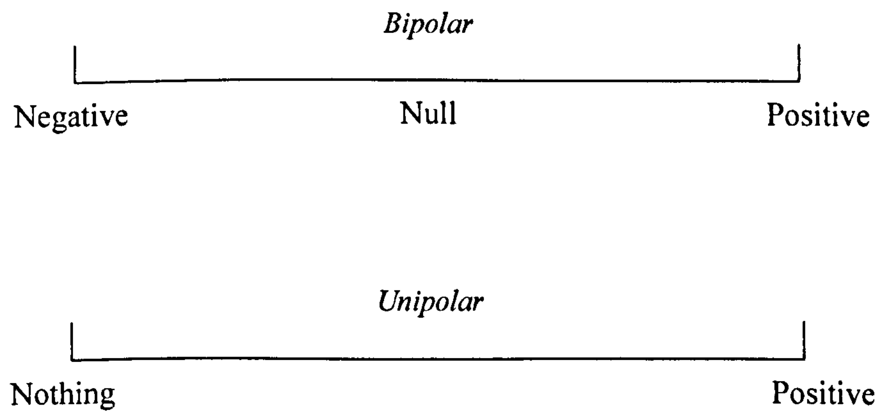


Figure 6.2 Bipolar and unipolar scale types

For example, rating 'pointing accuracy' would be bipolar, with a scale ranging from 'very inaccurate pointing' through a null mid-point to 'very accurate pointing' whereas rating 'headache' would require a unipolar scale with a range from 'no pain' through to 'a lot of pain'. Examination of previous work (Bass 1974, Spector 1976, Schriesheim 1974) showed that these did not present results for both bipolar and unipolar quantifiers. Hence for these reasons it was felt necessary to re-examine these quantifier terms and generate a new set of suitable bipolar and unipolar quantifiers.

6.4 Candidate quantifiers

In order to start with a wide and evenly distributed range of candidate quantifiers for the questionnaire scale design most of the quantifiers were derived from the very comprehensive range of 44 quantifiers previously tested (Bass 1974). Absolute quantifiers such as 'All' and 'None' were omitted, as they were not regarded as suitable for usability and workload questions. In addition, long phrases were truncated to their descriptive adverbs, for example 'A moderate amount of' was truncated to 'Moderately' and 'Quite a bit of' was truncated to 'Quite'. Finally an informal panel discussion between colleagues involved in Human Computer Interaction research reviewed the new set of candidate quantifiers and contributed three additional quantifiers not present in the original Bass set, plus the keyword (the subject described by the quantifiers, for example, 'easy', 'difficult' etc.) on its own, giving a total of 20 candidate quantifiers (Table 6.1).

Note that modification of existing quantifiers and inclusion of quantifiers contributed by a simple discussion is valid, as the validity of all of the quantifiers will be assessed later in this chapter to determine their actual subjective value.

Candidate quantifiers	
A bit	Not at all
A little	Not very
Considerably	Really*
Extremely	Pretty much
Fairly	Quite
Greatly	Scarcely
Just*	Somewhat
Moderately	Slightly
No quantifier*	Very
(just the keyword on its own)	Very much
	Very slightly*

* Added by panel discussion

Table 6.1 Candidate quantifiers[1]

[1] From Bass 1974.

6.5 Methods of quantifier estimation

The next step was to find an order, or subjective value, for the candidate quantifiers. Previous work (Oppenheim 1992, Edwards 1957, Bass 1974, Schriesheim 1978) showed that paired comparison between quantifiers and magnitude estimation of quantifiers were the two most commonly used methods of ordering and valuing a set of names. Magnitude estimation required a set of subjects to rate where on a magnitude rating line, or psychological continuum of infinitely small increments (Edwards 1957), labelled from a minimum to a maximum value, they felt each of the quantifiers would lie. These positions along the continuum then constituted the subjective values of the quantifiers (Figure 6.3).

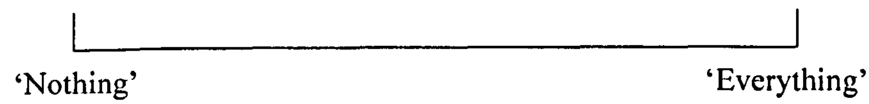


Figure 6.3 A 'psychological continuum'[1]

In contrast, paired comparison presented a set of subjects with all permutations of pairs of the quantifiers to be rated, the subjects then indicating which of each pair is the greater or smaller. The ordering and relative positions of the quantifiers could then be calculated. Typical examples of usage of these methods included (Bass 1974) with magnitude estimation, and (Schriesheim 1974) with paired comparison. Of these techniques, magnitude estimation appeared to give more valid results with (Schriesheim 1978) finding that questionnaire label generation by magnitude estimation gave fairly invariant interval points, whereas paired comparison gave poor interval points but did still preserve ranking order. Looking in detail at the techniques, (Edwards 1957) found that paired comparison with large candidate quantifier sets was also 'tedious' with $n(n-1)/2$ judgements required to pair all n quantifiers. Hence, with a large candidate set of 20 quantifiers, the method of magnitude estimation was chosen and a rating experiment devised.

6.6 Bipolar and unipolar rating continua

A psychological continuum (Figure 6.3) was required for the rating experiment to allow magnitude estimation of the values of the candidate quantifiers. This needed to be easy to understand and have a subject, or labelling, for the continuum that was easily valued or

[1] From Edwards 1957.

estimated by experimental subjects. Hence the end point labels or keywords 'Happy' and 'Sad' were chosen as it was felt that subjects could easily and quickly rate different levels of 'happiness' and 'sadness'. The continuum was a simple 10cm long line printed on a blank sheet of A4 paper. Bipolar and unipolar magnitude-rating lines were used. The bipolar line was anchored with 'Extremely sad', a negation of the highest ranked quantifier used on the left, 'Neither Happy nor Sad' as the null mid-point and 'Extremely Happy', the highest ranked quantifier (Bass 1974) on the right. The unipolar line was anchored with 'Not at all happy', the lowest ranked of the chosen quantifiers on the left and 'Extremely happy' the highest ranked of the quantifiers on the right (Bass 1974) (Figure 6.4).

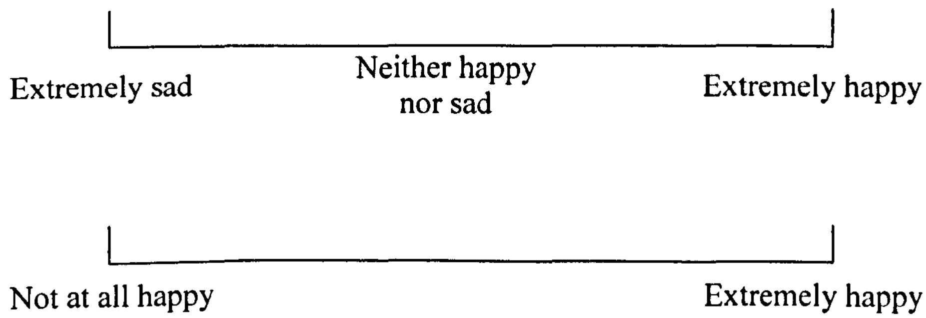


Figure 6.4. Bipolar and unipolar psychological continua

6.7 Experiment 1: Rating candidate quantifiers

An experiment was conducted with 50 test subjects, 33 male and 17 female, that were randomly chosen from volunteer students and staff at the university. Subject age ranges were 1 subject < 20, 35 subjects 20-29, 12 subjects 30-39, 1 subject 40-49, 1 subject 50+. Subjects were unpaid and were told that the tests were part of research work in the university and that they should complete the tests as accurately as possible. There was no time limit to complete the test and no penalty for non-participation, all subjects signed a consent form (Appendix Figure 6.1) and all data was anonymous. For the experiment subjects were given a randomly numbered list of the candidate quantifiers (Table 6.1) appended with the keywords 'Happy' and 'Sad' and asked to indicate where on the psychological continua lines (Figure 6.4) they felt each of the quantifiers fell by writing the number of the candidate quantifier on the psychological continua line. To eliminate order effects in the testing, the presentation order of the quantifiers and the continua types were randomly ordered on the test. To determine the value of the ratings for each quantifier the ratings of the bipolar quantifiers were scored from -100 on the left end of the psychological continuum line, through zero at the centre point, to +100 on the right end of the line and the unipolar quantifiers scored from zero on the left end of the line to +100 on the right end of the line. The results for each quantifier were ranked by the 50th percentile (median) of the ratings, with the spread of rating values represented by the 25th and 75th percentiles, giving the interquartile ranges. The results were calculated (Appendix Tables A6.5 and A6.6) and summarised graphically in order of the median points and ranges of the quantifiers (Figures 6.5 and 6.6).

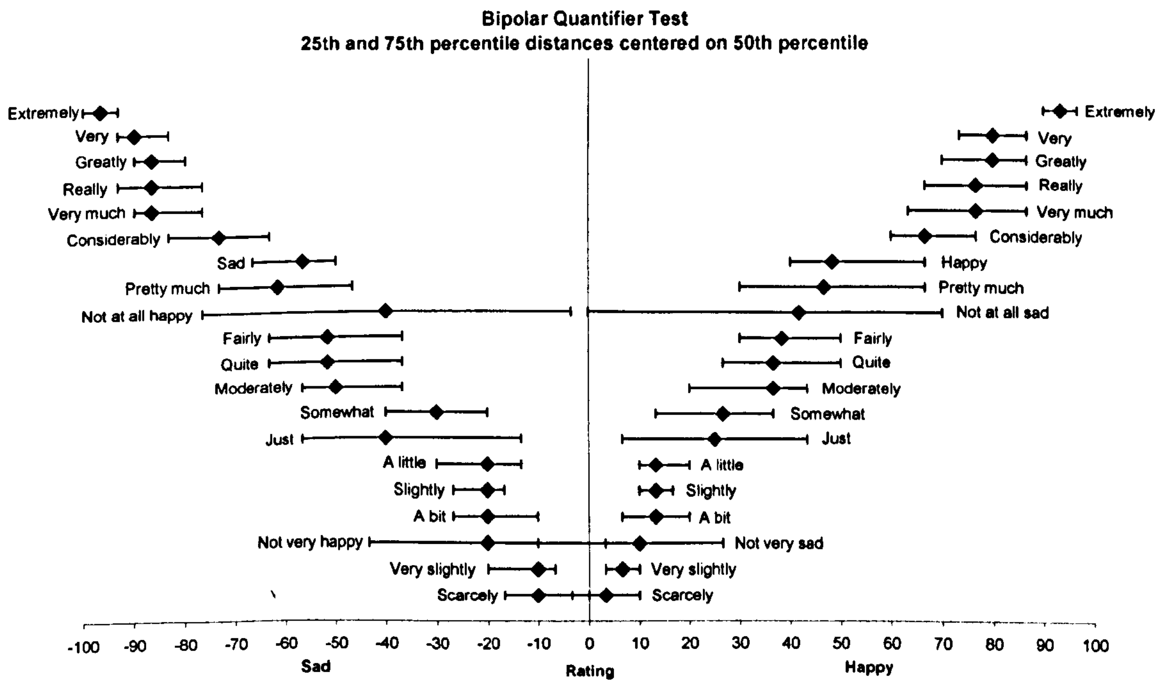


Figure 6.5 Bipolar quantifier ratings, median with interquartile range

Examining the ratings of the quantifiers on both the bipolar and unipolar graphs (Figures 6.5 and 6.6) showed that the original choice of candidate quantifiers (Table 6.1) was valid with the candidates fairly evenly spread across the psychological continuum. This suggested that the range of quantifiers chosen for the experiment were capable of covering the required range of ratings with no appreciable gaps. There was a strong suggestion of symmetry about the centre point on the bipolar continuum, indicating that the quantifiers could be used to construct a valid bipolar scale. In addition, the unipolar ratings suggested a similarity with the positive 'Happy' half of the bipolar continuum, indicating that the subjective ratings of the quantifiers were fairly stable when applied to both unipolar and bipolar continua. It was notable that the 'Not at all' and 'Not very' quantifiers had comparatively wide distributions, suggesting that the inclusion of a negation prefix to a quantifier confused the perceived value of the quantifier.

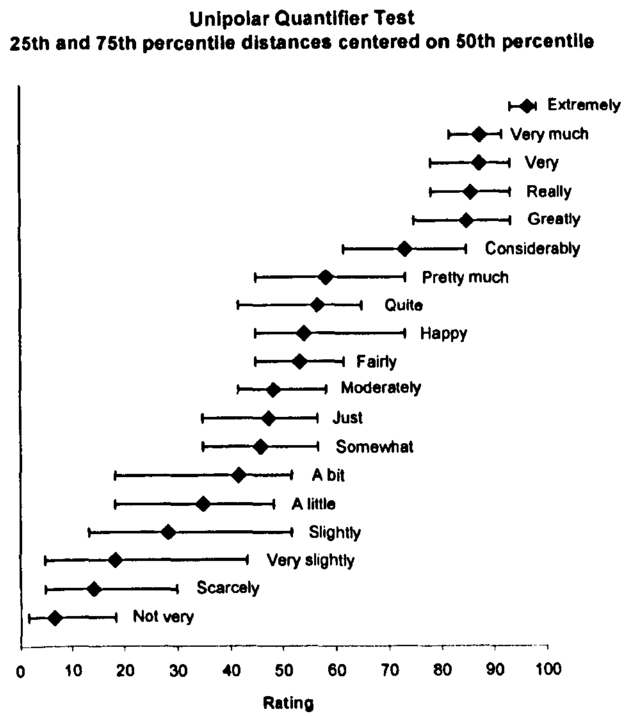


Figure 6.6 Unipolar quantifier ratings, median with interquartile range

6.8 The bipolar continuum in detail

Examining the rating of the bipolar quantifiers in detail confirmed a high degree of symmetry about the centre point of the psychological continuum (discussed in the following paragraph) with a high correlation between the ratings on the 'Happy' and 'Sad' halves of the line, Spearman's rank correlation test[1] (r_s=0.998, p<0.001) (Sprent, 1993). The grouping of quantifiers and the symmetry of the bipolar quantifier scale were further

[1] Spearman's rank correlation test is a non-parametric correlation equivalent to Pearson's correlation but based on the ranks of the values of the data pairs rather than the actual data values. It is particularly useful for data with non-normal distributions and scales not at interval levels.

investigated by performing agglomerative cluster analysis[1]. This method starts with each quantifier forming its own separate cluster (giving 20 clusters to start), the distance between individual clusters along the psychological continuum is then calculated and the two closest clusters are joined to form a single cluster. This process continues until all quantifiers are joined within a single cluster. The clustering 'height' (0 to 100) indicates the distance along the continuum between two clusters when they join to form a single cluster. (The quantifiers 'Not at all' and 'Not very' were not included due to their variability and to allow later comparison with the unipolar scale). This gave a dendrogram showing the clustering of the bipolar quantifiers on the psychological continuum (Figure 6.7).

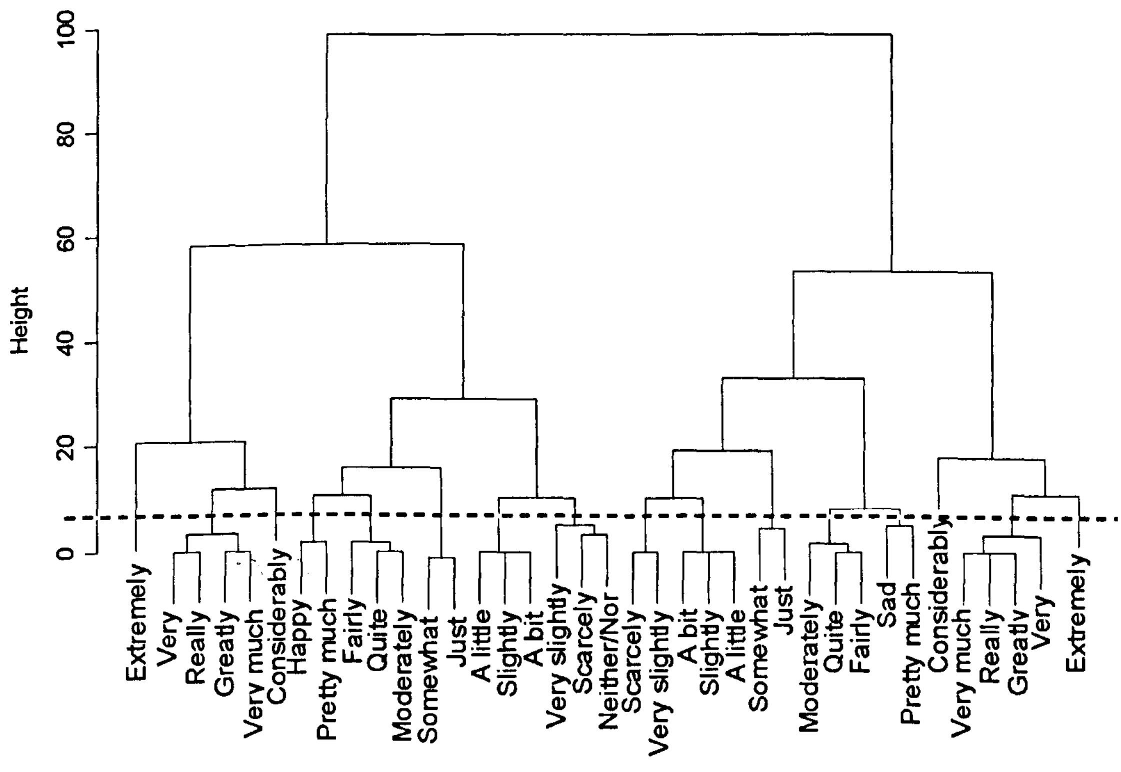


Figure 6.7 Grouping of bipolar quantifiers

[1] The agglomerative hierarchical cluster analysis used the Euclidean distance between object sizes as a dissimilarity metric. This method starts with each object size forming its own separate cluster, the distance between individual clusters is then calculated and the two closest clusters are joined to form a single cluster. This process continues until all object sizes are contained within a suitable number of clusters. Four clusters of object size were chosen as the analysis quickly settled at, and remained stable for, this number of clusters.

The symmetry of the bipolar quantifiers about the centre of the bipolar continuum was confirmed by the detail of the clustering on both the positive 'Happy' and negative 'Sad' halves of the dendrogram, where a high level of symmetry was apparent. This strongly suggested that a valid and stable positive-negative symmetrical bipolar scale could be derived from the data. The only exceptions were the reversals of 'Just' and 'Somewhat', 'Pretty much' and 'Sad' and 'Really' and 'Greatly' with respect to the order of the 'Happy', or positive side of the scale. However, looking at the distribution (Figure 6.7) in detail showed that these reversals were not highly significant. Here the quantifiers were within a small range of height on the clustering, were also within the same clusters, and in both cases the distributions of the ratings overlapped considerably.

There was a remarkable consistency and symmetry of detailed quantifier grouping. Taking a line at a similarity height of approximately 8 on the dendrogram, (Figure 6.7 dashed line) there were 16 (8 each side of the central anchor) definite, symmetrical groupings of similar-rated quantifiers. For example, quantifiers such as 'Somewhat' and 'Just' were closely rated and separately distanced from quantifiers such as 'Very', 'Greatly' 'Really' and 'Very much' on both the negative 'Sad' half and the positive 'Happy' side of the scale. This consistency strongly suggested that the subjective meanings of the quantifiers tended to be stable and equally distanced when applied to both negative and positive keywords, again showing that the data could be used to derive a valid symmetrical scale. In addition, this grouping suggested that quantifiers tended to be rated by subjects into a number of categories along the psychological continuum and that quantifiers within a given cluster could, to some extent, be used interchangeably if required. It was interesting to note that the scores for similar words on the 'sadness' half of the line were slightly further from the centre point of the line than equivalent scores on the happiness side. Since the ordering and groupings were highly consistent and separated on both sides of the scale, the selection of paired quantifiers at equal intervals should not be overly affected. It was not known why subjects rated the negative scale more highly. It was also notable that 'Considerably' had no similar rated quantifiers.

The importance of using clearly understandable and unambiguous quantifiers was shown by the high variability found in the 'Not at all' and 'Not very' quantifiers on the bipolar psychological continuum. The distributions of these quantifiers showed that 'Not at all happy' and 'Not very happy' were almost exclusively rated in the 'sad' half of the line and 'Not at all sad' and 'Not very sad' rated in the 'happy' half of the line. This showed that the majority of subjects reversed the meaning of the 'happy' or 'sad' keyword when prefixed by a 'not' quantifier. However, the wide nature of the distributions in each half of the continuum showed that subjects could not easily evaluate where to rate the quantifier after deciding which half of the continuum the quantifier belonged. These results suggested that care should be taken when selecting candidate quantifiers and that where possible confusing quantifiers or negation of quantifiers should be avoided.

6.9 The unipolar continuum in detail

The unipolar results suggested a strong similarity with the 'Happy' side of the bipolar psychological continuum and this was confirmed by a high correlation, Spearman's rank correlation test (r_s=0.980, p<0.001) (Sprent, 1993). Exceptions were the higher rating of 'A bit', the reversal of 'Just' and 'Somewhat', the disordering of 'Quite', 'Fairly', 'Pretty much' and 'Happy' and the disordering of 'Very much', 'Really', 'Greatly' and 'Very'. The higher ratings of 'Slightly', 'A little' and 'A bit' on the unipolar line were quite noticeable. It was notable that 'Not very happy' had moved from being rated quite 'Sad' on the bipolar line to being rated quite 'Happy' on the unipolar line, showing differences in rating dependent on scale type. Again performing an agglomerative cluster analysis on the unipolar data gave a dendrogram showing the clustering of the quantifiers for comparison with the positive side of the bipolar data scale (Figure 6.8).

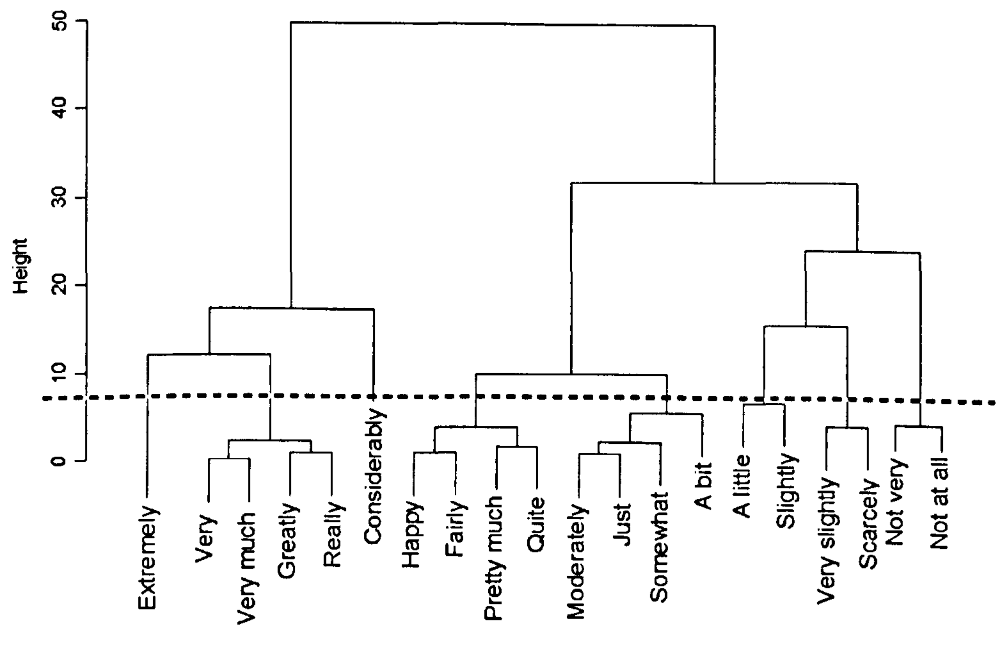


Figure 6.8 Grouping of unipolar quantifiers

Comparison of the unipolar dendrogram and the 'Happy' half of the bipolar dendrogram confirmed that there was an overall similarity between the two structures. As before, taking a line at a similarity height of approximately 8 on the dendrogram (Figure 6.8 dashed line) showed a total of 8 existing groups of similarly rated quantifiers on the unipolar dendrogram at this level of distance, the same number of groups as the bipolar dendrogram at the same similarity height (Figure 6.7 dashed line). However, looking at the detailed clustering of the dendrograms showed that the two scales were different.

The unipolar quantifiers 'Considerably' and 'Not at all' were clearly well distanced from their equivalents on the bipolar scale. Also there were pronounced differences between the quantifiers within some of these groups, confirming that the two scales were not the same. Notable differences between the ratings included the change of the quantifier grouped with 'Happy' and the movement of 'Moderately' and 'A bit' into different clusters on the unipolar scale.

These movements of the quantifiers and the dissimilarities between the results suggested that a unipolar scale is not equivalent to the positive half of a bipolar scale and that simple mirroring of a unipolar scale to create a bipolar scale or taking one half of a bipolar scale to create a unipolar scale was inadvisable. As with the bipolar ratings, the subjective meanings of the quantifiers on the unipolar scale tended to be rated into a number of categories along the psychological continuum. This again suggested that quantifiers within a cluster may to some extent, be used interchangeably.

When comparing the bipolar scale to the unipolar scale, there was a marked compression of the lower rated quantifiers towards the middle 'Neither/nor' anchor on the bipolar scale. This was probably due to an 'error of central tendency' (Oppenheim 1992), where subjects tend to mark toward the mid-point of a bipolar scale. This compression may have resulted in the 'Slightly', 'A little' and 'A bit' quantifiers being rated so closely that their true order is not apparent on the bipolar scale and only becomes clear when extended on the unipolar scale. It was notable again that 'Considerably' had no similar rated quantifiers.

6.10 Scale ranges

When calculating the intervals for a scale, the test end anchor point of 'Extremely' at the 'Happy' and 'Sad' ends of the bipolar continuum and the 'Happy' end of the unipolar continuum need not be used, instead alternative end anchor points may be selected such as 'Considerably' or 'Very'. The choice of anchor points affects the rating range of the scale such that there was a trade-off between the number of points on the scale and the range the scale covers (Hancock 1991). To cover a complete range the scale should include very high and very low rated anchor points, such as 'Extremely' and 'Not at all' at its ends, which entails the use of a larger number of points to avoid too coarse a scale and loss of fine resolution.

However, if the expected response range is known then a smaller section of the full scale may be used and the number of scale points correspondingly reduced. For instance, if a rated test will not induce responses above 'Moderately' then a section of the full scale starting with the anchor 'Not at all' and ending with the anchor 'Moderately' may be used. When calculating the optimal interval quantifiers given here, the end anchor points of 'Not at all' and 'Extremely' were chosen to give full-range scales. Full range scales were chosen for this work since the range of device subjective performance between the baseline hand mouse and the eye mouse was expected to be large based on previous experience (Bates 1999).

6.11 Generating optimal scales[1]

Typically scales found in human-computer interaction work have an odd number of intervals, probably to allow for a centre point if desired, and range from 5 to 11 intervals. Hence it was decided to generate bipolar and unipolar scales for odd-numbered intervals ranging from 5 to 11 intervals. This required quantifiers with suitable values to be chosen for set intervals along the psychological continua.

These interval quantifiers were simply generated from the detailed test data (Appendices Tables A6.1 and A6.2). This was accomplished by choosing the closest quantifiers to the desired intervals (5 interval, 7 interval and so on) for the scales. These intervals were calculated by dividing the range between the chosen anchor points by the number of intervals required. When a choice of candidate quantifiers was available then the quantifier with the minimum overlap with adjacent chosen quantifiers and the smallest distribution was selected. (An example of this process for a 5-point full-range bipolar scale is illustrated in Appendices, Figure A6.2).

[1] A note on the scales in this thesis:

Although questionnaire scales are ordinal scales, they are often treated as interval scales when used in attitudinal measurements. By treating this type of agreement scale or attitudinal measurement (as in this work) as interval, researchers can calculate mean scores which can then be compared, bearing in mind that the results originate from ordinal assessment. This is a very commonly used approach (For example: MacKenzie and Oniszczak 1998, Douglas and Kirkpatrick 1999, Smith 1996, Douglas et al. 1999, Brewster 1994, Hart et al, 1988), and has been adopted in this thesis.

Using these basic rules the optimal interval quantifiers for odd-numbered scales ranging from 5 to 11 points were calculated (Tables 6.2 and 6.3). These tables show the chosen quantifiers followed by the percentage points difference between the desired scale position and the actual quantifier rating noted in brackets after each quantifier. The percentage points separation between quantifiers is noted between adjacent selected quantifiers. These two metrics indicate the deviation of the closest quantifiers from their ideal positions on the psychological continua.

The reader may use the raw data (Appendix Tables A6.5 and A6.6) to calculate scales with different end points and different numbers of intervals by using the method described (Appendix Figure A6.7).

Optimal Bipolar Scales			
11 interval	9 interval	7 interval	5 interval
Extremely *Happy*	Extremely *Happy*	Extremely *Happy*	Extremely *Happy*
1.8%	7.0%	7.0%	12.3%
Really (+0.4%)	Considerably (-2.4%)	Considerably (+1.8%)	*Happy* (+0.4%)
0.0%	-3.5%	12.3%	21.5%
Happy (-4.4%)	*Happy* (+0.4%)	Somewhat (-2.7%)	Neither/Nor
-1.8%	1.8%	7.0%	26.3%
Moderately (-0.7%)	Somewhat (1.5%)	Neither/Nor	*Sad* (-4.9%)
1.8%	7.0%	11.0%	14.0%
Slightly (-3.0%)	Neither/Nor	Somewhat (+0.9%)	Extremely *Sad*
-8.6%	10.6%	12.3%	
Neither/Nor	Somewhat (-3.3%)	Considerably (-5.3%)	
8.8%	5.3%	5.3%	
Slightly (-0.5%)	*Sad* (-4.9%)	Extremely *Sad*	
5.3%	-1.8%		
Moderately (-6.3%)	Considerably (-1.1%)		
3.6%	5.3%		
Sad (+0.2%)	Extremely *Sad*		
5.3%			
Really (-5.7%)			
0.0%			
Extremely *Sad*			

Table 6.2 Optimal bipolar quantifiers

Optimal Unipolar Scales			
11 interval	9 interval	7 interval	5 interval
Extremely *Happy* 1.7% Very much (+0.5%) -12.0% Greatly (+8.2%) -10.3% Considerably (+5.9%) -3.4% Quite (-1.3%) -17.2% Moderately (0.0%) -6.9% A little (+3.8%) -34.5% Slightly (-0.7%) -31.3% Very slightly (-1.0%) -13.8% Not very (-3.1%) 1.8% Not at all *Happy*	Extremely *Happy* 0.0% Really (+1.2%) -6.9% Considerably (+0.8%) -3.4% Quite (-3.9%) -17.2% Moderately (0.0%) -6.9% A little (-1.3%) -34.5% Slightly (+4.3%) -17.3% Scarcely (+2.2%) 5.2% Not at all *Happy*	Extremely *Happy* 0.0% Really (+5.5%) -6.9% Considerably (+9.2%) 3.5% Moderately (0.0%) -6.9% A little (+2.9%) -12.1% Scarcely (-2.0%) 5.2% Not at all *Happy*	Extremely *Happy* 8.3% Considerably (+0.8%) 3.5% Moderately (0.0%) -10.3% Slightly (+4.3%) 13.8% Not at all *Happy*

Table 6.3 Optimal unipolar quantifiers

6.12 A summary of assessment scales

It was shown that there was a lack of any suitable questionnaire scale for the assessment method constructed previously. The design of scales was discussed and an experimentally derived set of bipolar and unipolar assessment quantifiers and a new set of bipolar and unipolar assessment scales determined. A method of generating interval scales was demonstrated and together with the results of the experiments in this chapter, tables were presented that allow the generation of diverse ranges of scales. Finally, the generation of a new set of bipolar and unipolar scales was shown.

Chapter 7

Choice of Scales

This chapter follows on from Chapter 6 and details the next step in constructing a questionnaire scale most suitable scale for hand, head and eye mouse assessment. As outlined in Chapter 3, the assessment method must be both sensitive to assess small variations in device assessment, but also have sufficient range to measure a wide range of possible device performances. This chapter first selects a range of candidate scales from the scales determined in Chapter 6, and then assesses each for their range of possible subjective responses, and their sensitivity to the smallest reliably detectable change in subjective response. Two experiments are conducted, the first to determine the smallest discrimination that can be reliably generated on a scale, and the second to use this discrimination level to test the sensitivity, and range, of a set of scales derived from Chapter 6. From this assessment, a single scale is selected that is both sensitive and wide-ranging for the subjective metric of the assessment method.

7.1 Candidate scales

The first task was to determine the labelling scheme (fully labelled or partially labelled) for the scale and the number of intervals (5, 7, 9 etc) the scale should have. When choosing the number of intervals there was a known trade-off between using too few, resulting in a loss of fine resolution due to coarseness of grouping, and too many, thus exceeding the rater's ability to discriminate between the intervals (Hancock 1991, Symonds 1924) although no optimum number of intervals was found. The choice of using a fully labelled or a partially labelled scale has been investigated previously (Frisbie 1979) but again no clear preference was found. Hence it was necessary to construct a range of scales in permutations of labelling schemes and interval ranges and to then test these scales to find an optimum scheme.

Choosing typical odd numbered interval ranges of 5 to 11 (to allow a central anchor point), and both fully and partially labelled scales, this gave a total permutation of 5, 7, 9 and 11 interval scales both fully labelled (all points) and partially labelled (end anchor points, and where applicable the centre anchor point) in bipolar and unipolar configurations. In addition to these scales, a 20 interval partially labelled unipolar scale was also included to examine the validity of the popular NASA-TLX workload questionnaire scale (Hart 1988) which can be used in the MUSiC method (Bevan et al. 1991 and 1995, MacLeod et al.1997).

7.2 A target acquisition discrimination test

Having decided upon the candidate scales, the next requirement was to devise a test that would allow subjects to generate a set of consistent subjective responses that could be used to test the *range,* and sensitivity or *discrimination,* of the candidate scales. To allow the evaluation of the candidate questionnaire scales to be as close as possible to their application in the assessment method, a questionnaire discrimination test needed to generate similar subjective responses to those that might be found during pointing device interaction with an interface. Simply conducting a period of pointing device interaction on an interface would not give the consistency between sessions as variation may occur during the interaction, instead a simple target acquisition test was devised that was similar to Fitt's Law target acquisition tests (thoroughly discussed by MacKenzie 1992). This would give consistency and would also be easy to perform.

To invoke a range of different subjective reactions from the test subjects, and so test the sensitivity of the scales, the 'difficulty' of the test needed to be varied. Typically this would be done by varying the target distance or target size during the test, with smaller targets or larger distances to be travelled being more 'difficult'. However, this variation in test difficulty would be visible, with differing target sizes and distances, and therefore could affect the subjective reaction of the subject in no relation to the actual subjective reaction experienced when performing the test. Instead, 'difficulty' was created by adding uncertainty to the cursor position by applying a level of 'jitter' to the cursor position that was visible as cursor displacement. This allowed the test to have a single fixed target size and distance, giving no visual clue as to the difficulty of the task, with small variation in cursor jitter giving variation in task difficulty. This method of added jitter, or positional uncertainty, of the cursor was also chosen in consideration of the assistive technology devices to be tested with the assessment method. Typically, head and eye mice exhibit positional inaccuracies of cursor position due to measurement inaccuracies and poorer bodily stability of head and eye position when compared to say hand position. This leads to a jitter in cursor position when using these devices. Hence mimicking the nature of these devices in a controlled way was an ideal method of selecting the questionnaire scales that would be used to assess these devices.

A Windows application was written in C++ for the target acquisition test (Figure 7.1). The application generated a circular target with a radius of 10mm at a random position on the screen. When the target was selected it was erased from the screen and a new target displayed at a random angle but a fixed distance of 80mm from the previous target position. This created a 'chase the target' form of target acquisition test (as discussed in Chapter 3) with a constant Fitts' Index of Difficulty (Fitts 1954) of 3.2 based on the

Shannon formulation (MacKenzie 1992). An aggregate score of the number of hits and misses was displayed to give subjects incentive to perform the test accurately. In addition, the application recorded the time taken to select each target during the test and the score for each session for later use (Chapter 8, 'Verification of the test method'). Finally, a Windows application was written in C++ to generate cursor jitter by displacing the current cursor by a random value in both the screen *x* and *y* directions at a fixed time interval. (This process is shown in Appendices, Figure A7.2).

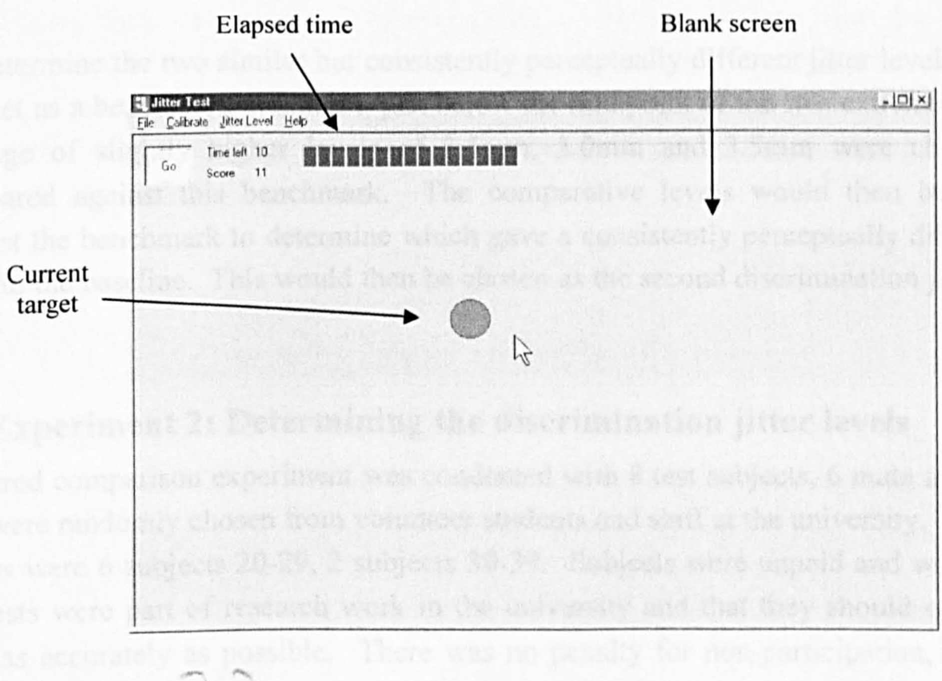


Figure 7.1 Jitter target acquisition test

7.3 Selecting jitter levels

Suitable jitter levels were required before the target acquisition test could be conducted. Both the *range* and *discrimination* of the scales needed to be tested. To do this a minimum of 4 levels of difficulty were required: two levels that were at extremes were needed to test the range of the scales, from a minimum to a maximum, and two that were very similar but consistently perceptually different were required to test the discrimination of the scales.

The extreme minimum level was logically set at zero jitter. To determine the upper extreme level of difficulty a range of cursor jitter levels were inserted into the jitter model (Appendices Figure 7.1) and informally evaluated by applying jitter to a standard desktop hand mouse when using the target acquisition test. This gave a 'feel' for the task difficulty for a range of jitter levels. From this it was felt that a displacement every 50ms gave a smooth rather than discrete and jerky cursor displacement period and the maximum jitter level that still allowed some control over cursor position at this timer interval was 6.0mm per displacement. Hence the extreme maximum jitter level was set at 6.0mm.

To determine the two similar but consistently perceptually different jitter levels, one level was set as a benchmark at 2.0mm, just below the mid point of the two extreme levels, and a range of slightly higher levels of 2.5mm, 3.0mm and 3.5mm were chosen to be compared against this benchmark. The comparative levels would then be compared against the benchmark to determine which gave a consistently perceptually different jitter level to the baseline. This would then be chosen as the second discrimination jitter level.

7.4 Experiment 2: Determining the discrimination jitter levels

A paired comparison experiment was conducted with 8 test subjects, 6 male and 2 female that were randomly chosen from volunteer students and staff at the university. Subject age ranges were 6 subjects 20-29, 2 subjects 30-39. Subjects were unpaid and were told that the tests were part of research work in the university and that they should complete the tests as accurately as possible. There was no penalty for non-participation, all subjects signed a consent form (Appendix Figure A7.2) and all data was anonymous.

For the experiment subjects were asked to perform the target acquisition test for 60 seconds, this was divided into two 30-second contiguous sequences, A and B. The sessions were identified by target colour, with a red target for one session and a blue target for the other session. The assignment of colours to sessions was randomised to eliminate learning effects or associations of task difficulty with target colour. In every test either session A or session B was assigned a jitter level of the baseline 2.0mm, with the remaining session having either the baseline jitter level of 2.0mm or one of the comparison jitter levels of 2.5mm, 3.0mm or 3.5mm (Appendices Table A7.1). To eliminate order effects in the testing, the presentation order of the jitter levels and session orders were prescribed in an incomplete Latin Square design (Appendices Table A7.2).

After each session the subjects were asked to state whether the red or blue target was more difficult to select or whether they were equally difficult to select. The object of the experiment was to determine at what magnitude of difference in jitter levels subjects could reliably discriminate between the baseline and higher level of jitter. An arbitrary discrimination accuracy of 90% was chosen as the threshold where a higher level was reliably discriminated from the baseline lower 2.0mm level. The results of the experiment were calculated as percentages and the prescribed order of session target colours removed so that the baseline was hence referred to as 'Red' and the variable comparators as 'Blue' to allow easier understanding of the results (Appendices Table A7.3).

The results were displayed graphically (Figure 7.2) and showed that the majority of subjects always discriminated correctly between the jitter levels, however the 90% reliability threshold was not reached until the comparator jitter level reached 3.5mm. Hence, the jitter levels of 2.0mm and 3.5mm were chosen as the closest possible levels that could be reliably discriminated.

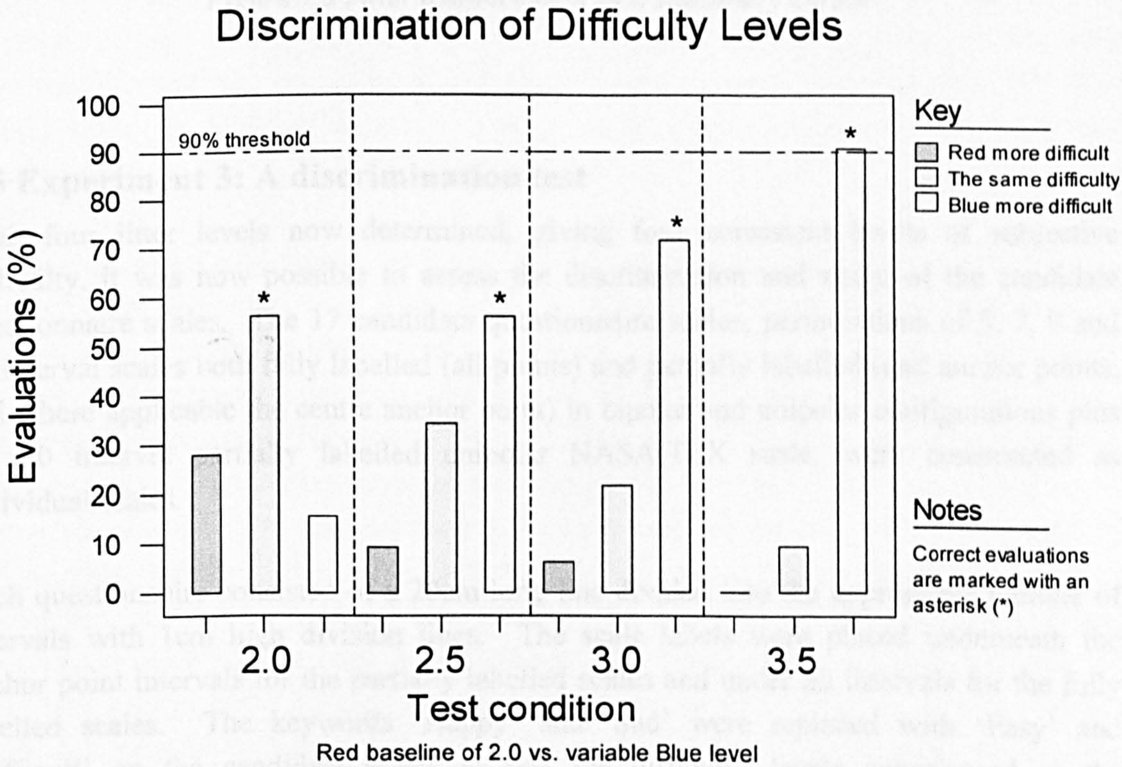


Figure 7.2 Discrimination of similar jitter levels

To illustrate the effects of applying the 4 levels of jitter to a stationary cursor, plots were recorded over 2 second periods of cursor movement (Figure 7.3). These plots clearly illustrate the large difference between the two levels at the extremities and also the similarity between the two middle levels.

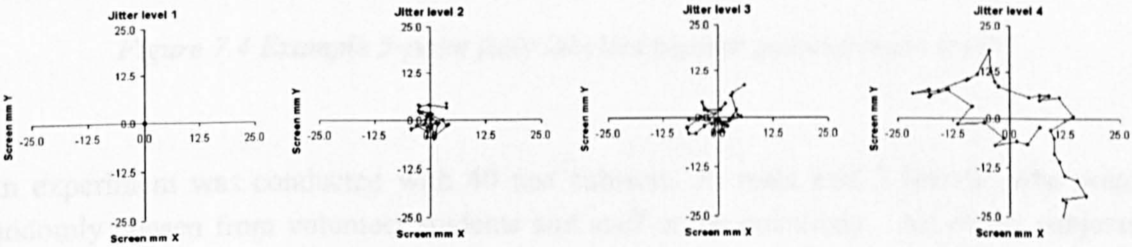


Figure 7.3 Jitter displacement of a stationary cursor

7.5 Experiment 3: A discrimination test

With four jitter levels now determined, giving four consistent levels of subjective difficulty, it was now possible to assess the discrimination and range of the candidate questionnaire scales. The 17 candidate questionnaire scales, permutations of 5, 7, 9 and 11 interval scales both fully labelled (all points) and partially labelled (end anchor points, and where applicable the centre anchor point) in bipolar and unipolar configurations plus the 20 interval partially labelled unipolar NASA-TLX scale, were constructed as individual scales.

Each questionnaire consisted of a 20cm long line divided into the appropriate number of intervals with 1cm high division lines. The scale labels were placed underneath the anchor point intervals for the partially labelled scales and under all intervals for the fully labelled scales. The keywords 'Happy' and 'Sad' were replaced with 'Easy' and 'Difficult' on the candidate scales to rate the difficulty levels experienced in the discrimination test. Each scale was printed separately on a blank sheet of A4 paper. A sample fully labelled 5-interval bipolar questionnaire scale is shown (Figure 7.4 not to scale).

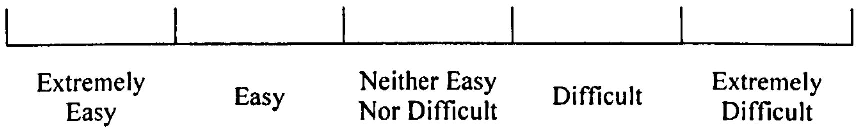


Figure 7.4 Example 5-point fully labelled bipolar questionnaire scale

An experiment was conducted with 40 test subjects, 35 male and 5 female, who were randomly chosen from volunteer students and staff at the university. All of the subjects spoke English as their first language. Subject age ranges were 1 subject < 20, 16 subjects 20-29, 16 subjects 30-39, 6 subjects 40-49, 1 subject 50+. Subjects were unpaid and were told that the tests were part of research work in the university and that they should complete the tests as accurately as possible. All subjects were regular computer users and were familiar with the operation of a desktop mouse. There was no penalty for non-participation, all subjects signed a consent form (Figure A7.3) and all data was anonymous.

For the experiment each subject performed the 30-second target acquisition test at a single selected jitter difficulty level (0.0, 2.0, 3.5, 6.0mm) followed by presentation of a questionnaire scale to the test subject (Appendices Tables A7.4 and A7.7). The subject was then asked to rate on the scale how easy or difficult they found the test. The subject then briefly rested before continuing with the next test until all permutations of difficulty levels and questionnaires were completed. To eliminate order or learning effects in the testing, all previous answers were hidden from the subjects and the presentation order of the candidate questionnaire scales and the levels of test difficulty were prescribed in incomplete Latin Square designs (Appendices Tables A7.5 and A7.8). In order to compare the different scale ratings from the experiment, the rating results were all converted to equivalent ratings based on the largest scale and placed on a 20-interval psychological continuum, with the interquartile ranges of the results used to assess the distribution of the ratings along the continuum. For example, a rating of 5 on a 7-interval scale was converted to a rating of (5 / 7) * 20 = 14.3 on the 20 interval scale (The results are shown in Appendices Tables A7.6 and A7.9).

7.6 Interpreting the discrimination test results

An idealised questionnaire would rate level 1 (0.0mm) and level 4 (6.0mm) of the discrimination test as narrow distributions on the far left 'Easy' and far right 'Difficult' parts of the continuum and well separated (Gaps A and B, Figure 7.5) from the two more central difficulty levels, with difficulty levels 2 (2.0mm) and 3 (3.5mm) narrow distributions not overlapping (Gap 'C' Figure 7.5) in the middle of the continuum.

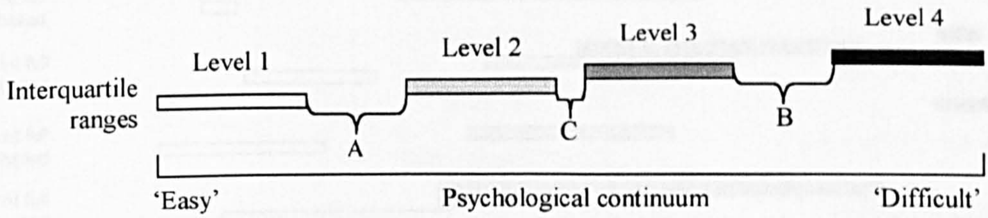


Figure 7.5 Idealised questionnaire scale ratings

7.7 Bipolar results

The interquartile ranges from the discrimination test results for the permutations of the bipolar scale were calculated (Appendices Table A7.6) and displayed graphically (Figure 7.5). This graph shows from the distributions that all of the scales correctly placed difficulty levels 1 and 4 at the correct ends of the psychological continuum and that all of the scales did discriminate between these two extreme levels.

Examining the distributions for difficulty levels 2 and 3 showed that the 5 and 9 interval part and fully labelled scales and the 11-interval fully labelled scale had poor discrimination between these levels (Appendices Table A7.6). This left the two 7-interval scales and the 11 interval part-labelled scale for further analysis. Mann-Whitney two-sample rank test[1] (Sprent 1993) were used to further examine these scales to determine if

[1] The Mann-Whitney two-sample rank test is a non-parametric equivalent to a two-sample t-test based on the ranks of the data. It is particularly useful for data with non-normal distributions where data sets have different sample numbers and the data do not have something in common – in this case they come from different test domains.

difficulty level 3 was significantly greater than level 2. The 7-interval fully labelled scale showed the largest significant difference between the levels (U=390, N=40, $p<0.0001$), followed by the 7-interval part-labelled scale (U=430, N=40, $p=0.0001$) and the 11-interval part-labelled scale (U=568, N=40, $p=0.0115$). This showed that the 7-interval fully labelled scale exhibited the highest level of discrimination.

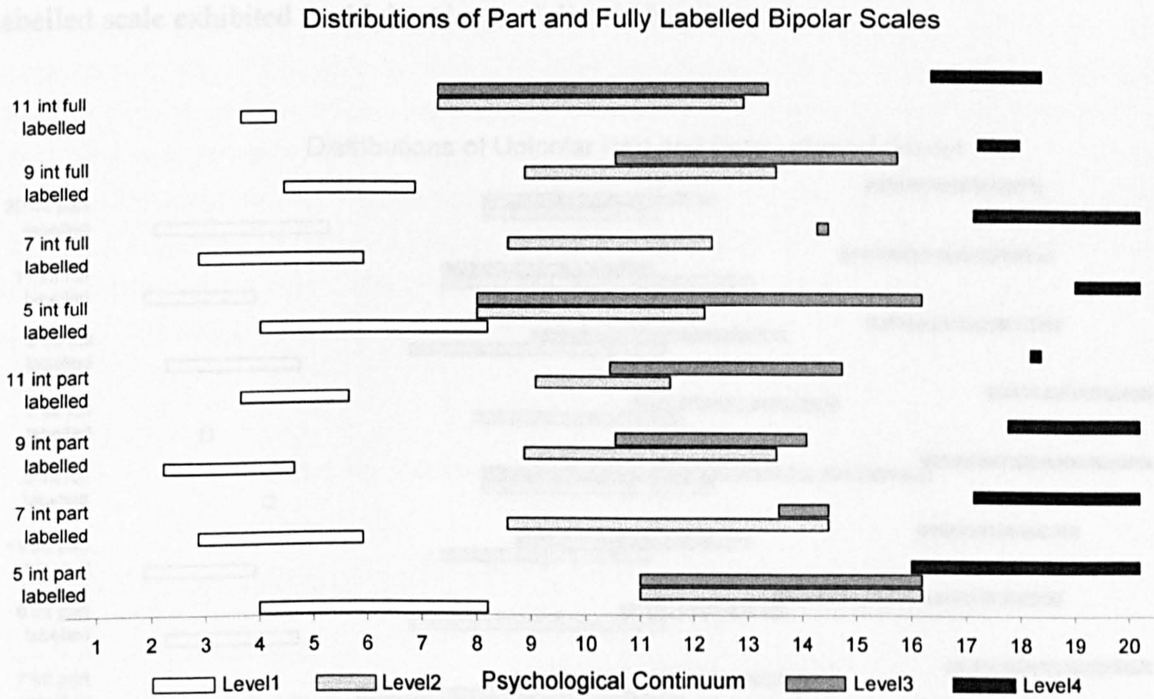


Figure 7.6 Distributions of bipolar scales

7.8 Unipolar results

The interquartile range discrimination test results for the permutations of the unipolar scale were calculated (Appendices Table A7.9) and displayed graphically (Figure 7.7). From this graph it is again clear from the distributions that all of the scales correctly placed difficulty levels 1 and 4 at the correct ends of the psychological continuum and that all of the scales did discriminate between these two extreme levels (Appendices Table A7.9). Examining the distributions for difficulty levels 2 and 3 on the unipolar scale showed that only the 7 and 9 interval fully labelled and 9-interval part labelled scale

Also see Appendices Notes Discussion 1. "Non-parametric tests in this work".

showed good discrimination between these levels. The 9 interval part-labelled scale must be discounted due to the overlap of levels 3 and 4. It was notable that the 20 interval part-labelled NASA-TLX scale exhibited poor discrimination. Mann-Whitney two-sample rank tests were again used to further examine these scales to determine if difficulty level 3 was significantly greater than level 2. Once again the 7-interval fully labelled scale showed the highest significance (U=402, N=40, $p<0.0001$), followed by the 9-interval fully labelled scale (U=590, N=40, $p=0.0207$). This showed that the 7-interval fully labelled scale exhibited the highest level of discrimination.

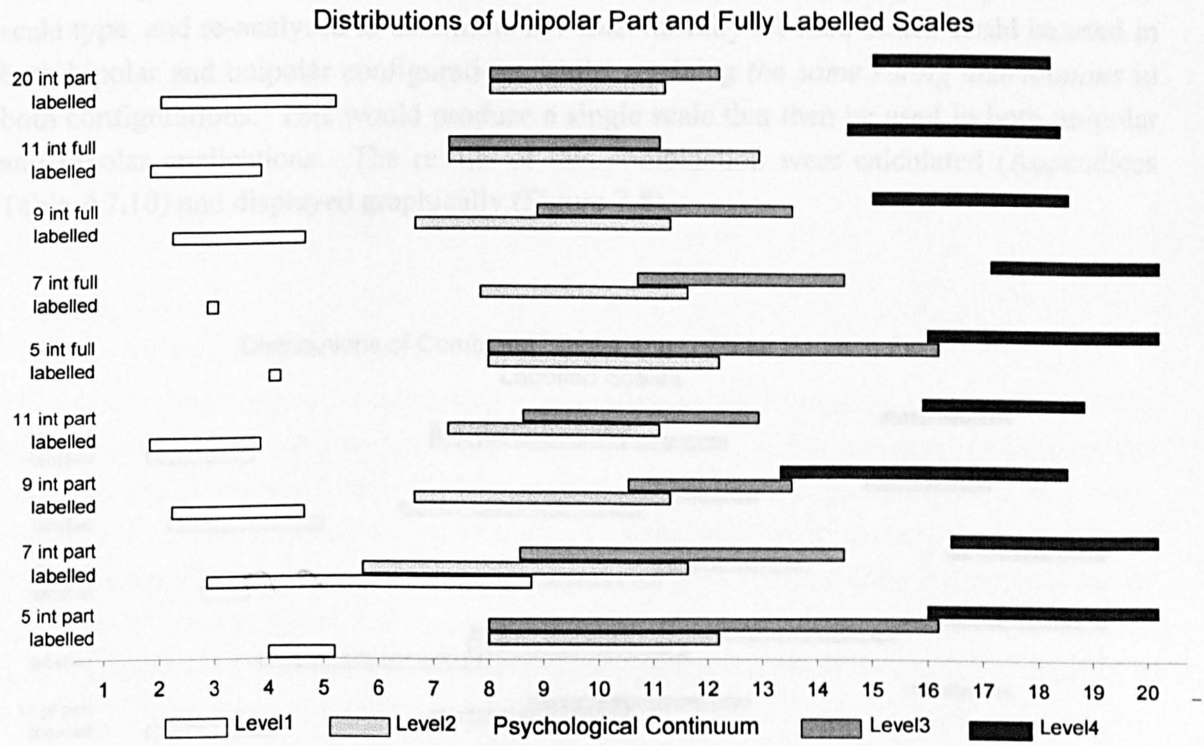


Figure 7.7 Distributions of unipolar scales

7.9 A single combined scale

Usability and workload questions and keywords may be either unilateral or bilateral in nature, lending themselves to unipolar or bipolar scales respectively. For example, the bilateral question 'How fast was the pointing device?' would suggest a bipolar scale with a range from 'Extremely slow' through 'Neither fast nor slow' to 'Extremely fast'. In contrast the unilateral question 'Do you feel tired?' suggests a unipolar scale with the range 'Not at all tired' to 'Extremely tired'. Attempting to change the scale types for these examples would result in 'Not at all fast' to 'Extremely fast' and 'Extremely not tired' through 'Neither tired nor not tired' to 'Extremely tired'- producing rather clumsy and confusing scales. Since both bilateral and unilateral types of questions may occur on a single questionnaire, it would be desirable to be able to use a single scale type with the same number of intervals and labelling type for both questionnaire scale designs.

The results of the previous tests indicated that the 7-interval fully labelled scales outperformed all other scales in both bipolar and unipolar configurations. This suggested that the bipolar and unipolar results could be combined, by simply pooling the data by scale type, and re-analysed to determine if 7-interval fully labelled scales could be used in both bipolar and unipolar configurations whilst retaining *the same rating distributions* in both configurations. This would produce a single scale that then be used in both unipolar and bipolar applications. The results of this combination were calculated (Appendices Table A7.10) and displayed graphically (Figure 7.8).

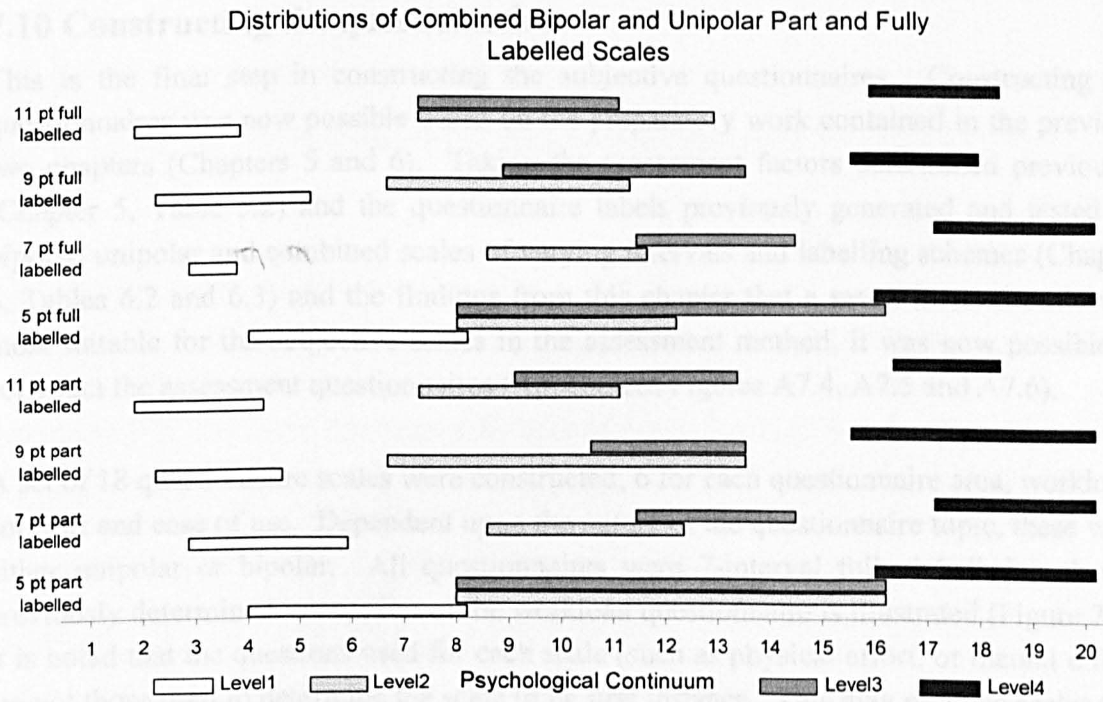


Figure 7.8 Distributions of combined bipolar and unipolar scales

As before, the distributions (Figure 7.8) showed that all of the combined scales correctly placed difficulty levels 1 and 4 at the correct ends of the psychological continuum and that all of the scales could discriminate between these two extreme levels. Examining the distributions for difficulty levels 2 and 3 on the combined scales showed that only the two 7 interval scales showed good discrimination between these levels, with the 9 interval fully labelled scale and the 11 interval part-labelled scale showing some discrimination. Again, Mann-Whitney two-sample rank tests were used to further examine these scales to determine if difficulty level 3 was significantly greater than level 2 for these combined results. Once again the 7-interval fully labelled scale showed the highest significance (U=430, N=40, p=0.0001), followed by the 7 interval part-labelled scale (U=440, N=40, p=0.0002), the 9 interval fully labelled scale (U=466, N=40, p=0.0005) and the 11 interval part-labelled scale (U=581, N=40, p=0.0161). This confirmed that the 7-interval fully labelled scale exhibited the highest level of discrimination with the combined distributions of the bipolar and unipolar scales, strongly suggesting its suitability for both bipolar and unipolar use within the same questionnaire.

7.10 Constructing the questionnaires

This is the final step in constructing the subjective questionnaires. Constructing the questionnaires was now possible based on the preparatory work contained in the previous two chapters (Chapters 5 and 6). Taking the assessment factors determined previously (Chapter 5, Table 5.2) and the questionnaire labels previously generated and tested on bipolar, unipolar and combined scales of varying intervals and labelling schemes (Chapter 6, Tables 6.2 and 6.3) and the findings from this chapter that a seven interval scale was most suitable for the subjective scales in the assessment method, it was now possible to construct the assessment questionnaires (Appendices Figures A7.4, A7.5 and A7.6).

A set of 18 questionnaire scales were constructed; 6 for each questionnaire area, workload, comfort and ease of use. Dependent upon the nature of the questionnaire topic, these were either unipolar or bipolar. All questionnaires were 7-interval fully labelled scales as previously determined. A section of the workload questionnaire is illustrated (Figure 7.9). It is noted that the questions used for each scale (such as physical effort, or mental effort) are not those used to determine the scale in he first instance. This may produce scales that are not as balanced as those derived previously (using Happy and Sad), however, this is regarded as an acceptable compromise (MacKenzie and Oniszczak 1998, Douglas and Kirkpatrick 1999) for example.

Workload Assessment Questionnaire

Please circle the 'X' closest to your opinion

← low workload ratings high →

1. How much *physical* effort or activity was required to operate the system?

X	X	X	X	X	X	X	
Extremely low physical effort	Considerably low physical effort	Somewhat low physical effort	Neither high nor low physical effort	Somewhat high physical effort	Considerably high physical effort	Extremely high physical effort	

X

2. How much *mental* effort or concentration was required to operate the system?

X	X	X	X	X	X	X
Extremely low mental effort	Considerably low mental effort	Somewhat low mental effort	Neither high nor low mental effort	Somewhat high mental effort	Considerably high mental effort	Extremely high mental effort

3. How much.....

Figure 7.9 Example questionnaire

7.11 A summary of choice of scales

The discrimination of fully labelled and part-labelled 5, 7, 9 and 11 interval questionnaire scales together with the popular NASA-TLX 20 interval part-labelled scale were determined. It was found that within the bounds of this work, in both bipolar and unipolar configurations, all of these scale types could correctly position and discriminate between widely spaced levels of difficulty. However, when trying to discriminate between closely spaced levels of task difficulty, the majority of the scales performed poorly. Within this work scales with few intervals exhibited a coarseness of grouping and scales with many intervals may exceed the rater's ability to discriminate between intervals. It was notable that the 20-interval part-labelled NASA-TLX scale showed very poor discrimination and that the use of this scale for usability and workload questionnaires in the Human Computer Interaction field is questionable[1].

It was found that 7 interval fully labelled scales gave the highest discrimination in both bipolar and unipolar configurations and that usability and workload questions and keywords may be either unilateral or bilateral in nature. Finally, it was shown that the 7 interval fully labelled scale could be used in both bipolar and unipolar formats, with the same rating distributions in both configurations, and hence was the most suitable scale for unipolar and bipolar question types.

[1] This finding is of some concern, as the TLX scale has been used previously. It is possible that the original design of the TLX, for assessing pilot workload, is valid but only for situations of extreme workload. The application of the TLX to input device assessment is not valid, probably as the range of workload generated by human computer interaction is not sufficiently wide.

Chapter 8

Validation of the Assessment Method

This chapter validates the work contained in the previous chapters that developed the assessment method for head and eye based interaction. The chapter first defines a benchmark for comparison based on the 'standard' target acquisition test discussed in Chapter 3. It then shows how the performance of a standard hand mouse can be varied in a controlled way to produce a device with differing levels of known benchmark performance. The proposed assessment method constructed in the previous chapters is then used to assess the performance of this hand mouse when varied to different levels of known benchmark performance. The results of these trials on the method are then compared to the known performance variation of the benchmark hand mouse, with the aim of validating the assessment method against the known benchmark.

8.1 A benchmark for comparison

In order to ensure that the assessment method would give valid and consistent results, it was necessary to validate the method against a known standard before it could be used. As discussed in Chapter 3, the most commonly used and consistent and proven standard for pointing device assessment is the Fitts Law (Fitts 1954) target acquisition test, giving an *Index of Performance* (IP) for a device on any given test task (For example, MacKenzie 1992, MacKenzie 1991; MacKenzie and Buxton 1992, Accot and Zhai 1997, Sibert and Jacob 2000, Douglas and Kirkpatrick 1999, Murata 1991, Istance 1993, Bates 1999).

To briefly review this metric, the Index of Performance is calculated based on the difficulty of the task, or *Index of Difficulty* (ID) (where Index of Difficulty is calculated from a combination of the size of the target and the cursor distance to be moved to the target) and the time taken to select the target, or *Movement Time* (MT) (where Movement Time is the movement time of the cursor) (MacKenzie 1991) (Equation 8.1).

As the 'performance' of a device on a target acquisition task increases (the time taken to complete the task reduces) so the measured Index of Performance for that device will increase in a linear fashion. This gave a simple performance baseline. Hence one method of validating the proposed assessment method would be to test a series of devices of known IP on the method and examine the test results from the method in comparison to the known Index of Performance. For the assessment method to be valid the results for the devices should preserve the order or *ranking* of the devices based on their known

Index of Performance. In addition, the *range* of the method (how well can it measure both very low and very high performance) and the *sensitivity* (how well can it discriminate between very similar performances) should be similar to the results for known device Indices of Performance. Some variation in the relationship between the Fitts benchmark and the results of the method would be expected as they are different measuring techniques, but the rankings, ranges and sensitivities should be reasonably comparable to validate the method.

$ID = \log_2(A / W + 1)$ *Index of Difficulty (ID) of the task (dimensionless)*

$MT = a + b\,ID$ *Movement Time (MT) for the task (seconds)*

$IP = ID / MT$ *Index of Performance (IP) in Bits (of information generated) / second*

Equation 8.1 Calculation of Index of Performance[1,2]

8.2 The jitter test revisited

During the questionnaire jitter test experiment (Chapter 7) the individual movement times for each target acquisition were recorded. Since the target acquisition test (Chapter 7) had a fixed Index of Difficulty these times were used to calculate the Index of Performance of the standard desktop hand mouse with the range of jitter levels applied. In addition, the four jitter levels used, two to test a wide range of difficulty and two closely spaced to test sensitivity or discrimination generated pointing devices, would be expected to produce both widely spaced low and high Indices of Performance and closely spaced similar Indices of Performance.

The data for the four jitter levels for all test subjects in the jitter test experiment were analysed and the Indices of Performance of the four jitter levels calculated (summary Table 8.1 and Appendices Table A8.1) and displayed graphically (Figure 8.1).

[1] From MacKenzie 1991.
[2] See Chapter 3 for definitions

Jitter level and Index of Performance	
Jitter level (mm)	*IP*
0.0	3.339
2.0	2.433
3.5	1.858
6.0	0.080

Table 8.1 Jitter level and Index of Performance

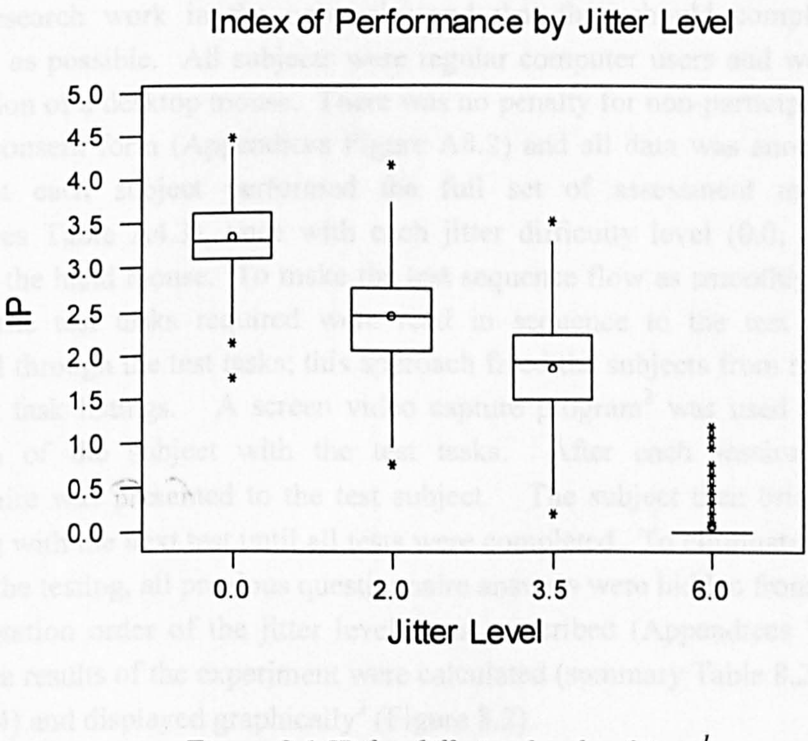


Figure 8.1 IP for differing levels of jitter[1]

The results showed a logical *ranking* of Index of Performance for the jitter levels and supported the premise that the two widely spaced jitter levels would produce low and high Indices of Performance and the two closely spaced jitter levels would produce closely spaced similar but statistically different Indices of Performance (Wilcoxon two-sample

[1] See Appendix Notes Figure 1 for an explanation of the graph symbols.

matched pairs signed rank test tests[1] (Sprent 1993) between levels 2.0 and 3.5, W=72318, N=394, $p<0.001$). The use of jitter to modify the Index of Performance of a standard desktop mouse in a consistent way was further supported by a strong regression relationship (IP = 3.47384 - 0.537788 Jitter, $R^2 = 0.85$) between raw Index of Performance data and jitter level (Appendices Figure A8.1).

8.3 Experiment 4: Testing the method with jitter

An experiment was conducted with 6 test subjects, 4 male and 2 female, which were chosen from volunteer students and staff at the university. Subject age ranges were 5 subjects 20-29, 1 subject 30-39. Subjects were unpaid and were told that the tests were part of research work in the university and that they should complete the tests as accurately as possible. All subjects were regular computer users and were familiar with the operation of a desktop mouse. There was no penalty for non-participation, all subjects signed a consent form (Appendices Figure A8.2) and all data was anonymous. For the experiment each subject performed the full set of assessment method test tasks (Appendices Table A4.3), once with each jitter difficulty level (0.0, 2.0, 3.5, 6.0mm) applied to the hand mouse. To make the test sequence flow as smoothly and naturally as possible, the test tasks required were read in sequence to the test subjects as they progressed through the test tasks; this approach freed the subjects from repeated reference to the test task listings. A screen video capture program[2] was used for recording the interaction of the subject with the test tasks. After each session the assessment questionnaire was presented to the test subject. The subject then briefly rested before continuing with the next test until all tests were completed. To eliminate order or learning effects in the testing, all previous questionnaire answers were hidden from the subjects and the presentation order of the jitter levels was prescribed (Appendices Tables A8.2 and A8.3). The results of the experiment were calculated (summary Table 8.2 and Appendices Table A8.4) and displayed graphically[3] (Figure 8.2).

The results confirmed a correct *ranking* of task efficiency for the jitter levels and validated the range of the method with two widely spaced jitter levels producing low and high task efficiencies. In addition, the two closely spaced jitter levels produced closely spaced similar but statistically different task efficiencies (Wilcoxon two-sample signed rank test

[1] The Wilcoxon matched pairs signed rank non-parametric test is used when distributions are not normal, sample sizes are equal and both data sets have commonality – in this case both sets of data originate from the same set of test tasks and each data sample from one device can be paired with a corresponding sample from the another device. See Appendices Notes Discussion 1. "Non-parametric tests in this work"

[2] Hypercam: www.hyperionics.com

[3] See Appendix Notes Figure 1 for an explanation of the graph symbols.

between levels 2.0 and 3.5, W=72318, N=394, $p<0.001$) validating the sensitivity and discrimination of the test method.

Jitter level and Task efficiency	
Jitter level (mm)	Task efficiency (%)
0.0	83.3
2.0	80.6
3.5	71.4
6.0	44.1

Table 8.2 Jitter level and Task efficiency

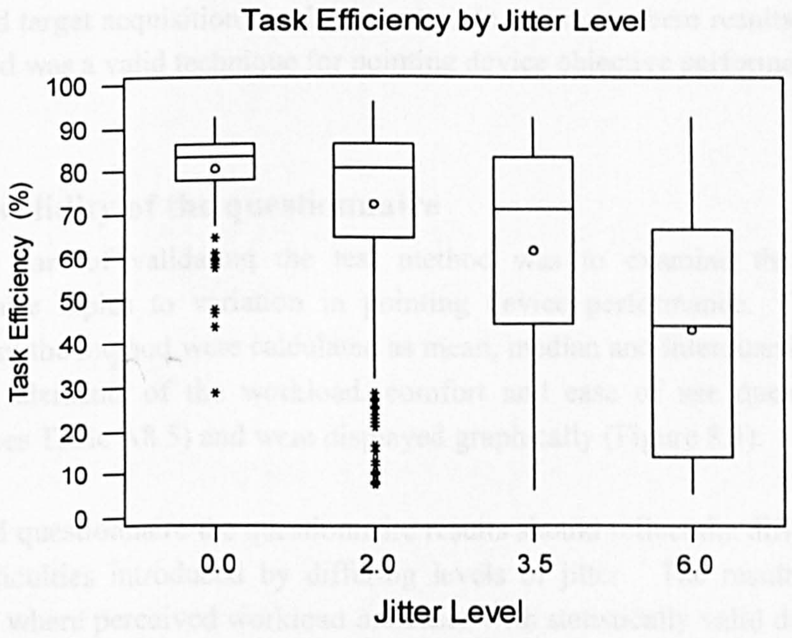


Figure 8.2 Task efficiency for differing levels of jitter

The validity of the test method to accurately reflect variation in standard desktop mouse pointing performance in a consistent way was further supported by a (somewhat weaker than the jitter target acquisition test, Chapter 8.1) regression relationship of (Task efficiency = 87.67 – 7.44 Jitter, $R^2 = 0.30$) between raw task efficiency data and jitter (Appendices Figure A8.2). This weaker regression fit ($R^2 = 0.30$) can be explained by the nature of the scale and is discussed later (Chapter 8.3).

8.4 Comparison of jitter test validation results

Although not identical, comparing the baseline jitter target acquisition test results (Table 8.1) with the results of the same jitter levels applied to the test method (Table 8.2) confirmed that the test method produced the same *ranking* of device performance to the baseline target acquisition test. The test method showed a weaker regression relationship between efficiency and jitter level than the target acquisition test, however this can somewhat be explained by the compression at the ends of the efficiency scale and the wide interquartile ranges of the test method data. These wide ranges are due to the far more complex nature of the test tasks; for a given jitter difficulty level the target acquisition test had simple tasks that would tend to produce little variation in performance; however the test method had complex tasks that would tend to produce a wider range of efficiency results. The test method does maintain performance *ranking* when compared to the established target acquisition baseline results. The method can accommodate a wide *range* of performance and also shows very similar *sensitivity* or discrimination to the established target acquisition baseline results. In summary these results indicated that the test method was a valid technique for pointing device objective performance assessment.

8.5 The validity of the questionnaire

The final part of validating the test method was to examine the response of the questionnaire topics to variation in pointing device performance. The questionnaire results from the method were calculated as mean, median and interquartile ranges from the individual elements of the workload, comfort and ease of use questionnaire sections (Appendices Table A8.5) and were displayed graphically (Figure 8.3).

For a valid questionnaire the questionnaire results should reflect the differing pointing and usage difficulties introduced by differing levels of jitter. The results support this for workload, where perceived workload increases with statistically valid differences for each increase in pointing difficulty (Wilcoxon two-sample signed rank test between levels). However, perceived comfort remains essentially unchanged for varying pointing difficulty with statistically the same result for all levels of difficulty. This result reduced the validity test of the questionnaire, but can be explained as the actual physical comfort of using the device is essentially unchanged by the application of jitter. Finally perceived ease of use correctly reduced with increasing difficulty, although with some statistically identical overlapping results between adjacent levels of difficulty, indicating a slight loss of discrimination in the questionnaire. However, considering the small sample size of 6 subjects and hence 6 responses to each questionnaire topic at each level of difficulty, some loss of discrimination was to be expected in the questionnaire results. Overall, the questionnaire results do support validation of the questionnaire, with correct ordering of workload, arguably correct equal and high results for comfort (with the hand mouse being a comfortable device to use) and a correct trend for ease of use.

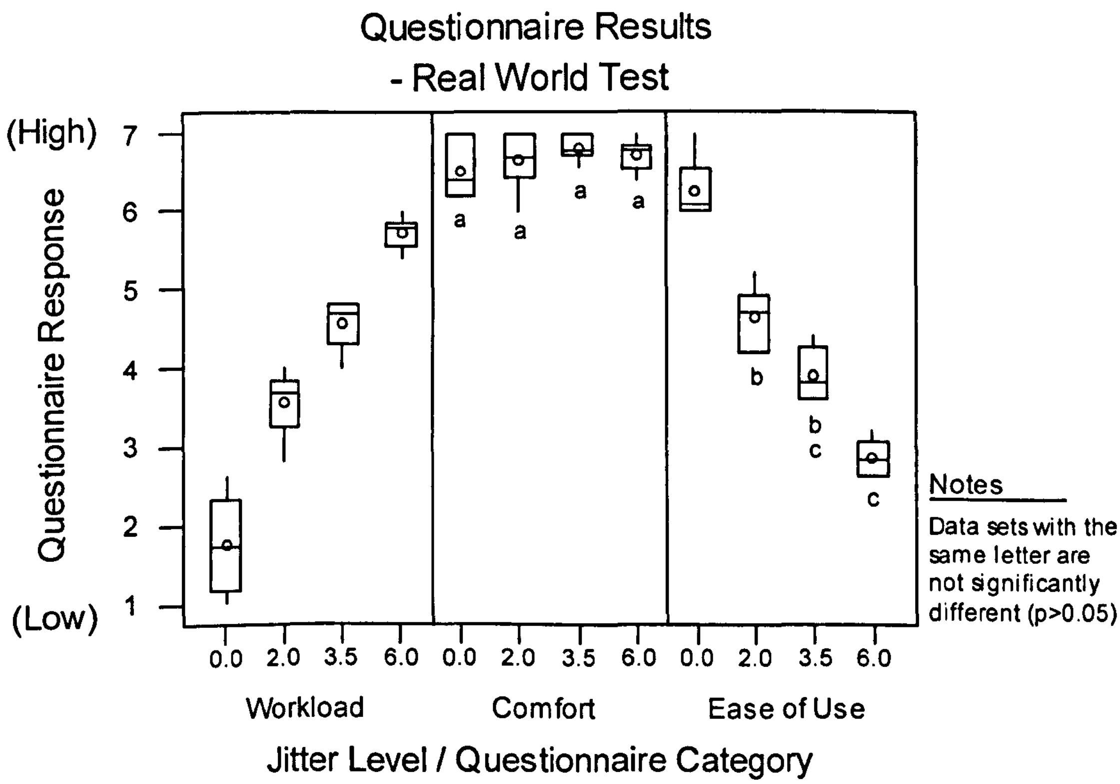


Figure 8.3 Questionnaire response for differing levels of jitter[1]

8.6 A summary on validation of the assessment method

The preceding chapters proposed a complete method for the assessment of assistive technology pointing devices. This chapter tested the validity of that method by comparison with a known baseline of pointing performance, and found the method to valid. This chapter concludes the construction and validation of a comprehensive assessment method based on real world test tasks, objective performance measurements and subjective satisfaction measurement that is suitable for the detailed assessment of hand, head and eye mouse pointing devices.

[1] See Appendix Notes Figure 1 for an explanation of the graph symbols and statistical markings.

Chapter 9

Creating Eye and Head Mice

This chapter first discusses the limitations of currently available eye and head mouse systems introduced in Chapter 2. It then shows the development of new eye and head mouse systems designed to give repeatable and accurate performance measurements, and also allow modification to their operation so that any performance enhancements proposed by this work could be applied to the devices. This chapter then shows the development or modification of associated object selection and text entry systems, suitable for supporting the assessment of head and eye mice performance. Finally, the chapter shows the development of a simple calibration test screen to assess the calibration and pointing accuracy of the head and eye mice created in this chapter.

9.1 Limitations of available head mice

As introduced in Chapter 2, there were a range of commercially available head mouse systems that could be used for this work. However, all of these systems showed some limitations either in pointing accuracy, latency or ease of set up (Chapter 2, Table 2.1), with differing systems showing differing operational characteristics and with no one system regarded as being better or more optimal than the others. These limitations will tend to reduce the performance of these devices in terms of calibration consistency, pointing accuracy, pointing speed and pointing lag. It was felt that, by assessing any of these commercial devices, these characteristics could influence, or skew, the results of the assessment depending on the device chosen.

This work was intended to be generic and not specific to the properties of any individual head mouse device, so for this reason the commercial devices surveyed in Chapter 2 were rejected due to their compromised performance. Instead, a simple to use but very accurate, consistent and responsive head mouse device was required based solely on the best practically available tracking technology. By constructing and using such a device this would produce results that would give the best possible performance for the assessment, with the least influence on the assessment results caused by any specific characteristic(s) from a given commercial device.

9.2 Creating a Head Mouse

To construct a highly accurate but simple to use head mouse the optimum approach was to track the head position of the user as accurately as possible in 3-dimensional space in front of the computer screen displaying the interface on which the assessment was to be conducted. Since the restrictions of cost and 'hands free' operation were lifted, the most accurate, consistent and responsive systems commonly used for tracking were electromagnetic 6-degree of freedom (*x, y, z, roll, pitch, yaw*) systems. These systems relied on a transmitter/receiver pair coupled electro-magnetically, with the orientation of the receiver known with a very high degree of accuracy in relation to the transmitter, typically with a resolution of 0.5mm linear and 0.1° angular, with a lag of <10ms, a sampling rate of >20Hz and simple calibration[1,2]. Two such systems are widely used, the Ascension 'Flock of Birds[1]' and the Polhemus 'Fastrack[2]' systems. With little to choose between the two, and with easy availability within the University, a Polhemus system was chosen.

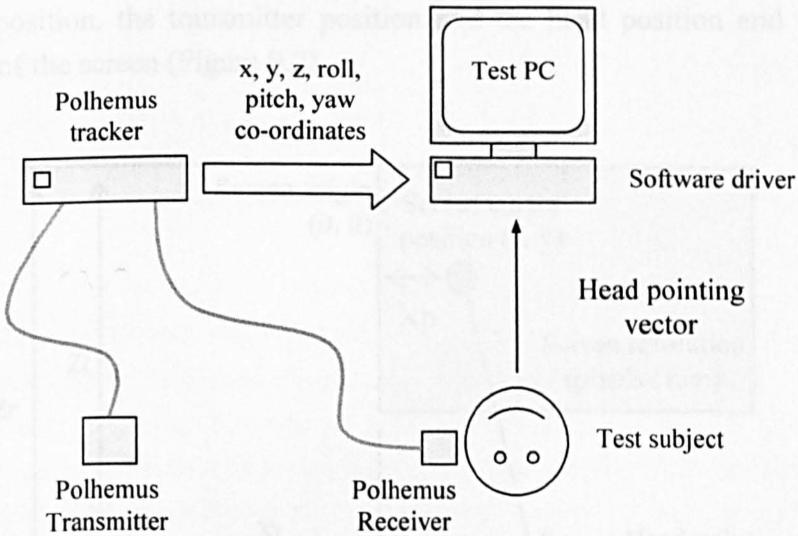


Figure 9.1 Head tracking equipment arrangement

The Polhemus head tracking equipment was set up in close proximity to the position of the test subject to maximise the tracking accuracy of the equipment (Figure 9.1). Close

[1] http://www.ascension-tech.com/
[2] http://www.polhemus.com/

proximity was required as the system was electromagnetic, with the power of the coupling field dropping dramatically with the distance (for example, doubling the distance results in a four-fold drop in power) between the transmitter and receiver. In operation, a test subject wore the small (2cm cube) Polhemus receiver on a wide fabric elasticised band around the head so that the orientation of the receiver closely tracked the orientation of the head. The transmitter was placed in a fixed position close (within 30cm) to the receiver. The area was cleared of any metallic objects (within 2m) as far as possible to avoid any distortion of the electromagnetic coupling field, hence minimising any tracking inaccuracy caused by the environment.

In operation the Polhemus tracker delivered spatial coordinates between the receiver and transmitter to the test PC. A software driver was written in C++ (Appendices, Figure A9.1) that translated these coordinates in real-time into a head pointing vector between the head of the test subject and the plane of the screen, and then moved the screen cursor position to the intersection of the pointing vector and screen plane. Determining the head-pointing vector on the target screen required only simple geometry with knowledge of the screen plane position, the transmitter position and the head position and orientation in space in front of the screen (Figure 9.2).

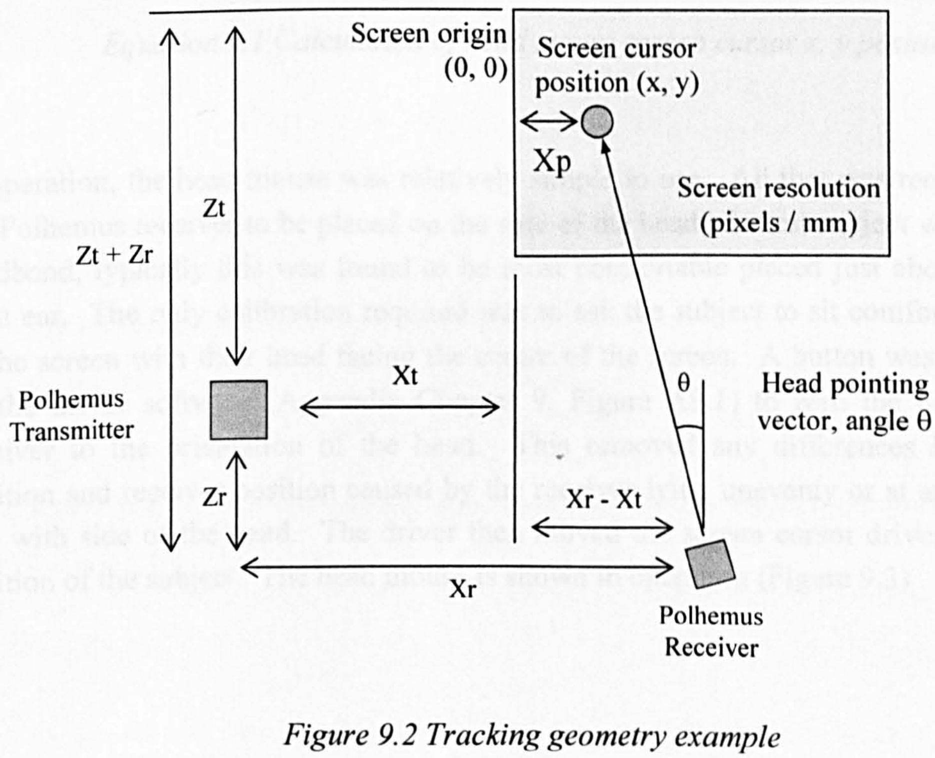


Figure 9.2 Tracking geometry example

To calculate, for example the *x* screen cursor position (Figure 9.2, Y dimension not shown for clarity), the position of the transmitter was measured (Xt, Zt) in relation to the screen origin. This position was permanently fixed. With knowledge of the transmitter position in relation to the screen origin, the position of the receiver was then calculated in relation to the screen origin (Xr – Xt, Zt + Zr). Next the horizontal attitude (Roll) of the receiver to the screen plane was measured. A head-pointing vector (θ) was then constructed from the receiver position to the screen plane. Knowing the position of the receiver (Xr – Xt, Zt + Zr), the attitude of the receiver (θ), the position of the screen plane origin (0, 0) and the screen resolution (pixels /mm) then gave a simple right-angled triangle solution for the screen *x* position of the cursor (Xp) that was scaled in screen pixels from the origin (Equation 9.1). The screen cursor *y* position was calculated in the same way but using the vertical attitude (Pitch) of the Polhemus receiver.

$$\text{Screen cursor } x \text{ position (Xp)} = [\,(Xr - Xt) - (Zt + Zr) \times (\tan\theta)\,] \times (\text{pixels / mm})$$

$$\text{Screen cursor } y \text{ position (Yp)} = [\,(Yr - Yt) - (Zt + Zr) \times (\tan\theta)\,] \times (\text{pixels / mm})$$

Equation 9.1 Calculation of head mouse screen cursor x, y position

In operation, the head mouse was relatively simple to use. All that was required was for the Polhemus receiver to be placed on the side of the head of a test subject with the elastic headband, typically this was found to be most comfortable placed just above the left or right ear. The only calibration required was to ask the subject to sit comfortably in front of the screen with their head facing the centre of the screen. A button was then selected on the driver software (Appendix Chapter 9, Figure A9.1) to zero the position of the receiver to the orientation of the head. This removed any differences between head position and receiver position caused by the receiver lying unevenly or at an angle not in line with side of the head. The driver then moved the screen cursor driven by the head position of the subject. The head mouse is shown in operation (Figure 9.3)

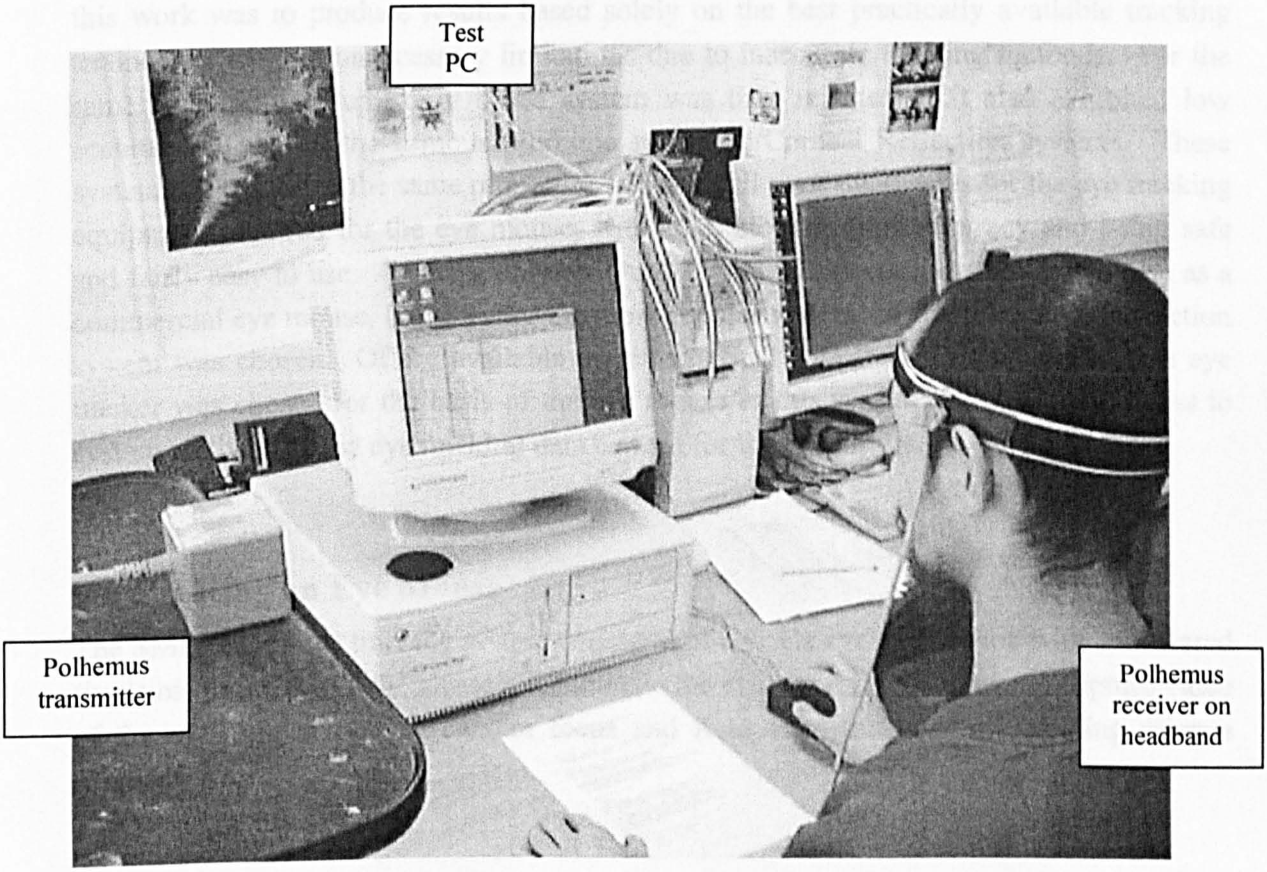


Figure 9.3 The head mouse in operation

9.3 Limitations of available eye mice

An eye tracker was required to generate eye gaze positional data for the eye mouse that was accurate, that was easy and safe to set up and use, readily available, and that was similar in operation to typical eye mice currently available. As introduced in Chapter 2, there were a range of commercial eye tracking devices that could be considered as candidates for the eye mouse in this work (Chapter 2, Table 2.2). As with selecting a suitable head mouse system, the differing eye mouse systems showed differing operational characteristics.

Of the available systems, the Scleral Coil and Contact Lens systems were rejected, despite their accuracy and sampling rate, as they all required invasive contact with the test subject, and all had a low ease of use. The Electro-oculography systems were rejected, despite being in use as an eye mouse commercially, as they are fairly inaccurate and the aim of this work was to produce results based solely on the best practically available tracking technology without unnecessary limitations due to inaccurate tracking methods. For the same reasons, the Pupil only based system was then rejected as it also exhibited low accuracy. This left the Limbus, Purkinje and Pupil/Corneal Reflection systems. These systems all exhibited the same properties and were all good candidates for the eye tracking equipment required for the eye mouse, with all exhibiting good accuracy and being safe and fairly easy to use. Of these only the Pupil/Corneal Reflection system was in use as a commercial eye mouse, being by far the most popular system. A Pupil/Corneal Reflection system was chosen. Of the available systems, a SMI[1] 'RED II' Corneal Reflection eye tracker was chosen for the basis of the eye mouse as this system allowed open access to system calibration and eye tracking data control for the eye mouse.

9.4 Creating an Eye Mouse

The SMI RED II eye tracking system consisted of a single eye tracker box with an infrared (invisible to the eye) light source to illuminate the eye, an infrared camera to capture video of the eye, and automated camera focus and field of vision lens and steering mirrors (Figure 9.4).

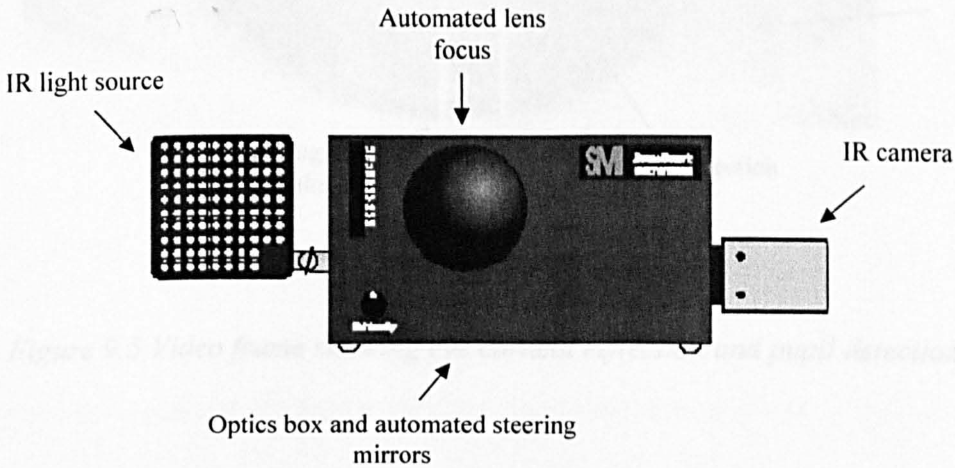


Figure 9.4 SMI RED II eye tracker[1]

[1] From SensoMotoric Systems, www.smi.de

In addition to the eye tracker box a video processing board was provided that processed the video from the camera to detect the pupil of the eye and the corneal reflection from the eye and convert these locations into a gaze position. In this system gaze position was calculated by the changing relationship between the moving dark pupil of the eye and the essentially static reflection of the infrared light source back from the cornea. This approach relied on shining infrared light (to avoid the tracked subject squinting) at an angle onto the cornea of the eye, with the cornea producing a reflection of the illumination source (Figure 9.5).

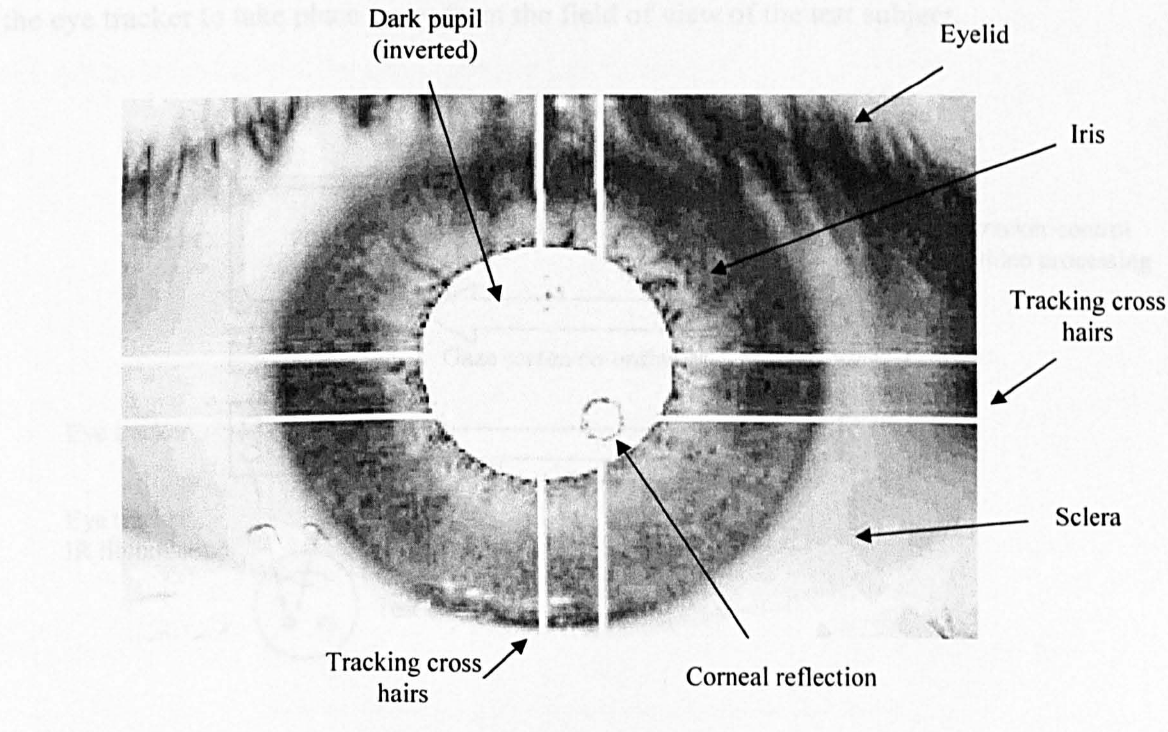


Figure 9.5 Video frame showing eye corneal reflection and pupil detection

The corneal reflection remains approximately constant in position during eye movement hence the reflection will remain static during rotation of the eye and changes in gaze direction, thus giving a basic head position reference. This reflection also provides a simple reference point to compare with the moving pupil and so enables calculation of the gaze direction vector of the eye (for a more detailed explanation see Duchowski, 2000).

The system was set up by placing the eye tracker box on the desk in front of the test PC monitor and feeding the infrared eye video to the video processing card fitted in a dedicated eye tracking PC (Figure 9.6). This PC, PC1, calculated the eye gaze position as screen coordinates on the test PC monitor. A software driver was written in C++ (Appendix Chapter 9, Figure 9.2) to filter the eye-gaze position data in real-time with a simple 4 point rolling average filter running on PC1 to damp gaze point jitter due to small natural saccadic movements of the eye. These damped coordinates were then fed to the test PC, PC2, via a serial cable and then used to move the test PC mouse cursor to the eye gaze position on the test PC screen. Using PC1 as a dedicated machine for all eye data processing removed load from the main machine, PC2, and allowed set up and control of the eye tracker to take place away from the field of view of the test subject.

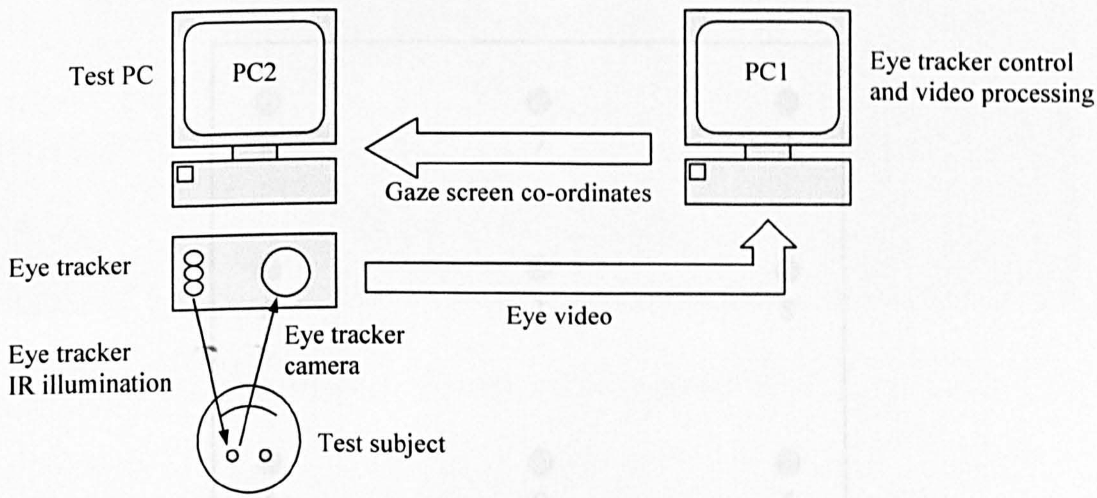


Figure 9.6 Eye tracking equipment arrangement

The eye tracker required calibration to each test subject before use. The subject was seated in front of the test PC, the camera mirrors and lens were adjusted under the control of software to capture one eye (it did not matter which eye was tracked as both eyes move monoscopically). The infrared light source was then adjusted to fully illuminate the chosen eye to create a corneal reflection. This required some skill to correctly position the light source to create a reflection on the cornea rather than the sclera, and to position the reflection away from any obscuration caused by the eyelid. Once this was done the eye tracker could be calibrated. To achieve this a calibration application was written in C++ that displayed a set of nine calibration points on the test PC screen (Figure 9.7).

To calibrate the eye tracker, the test subject was required to simply gaze at each point in numerical order. The eye tracker then mapped the eye corneal reflection and pupil locations of the subject onto the actual *x, y* screen locations of each target on the calibration screen. Once one point was calibrated (typically within 1 second), the software removed that point and illuminated the next target point until all points were calibrated. Points on the screen between the target locations were interpolated from the location data for each of the calibration targets. After calibration, the calibration software invoked the driver software (Appendix Chapter 9, Figure 9.2) that drove the test PC mouse cursor to the eye gaze position of the subject on the test screen.

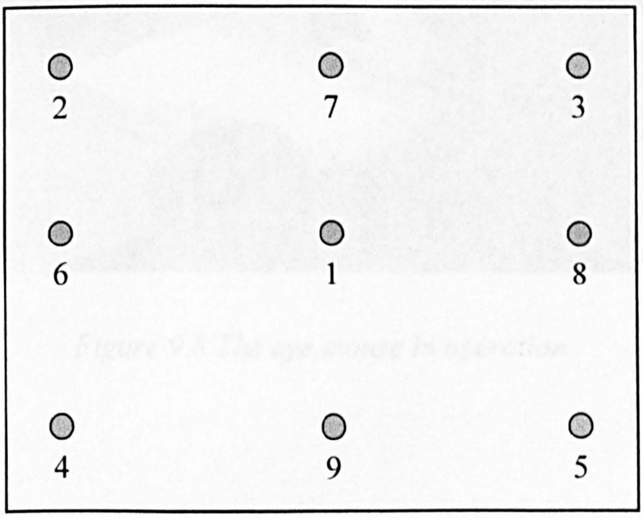


Figure 9.7 Eye tracker calibration screen

During practice sessions it was found that subjects tended to move out of the field of view of the camera, so a chair with a headrest was used to aid subjects to keep their head position within the field of view of the camera. The eye mouse is shown in operation, together with the headrest (Figure 9.8)

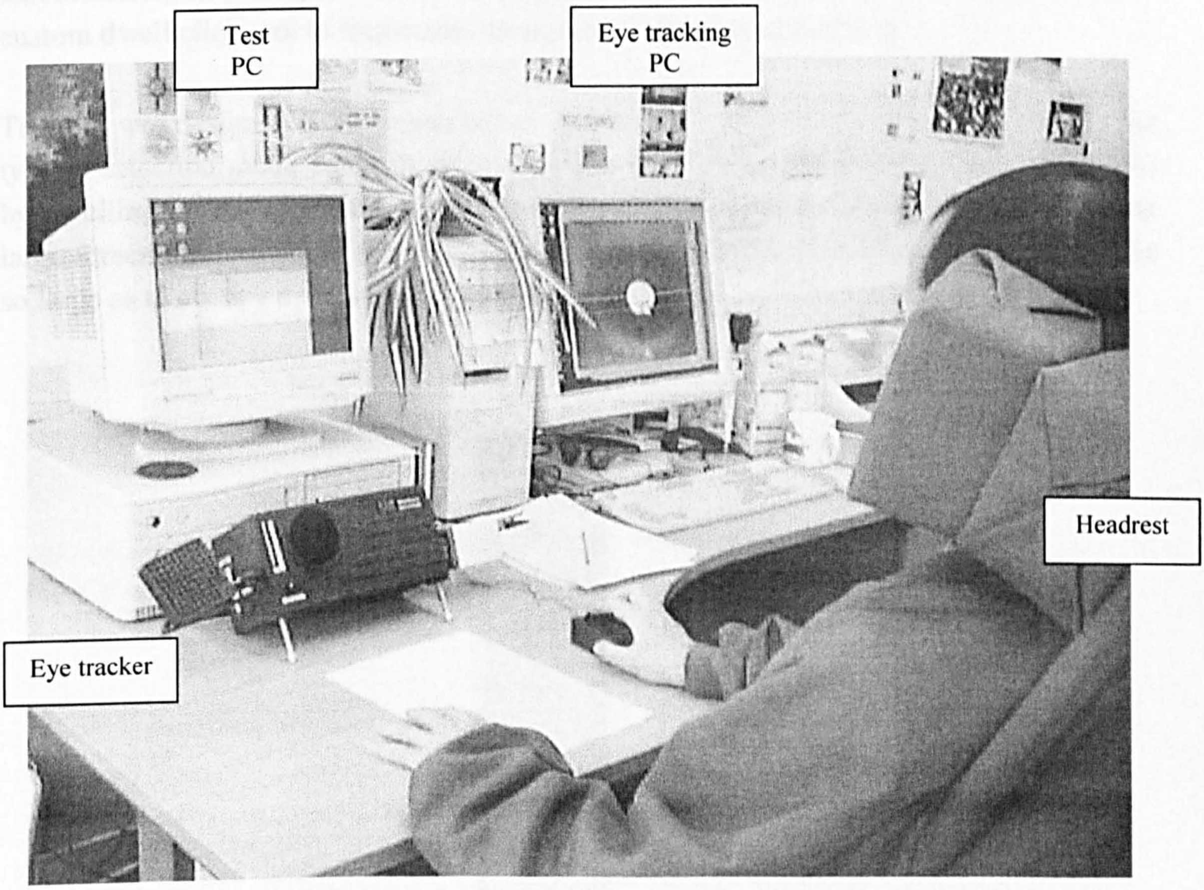


Figure 9.8 The eye mouse in operation

9.5 Creating object selection software

As discussed in Chapter 2, an on-screen dwell click tool was required to allow monomodal selection for users who had no supporting modality available to operate a switch for object selection. It was feasible to use a commercially available application for the dwell click tool; however, none of these applications were available with an open source code. Although the dwell click time could be changed via controls on these applications, it would not be possible to determine what method was used to calculate when the cursor was actually 'dwelling', nor would it be possible to modify this method. The possibility of modifying dwell behaviour, either within this work or during future research, was regarded as most important as previous work had shown that dwell behaviour and error rates could be improved if sophisticated dwell algorithms were used based on patterns of cursor behaviour (Stampe 1993, Jacob 1995, Jacob 1991). Hence it was decided to write a custom dwell click tool to implement monomodal object manipulation.

The tool was designed to be simple to use and intuitive. It allowed subjects to change the type of selection mode currently in operation (single click, double click, drag, no action) by dwelling on the appropriate button, with each button on the device having the same large screen dimensions as a desktop icon or soft keyboard key to aid selection but not be so large as to occupy a disproportionate part of the available screen area (Figure 9.9).

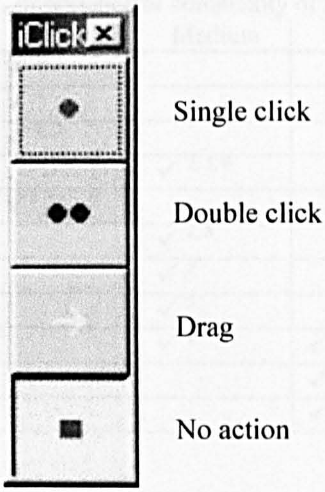


Figure 9.9 Dwell click tool

In operation, the tool sampled the screen cursor position every 50ms and generated a dwell click if the position remained within 1° head or eye visual angle (when seated at 60cm from the screen - equivalent to approximately 10mm on the screen) of the first sample position in a rolling buffer of 20 samples, giving a dwell click time of 1000ms. Note that eye blinks and other losses of tracking during dwell timing would interrupt and restart the timing.

A survey of previous work with, particularly, eye-based interaction using dwell selection was used to determine the dwell time for the tool. Dynamic or 'intelligent' dwell times, as discussed in Chapter 2, were rejected for this work as these would add additional permutations of the results beyond the scope of this work (but which may be the subject of future work), giving permutations of performance not just of head or eye mouse and interface, but also of selection of dynamic dwell time and technique. However, these are hoped to be the subject of future work. Instead, it was decided to use a simple fixed dwell time. Previous work had used dwell times ranging from 100ms to 1500ms depending on the importance or consequence of the action commanded by the selection, the complexity of the arrangement of interface targets and the level of errors tolerable during the interaction. Due to this, longer times had typically been used for more 'important' or complex selections (Table 9.1).

Typical dwell times for importance of selection			
Dwell time (ms)	Importance or complexity of selection		
	Low	Medium	High
100	✓[2,3]		
150	✓[4]		
250	✓[2,3]		
400		✓[2,3,6]	
500	✓[1]		
600		✓[2,3]	
750		✓[5]	
850		✓[5]	
1000		✓[1]	✓[1,2,5]
1500			✓[1]
>1500			✓[1]

Table 9.1 Typical dwell selection times and importance or consequences of selection

Typically error rates had been found to rise quite considerably with shorter dwell times (Istance and Howarth 1993) and more difficult selections (Velichkovsky et al. 1997). It had also been suggested that dwell should not be used for highly consequential selections as the dwell time required for a reliable selection would be excessive (Jacob 1991). It was clear (Table 9.1) that longer dwell times were regarded as more desirable for selection when interacting with a more complex interface and when the consequences of incorrect selection were important. Since it was desirable to achieve close to error free interaction on a complex 'real world' interface, a longer dwell time of 1000ms was selected for the dwell click application. This choice represented a compromise between a short dwell time that could cause inadvertent dwell selections and an excessively long dwell time that would unnecessarily slow-down interaction and require the test subjects to fixate on targets for unnaturally long periods.

[1] Istance et al. (1996), [2] Jacob (1990), [3] Jacob (1991), [4] Sibert and Jacob (2000), [5] Stampe and Reingold (1995), [6] Ware and Mikaelian (1987).

9.6 Creating object selection hardware

For multimodal object manipulation, where the user has an available supporting modality to operate a switch, for example, conventional object selection was adopted. A small box containing a micro-switch was constructed, with the user holding the switch in the left or right hand (Figure 9.10). The use of the hand as a supporting modality for the multimodal eye mouse gave a multimodal input system to contrast with the monomodal dwell click system. Although in practice a disabled person may not have sufficient motor function to use a hand switch, any other form of switch control, as discussed in Chapter 2, such as sip/puff switches or blink switches, may be possible and offer the same performance. In this case the hand was chosen as the switch control modality as it offered the most familiar, highest performance and most reliable supporting modality for comparison.

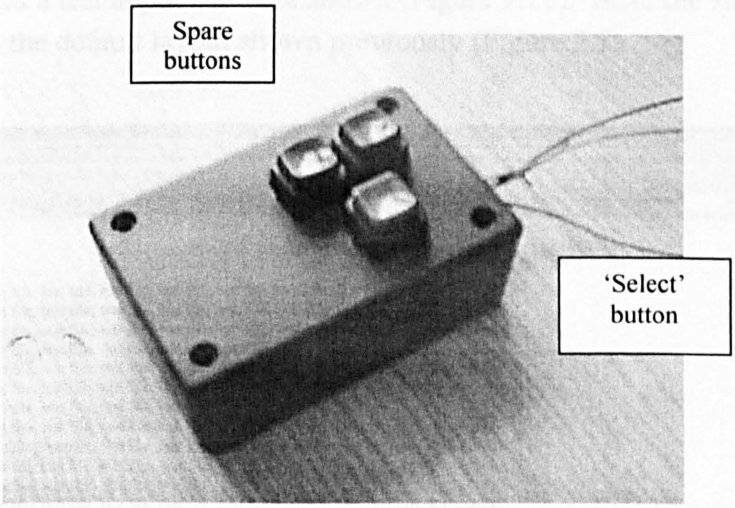


Figure 9.10 Switch click tool

9.7 Creating an on-screen keyboard

As with the on-screen dwell click tool discussed previously, Chapter 2 also discussed the need for, and diversity of, on-screen dwell click tools that would be required to allow text generation. Unlike the dwell click tool, there was no need to build a custom application for the on-screen keyboard, as any dwell selections on the keyboard keys would be generated by the dwell click tool, not the keyboard, so no customisation of the internal working of the keyboard would be necessary. As discussed in Chapter 2, there were a range of possible keyboards that could be chosen for this work, ranging from very simple full-screen keyboards to complex dynamic keyboards, each with their own strengths and weaknesses. (The use of an on-screen keyboard does move the work away from 'direct' interaction with he interface as the keyboard is an adjunct to the standard interface, however without it text entry would not be possible). Since this work was designed to be generic, and not influenced or specific to any strength or weakness of a keyboard, it was decided to use a simplified standard keyboard for text entry. The keyboard layout and available keys were modified with simplified reductions of the comprehensive layout found on standard on-screen keyboards[1] (Figure 2.5). Keys that would not be used during the assessment method tasks (Chapter 4) were removed to simplify the layout and reduce the number of keys that needed to be displayed. This reduction from 92 keys to 73 keys was done to aid test subjects in finding the required keys during the test and also to allow the remaining keys to occupy as much screen area as possible, making selection easier. The appearance and placement on the interface of the simplified on-screen keyboard, the dwell click tool, and a test application are shown (Figure 9.11). Note the simplification of the keyboard from the default layout shown previously (Figure 2.5).

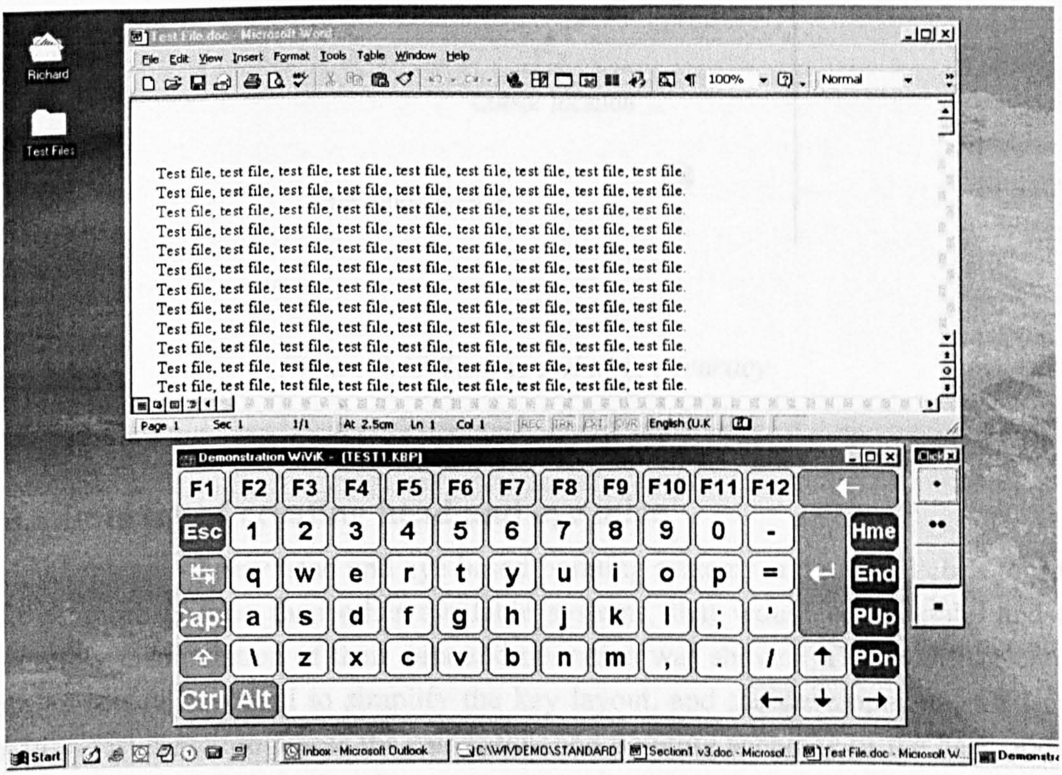


Figure 9.11 Placement of the keyboard and dwell click tool

[1] 'WiViK' on-screen full-function keyboard from Prentke Romich, www.prentrom.com

9.8 Testing pointing accuracy

A variant of the eye tracker calibration screen (Figure 9.7) was written to test the pointing accuracy of the head and eye mice after calibration and before a test would be started (Figure 9.12). This software was written in C++ and displayed nine targets on the test PC screen. Subjects pointed at each target in numerical order for 1 second per target using either the head or eye mouse. The software then recorded the cursor location during this time and calculated the distance between the target centre and screen cursor location for each target, storing the results as a text file. This software was used to check the accuracy of calibration of the head and eye mice against an expert user baseline before subjects undertook tests with the devices, and avoided undertaking tests with poor calibration of the devices.

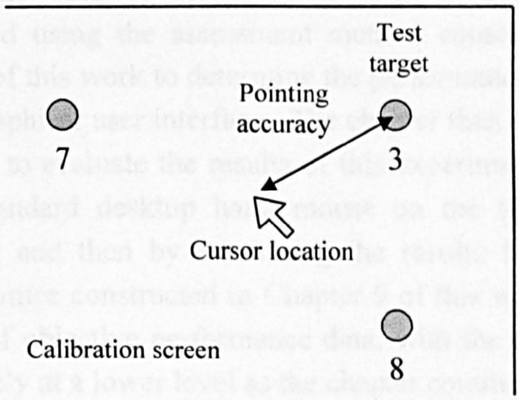


Figure 9.12 Testing pointing accuracy

9.9 A summary of creating head and eye mice

The development of new head and eye based pointing systems and a dwell click tool that would be more generic than other available systems, that would be accurate, and that would allow manipulation of their data and operation was shown. The modification of a simple on-screen keyboard to simplify the key layout, and the development of a simple calibration test screen to assess the calibration and pointing accuracy of the head and eye mice created in this chapter was shown.

Chapter 10

The Performance of Hand, Head and Eye Mice

The aim of this chapter is to compare the performance between eye based pointing and head based pointing during interaction with a normal, unmodified graphical user interface. The chapter determines if direct eye and head based interaction can be effective with an unmodified graphical user interface, what are the main factors influencing direct eye and head based interaction on an unmodified graphical user interface, and to what extent can a eye and head mouse achieve the same performance as the benchmark hand mouse on an unmodified graphical user interface.

This chapter builds on the work of the previous chapters, it first describes how an experiment was conducted using the assessment method constructed in the preceding chapters, chapters 4 to 8, of this work to determine the performance of hand, head and eye mice on an unmodified graphical user interface. The chapter then uses the metrics devised in Chapter 5 of this work to evaluate the results of this experiment, firstly by examining the performance of a standard desktop hand mouse on the test method to create a benchmark for the work, and then by examining the results for the monomodal and multimodal eye and head mice constructed in Chapter 9 of this work. This chapter starts with high-level analysis of objective performance data, with the level of analysis of this data becoming progressively at a lower level as the chapter continues. Finally, the chapter ends with an analysis of the subjective user reaction to the devices.

10.1 Experiment 5: The performance of hand, head and eye mice

Five devices were to be assessed; the baseline hand mouse, and the monomodal (dwell click) head and eye mice and multimodal (switch click) head and eye mice described previously in chapter 9, by using the 'real world' assessment method devised and validated previously in chapters 4 to 8 of this work.

A within subjects test design was adopted, with all subjects using all devices, to enable detailed comparison between the performances of each subject with each device. To compensate for order effects in the testing, the presentation order of the devices was prescribed with an incomplete Latin Square design (Appendices Tables A10.1 and A10.2).

Independent variables were:

- Device (Hand mouse, monomodal head and eye mouse, multimodal head and eye mouse) (Chapter 9).
- User experience (Low, Medium, High) (This chapter).
- Task target size (S1, S2, S3, S4 in pixel/mm/visual angle) (Chapter 4).
- Task interaction type (Single click, Double click, Drag, Restricted/Unrestricted) (Chapter 4).

Dependent variables were:

- Task Efficiency (%) (Chapter 5).
- Task Quality (1-5) (Chapter 5).
- Task Time (mS) (Chapter 5).
- Task time taken by non-productive actions (mS) (Chapter 5).
- Device pointing accuracy (pixel/mm/visual angle from test targets) (Chapter 9).
- Device assessment questionnaire (workload, comfort, ease of use, 1-7) (Chapter 5).

10.2 Test subjects

One important element of the experiment, and hence choice of test subjects, was to investigate how experience with the devices affected performance. The number of available participants with a wide range of experience of both test devices was limited due to the rarity of eye mouse devices in general usage. Hence six test subjects with a wide range of experience (from very experienced users through to novice users with little previous experience with the assistive technology devices) were chosen for the experiment (Table 10.1). The experience rating of the test subjects was determined by counting the hours of use with each of the devices, and also comparing their general level of performance with the devices in comparison to a very highly experienced subject with each of the devices. Of note was that test subjects required considerable time with the eye based system to reach medium and high experience ratings, this was in contrast to the head based system, which required only short periods of time to achieve a high experience rating. Clearly the eye mouse required more learning time than the head mouse, leading to further investigation of the effects of experience on performance in this work.

All of the subjects were staff or students from the University with experience using the test applications and Windows interfaces with a hand mouse, and all subjects were able-bodied[1]. There was no penalty for non-participation, all subjects signed a consent form (Appendices Figure 10.1) and all data was anonymous. Subjects were unpaid and were told that the tests were part of research work in the university and that they should complete the tests as accurately as possible. Subjects 1-4 were male and subjects 5-6 female, subjects 2 and 3 wore vision correction glasses and subject 5 wore contact lenses during the experiment. The mean age was 28 years with the oldest subject 34 and the youngest 24.

Subject experience			
Subject numbers	*Typical pre-test hours of experience with head-tracking system*	*Typical pre-test hours of experience with eye-tracking system*	*Experience rating*
1, 2	2 - 3	15 - 30	High
3, 4	0.5 - 1	6 - 8	Medium
5, 6	0.25	1 - 2	Low

Table 10.1 Test subject experience with the head and eye tracking systems

10.3 Test procedure

A standard Pentium II PC running Windows 98 with a standard keyboard and mouse and a 17" monitor were selected for the test PC. The test PC was prepared by loading the appropriate Word and Internet documents for the test tasks (Chapter 4, Appendices Figure 4.1) and the soft keyboard (Chapter 9). The screen was laid out with the soft keyboard

[1] Subjects with a disability were not used as: 1) Eye control is generally not affected by high level motor disability; hence subjects may be either disabled or non-disabled. 2) It was not considered ethical at this stage to allow the use of potentially enabling systems with no possibility of then being able to provide those systems to the disabled test subjects (after consultation with a specialist in the field - Dr Clive Thursfield, Access to Communication and Technology, West Midlands Regional Rehabilitation Centre, Birmingham).

occupying the bottom 1/3 of the screen leaving the top 2/3 of the screen free for the test applications, and when required for the head and eye mice only the dwell click tool was started and placed to the bottom right of the screen to enable monomodal selection for the multimodal head and eye devices (Chapter 9, Figure 9.11). Note that dwell click was not tested with the hand mouse as this device was used as a baseline in unmodified form. Finally, a screen capture program was loaded on to the test PC that would capture the entire contents of the test computer screen, including the cursor position and a visible marker of any selection actions, at a rate of 5 frames per second using the commercial 'Hypercam' screen capture application[1] recording to hard disk. Paperwork was prepared by printing copies of the test tasks (Chapter 4, Appendices Table A4.3), the quality-rating scheme (Chapter 5, Table 5.1), a set of test marking sheets (Appendices Table A4.4), a set of questionnaire sheets (Chapter 5, Appendices Tables A7.4, A7.5 and A7.6) and finally consent forms (Appendices Figure A10.1).

The nature of the tests and the usability questionnaires were explained to the test subjects. Subjects were then asked to sign consent forms guaranteeing confidentiality and explaining that they could rest during tests and withdraw from the tests at any time and for any reason without suffering penalty. No subjects opted to withdraw from the experiment. The test subjects were all initially familiarised with the test tasks by performing the assessment method test tasks with the hand mouse under direction of the test administrator until the subjects felt confident performing the test sequences. To ensure the test subjects performed the correct test tasks in the correct order, the test administrator verbally narrated the required actions to the test subjects at the start of each test task.

Subjects were allowed practice calibrations of the head and eye mice (Chapter 9) until they consistently achieved an accuracy of 75% of the calibration accuracy of an expert user. A level of 75% was chosen as the level above which reasonable interaction was possible without subjects objecting to 'poor calibration'. The pointing accuracy of the subjects with the devices was recorded after device calibration and before each test by asking the subjects to point at 9 equally spaced targets on the screen (Chapter 9) with the overall mean distance of the cursor from the targets recorded. From this, tests were only conducted with calibrations exceeding 75% of the accuracy obtained by expert users with the devices to remove the possibility that a poor calibration would affect the test results. The average calibration accuracy achieved by an expert user was previously determined to be an accuracy of within 0.9° (visual angle at 60cm screen to eye distance) from each of the 9 calibration screen targets for the eye mouse and 0.3° for the head mouse. Once the subjects had consistently achieved the required calibration accuracy they practised using

[1] Hypercam from www.hyperionics.com

the head or eye mouse (depending upon which device was to be used for the following test session) by playing the Windows 98 Solitaire card game until they felt comfortable with the operation of the system. Typically this took 5 minutes with each device.

For the test, subjects were seated at a screen to head or eye distance of 60cm (from practice sessions this distance gave the minimum head or eye to screen distance, and hence highest pointing accuracy, whilst allowing reasonably comfortable head or eye range of movement to cover the whole screen), using the headrest with the eye mouse, and were instructed to work at their normal pace and maintain a high level of accuracy. At any time during the test the subjects were allowed to pause to recalibrate the device. This time taken by additional calibrations was included in the total calibration time for each test. Once the subjects had finished a test they completed the satisfaction questionnaires. Subjects were asked to rest for a minimum of 20 minutes between tests. The test procedure is summarised as follows (Table 10.2):

Test procedure				
Test step	*Test stage*	*Task*	*Time taken (typically)*	
1	Set up	Take subject details	2 mins	
2		Explain nature of tests	4 mins	
3		Explain questionnaires	2 mins	
4		Read and sign consent form	2 mins	
5		Familiarise with test tasks using hand mouse	15 mins	
6		Practice head and eye mouse calibrations	10 mins	
7		Familiarisation practice with head and eye mice playing game	10 mins	
8	Test tasks	Calibrate head or eye mouse	2 mins	Repeat for each device (steps 8–10)
9		Undertake test tasks	20 to 40 mins	
10	Rest	Rest, or return later	>20 mins	

Table 10.2 Test procedure for a subject

10.4 Analysis

Object sizes from the assessment method test tasks (Chapter 4) were converted from screen pixels to the visual angle subtended by the object size categories at the test screen to head or eye distance of 60cm, giving object size categories of 0.3°, 0.6°, 0.9° and 1.2° of visual angle at 60cm. This conversion was carried out since both the head and eye devices are angular pointing devices, with their apparent screen accuracy in pixels or mm dependent upon the angle from the head or eye and distance the head or eye was from the screen. Giving results in visual angle was felt to be more suitable as this approach allowed calculation of equivalent object sizes for any head or eye to screen distance, screen resolution or object size in pixels or mm.

The data were analysed by stepping through the captured video files and noting the quality and time taken to perform each task using the assessment method marking sheet (Appendices Figure A4.4) in the method described previously (Chapter 5). In addition, the time taken by any non-productive actions during each task was measured and the nature of the non-productive action was recorded. The time taken for each task was measured as the time from the initial movement of the cursor toward the target to the selection of the target or the failure of the task. This approach removed any additional time taken by the subject to understand and respond to the task narration from the test administrator. Dependent variable performance metrics were calculated for each of the independent variables (This chapter). The device assessment questionnaire results were calculated for each independent variable (Chapter 5). All data was displayed in terms of medians (mid-bars on graphs), quartile ranges (boxes on graphs) and outliers (asterisks on graphs), with means also included (circles on graphs) to give a fuller picture of the ranges and distributions of the data[1]. Finally, the statistical significance of any difference between metrics was determined with appropriate statistical tests.

10.5 Baseline hand mouse efficiency

The overall median task efficiency (defined in Chapter 5.6) for the standard desktop hand mouse on the assessment method was 83.3% for pooled data across the two assessment domains (Appendices Table A10.3, Graphed in Figure 10.1). This high level of performance was a result that was expected (only a device that produced no errors and took no time to complete a task would achieve 100% performance on the test tasks), as a hand mouse is the device of choice for manipulation of the test interface and so should perform well. Examining the range of performance indicated by the inter-quartile range

[1] See Appendix Notes Figure 1 for an explanation of the graph symbols.

markers and outlier markers for the hand mouse on the graph of results (Figure 10.1) showed a tight distribution at a high level of efficiency with only a few outliers for both test domains[1], indicating that the device was capable of both achieving and maintaining a consistently high degree of efficiency across a wide range of tasks.

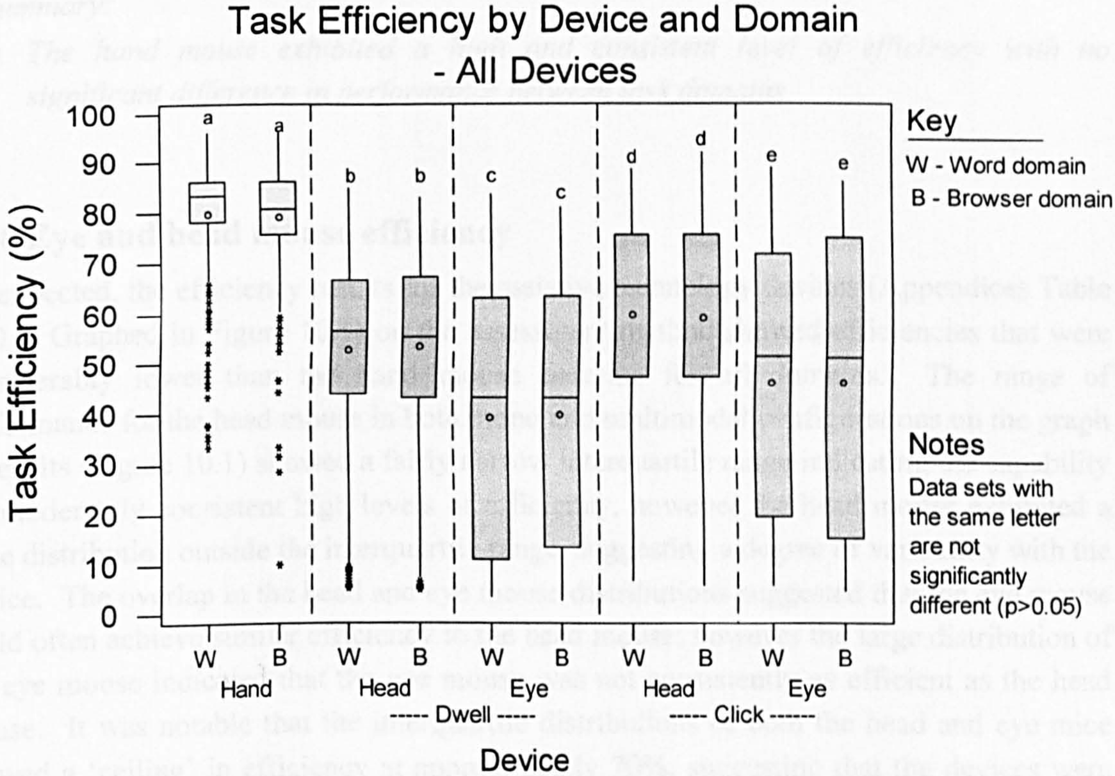


Figure10.1 Device task efficiency by domain[2]

A Mann-Whitney two-sample rank test[3] (Sprent 1993) was used to show that this difference between the two domains was not statistically significant (p = 0.671) (comparison shown in Appendices Table A10.4). The similarity of device performance between the word and web task domains showed that the context or nature of the tasks had

[1] Although there appear to be a moderate number of outliers on the graphical representation of the data, these are few in comparison to the number of samples in each domain, with 6 subjects * 82 tasks = 492 samples in the Word domain and 6 subjects * 68 tasks = 408 samples in the browser domain.

[2] Note that monomodal operation uses a *dwell* tool for selection, and multimodal operation uses a *click* device.

[3] The Mann-Whitney two-sample rank test is a non-parametric equivalent to a two-sample t-test based on the ranks of the data. It is particularly useful for data with non-normal distributions where data sets have different sample numbers and the data do not have something in common – in this case they come from different test domains. See Appendices Notes Discussion 1. "Non-parametric tests in this work".

little effect on the performance of the hand mouse device. The frequencies of target sizes present in each of the tasks in the domains were similar (Chapter 4), and it was probable that this accounted for the similarities in measured performance much more than the context of the tasks.

In summary:

- *The hand mouse exhibited a high and consistent level of efficiency with no significant difference in performance between task domains.*

10.6 Eye and head mouse efficiency

As expected, the efficiency results for the assistive technology devices (Appendices Table A10.3, Graphed in Figure 10.1) on the assessment method showed efficiencies that were considerably lower than the hand mouse baseline for all domains. The range of performance for the head mouse in both mono and multimodal configurations on the graph of results (Figure 10.1) showed a fairly narrow interquartile range indicating the capability for moderately consistent high levels of efficiency, however the head mouse exhibited a wide distribution outside the interquartile range, suggesting a degree of variability with the device. The overlap in the head and eye mouse distributions suggested that the eye mouse could often achieve similar efficiency to the head mouse; however the large distribution of the eye mouse indicated that the eye mouse was not consistently as efficient as the head mouse. It was notable that the interquartile distributions of both the head and eye mice showed a 'ceiling' in efficiency at approximately 70%, suggesting that the devices were unlikely to achieve the same levels of efficiency as the hand mouse under normal circumstances.

The performance of the assistive technology devices appeared to show little difference within device and selection mode between the two test domains (Figure 10.1). This similarity was confirmed with Mann-Whitney two-sample rank tests showing no statistically significant difference between domain performances within each device (comparison shown in Appendices Table A10.4). As with the hand mouse baseline results, this similarity between domains confirmed that the context or nature of the tasks had little effect on the performance of the devices. As there were no statistically significant differences between domain performances within all of the devices, the domain results were pooled for each device to give greater clarity and ease of comparison of the performances of the devices to each other (Figure 10.2).

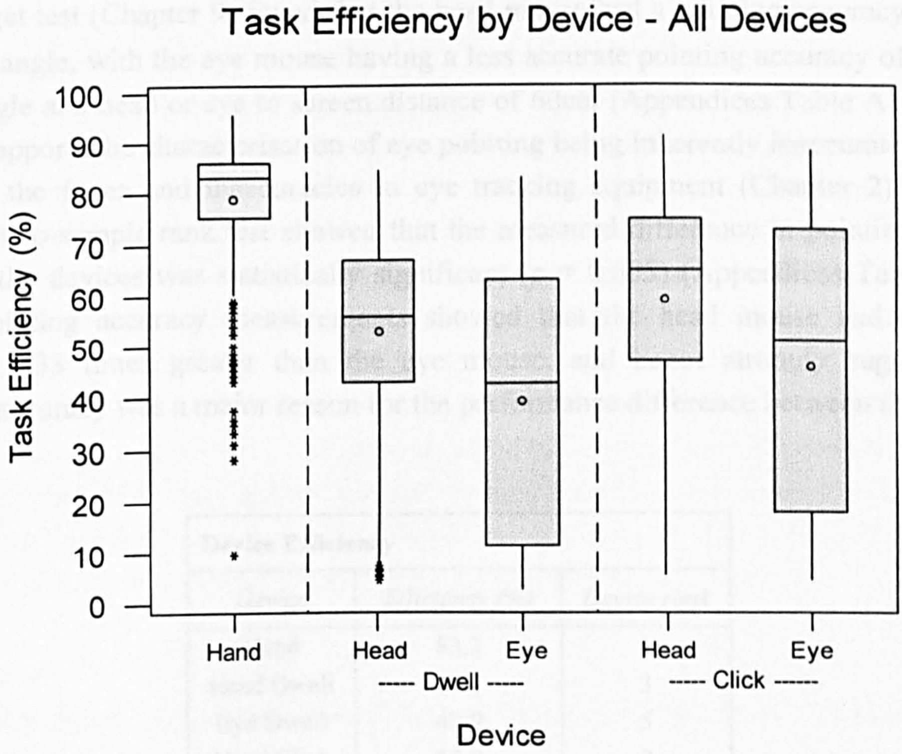


Figure 10.2 Device overall task efficiency

Wilcoxon matched pairs signed rank tests[1] (Sprent 1993) were then used to investigate the significance of differences between the pooled domain efficiencies within each device (Appendices Table A10.5). The comparisons showed that the pooled performances of all devices were statistically significantly different from each other, showing that both device and selection modality affected device efficiency. Ranking the devices by efficiency (Table 10.3) showed that the hand mouse outperformed all devices and that the head mouse outperformed the eye mouse in both monomodal and multimodal configurations, hence the choice of device had more influence on measured efficiency than the choice of selection modality. However, multimodal selection was more efficient than monomodal selection for both devices, so clearly the method of target selection also had a strong effect on device performance as monomodal selection task time includes an additional 'dwell' time for the dwell selection tool.

[1] The Wilcoxon matched pairs signed rank non-parametric test is used when distributions are not normal, sample sizes are equal and both data sets have commonality – in this case both sets of data originate from the same set of test tasks and each data sample from one device can be paired with a corresponding sample from the another device. See Appendices Notes Discussion 1. "Non-parametric tests in this work".

Examining the pointing accuracy of the devices measured before each test using the 9-point target test (Chapter 9) found that the head mouse had a pointing accuracy of 0.286° of visual angle, with the eye mouse having a less accurate pointing accuracy of 0.955° of visual angle at a head or eye to screen distance of 60cm (Appendices Table A10.6). This finding supports the characterisation of eye pointing being inherently inaccurate due to the width of the fovea and inaccuracies in eye tracking equipment (Chapter 2). A Mann-Whitney two-sample rank test showed that the measured difference in pointing accuracy between the devices was statistically significant (p = 0.005) (Appendices Table A10.6). These pointing accuracy measurements showed that the head mouse had a pointing accuracy 3.33 times greater than the eye mouse, and hence strongly suggested that pointing accuracy was a major reason for the performance difference between the devices.

Device Efficiency		
Device	*Efficiency (%)*	*Device rank*
Hand	83.3	1
Head Dwell	56.4	3
Eye Dwell	42.9	5
Head Click	65.2	2
Eye Click	51.1	4

All data sets are significantly different ($p < 0.05$)[1].

Table 10.3 Device efficiency comparisons and rankings

Comparing the efficiencies of the devices pooled across the two test domains and comparing within modalities (Table 10.3) showed that the monomodal head mouse had efficiency 1.31 times greater than the monomodal eye mouse, and the multimodal head mouse had efficiency 1.28 times greater than the multimodal eye mouse. These efficiency ratios were considerably smaller than the ratio of pointing accuracy between the devices at 3.33 times, and suggested that pointing accuracy alone, measured by an abstract target acquisition test, tended to give an inaccurate and exaggerated difference between the devices. Such simple measurement clearly did not truly reflect the actual performance of the devices on a 'real world' interface as determined by the test method. This finding alone strongly supported the choice (Chapter 3) of a 'real world' test method over more

[1] Note these results are not corrected for alpha, see Notes Discussion N2. Multiple comparisons in this work in the Appendices for a discussion on multiple comparisons.

conventional abstract target acquisition tests when assessing devices to be used on 'real world' interfaces and tasks.

The relationship between pointing accuracy and efficiency was further investigated by examining the correlation between measured pointing accuracy and device efficiency across test subjects (Appendices Table A10.6). Here there were no statistically significant correlations between head mouse pointing accuracy and monomodal (p = 0.787) or multimodal (p = 0.872) head mouse efficiency, indicating that the efficiency of the head mouse was not solely influenced by the pointing accuracy of the device. In contrast, there were statistically significant correlations between eye mouse pointing accuracy and monomodal (p = 0.042) or multimodal (p < 0.001) eye mouse efficiency, indicating that pointing accuracy had a strong influence on eye mouse efficiency. In order to explain the differences found in this section further investigation into the factors influencing the efficiency of the devices was required based on the detailed task time and task quality elements that comprised the efficiency metric (from Chapter 5).

In summary:

- *There are no differences in performance between the Word and Browser domains for each device*
- *The hand mouse outperformed all assistive technology devices*
- *The head mouse outperformed the eye mouse in both monomodal and multimodal configurations*
- *Multimodal click selection outperformed monomodal dwell selection for all devices*
- *The head mouse had a greater pointing accuracy than the eye mouse*
- *Pointing accuracy did not have a strong influence on head mouse efficiency but had a strong influence on eye mouse efficiency*

10.7 Device task time

The test method design allowed further investigation into the total task times and task time elements (Chapter 5) that influenced the differences in efficiency between the devices. With device efficiency proportional to device task time (efficiency = device task quality / device task time), an investigation into task times was essential to understand the underlying factors influencing device efficiency. Total task time results for the devices were calculated (Appendices Table A10.7, summary Table 10.4) and any statistical significance of any difference determined (Wilcoxon matched pairs signed rank test, Appendices Table A10.8).

Device task time	
Device	*Task time (ms)*
Hand	1246
Head Dwell	3489 ^{a}
Eye Dwell	3668 ^{a}
Head Click	2537
Eye Click	3289

Data with the same letter are not significantly different ($p > 0.05$)[1]

Table 10.4 Device task time comparisons

From this (Table 10.4) all of the total task times were found to be statistically significantly different from each other except for the head and eye monomodal dwell selection devices. This may be caused by the dwell click soft tool (Chapter 9) causing the test subject to wait in order to generate a dwell click. Using dwell click had the penalty of extending task times. Further investigation (Appendices Table A10.7, Graphed in Figure 10.3) showed that the hand mouse easily outperformed the other devices, suffering no appreciable time delays from corrections or errors in interaction. The analysis showed that the largest unproductive factor for the head and eye mice in both mono and multimodal configurations was the time lost in positional cursor control corrections (defined in Chapter 5), with the eye mouse losing 44.5% of interaction time in monomodal configuration and 39.4% of time in multimodal configuration (shown in blue on Figure 10.3). The head mouse lost considerably less time in comparison with 8.6% in monomodal and 16.5% in multimodal configuration. All other losses in task time were less than 5% of total task time for each factor.

It was notable that the eye mouse suffered from repeated movements of the cursor from the interaction point to the part of the screen that was providing feedback to the user (the 'feedback point'), such as moving from the on-screen keyboard to the text on the word-processor when typing (shown in green on Figure 10.3). Here the real world test approach clearly showed advantages over simpler abstract target acquisition tests, as it would be unlikely that such feedback point movements would be detected by these simple tests. This feedback point factor was caused by the association of the cursor location to the point of gaze of the user, meaning that, unlike the other devices, the cursor followed the point of

[1] Note these results are not corrected for alpha, see Notes Discussion N2. Multiple comparisons in this work in the Appendices for a discussion on multiple comparisons. Overall mean times shown for task time, non-parametric statistical comparisons used raw task time data.

attention of the test subject. This was a disadvantage, as the cursor could not be left on a target whilst monitoring the response of the interface, meaning that the target must be continually reacquired to continue interaction. This factor was unique to the eye mouse and used 4.8% of interaction time for the monomodal eye mouse and 3.8% for the multimodal eye mouse.

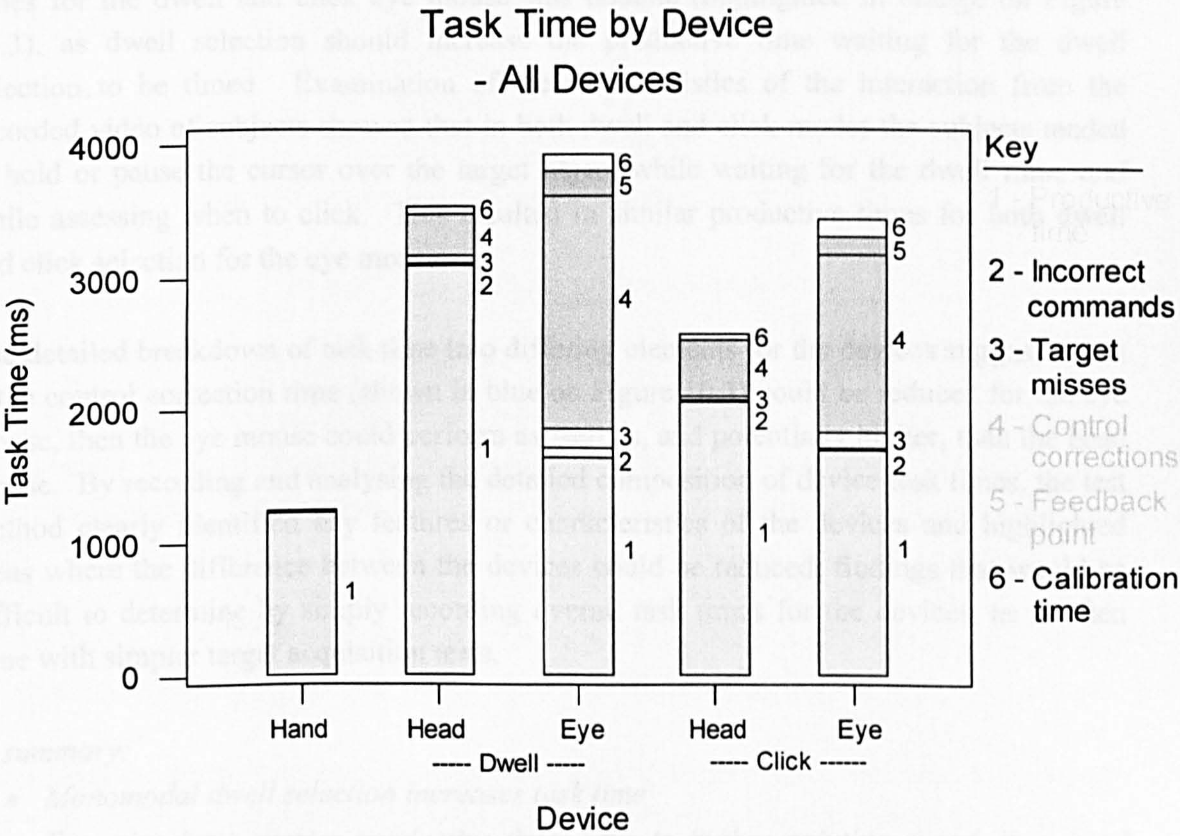


Figure 10.3 Composition of device task time[1]

This level of analysis revealed that the head mouse had longer productive times than the eye mouse when comparing the head and eye devices in either mono or multimodal configuration. This indicated that the head mouse exhibited slower productive interaction than the eye mouse but outperformed the eye mouse overall due to the higher non-productive times of the eye mouse. This was explained by the rapid and productive cursor movement of the eye mouse into the region of the target object being followed by non-productive positional control corrections of the gaze position on the target due to the

[1] The elements of task time as shown on the graph are defined in Chapter 5.

inaccuracy of the eye mouse. These movements degraded eye mouse performance quite considerably. These results coupled with the measured pointing accuracies of the head and eye mice (Appendices Table A10.6) suggested that the head mouse could be characterised by slower movement times but greater pointing accuracy due to lower control correction times, and the eye mouse by more rapid movement times but poorer pointing accuracy due to higher control correction times. The similarity in productive times for the dwell and click eye mouse was notable (highlighted in orange on Figure 10.3), as dwell selection should increase the productive time waiting for the dwell selection to be timed. Examination of the characteristics of the interaction from the recorded video of subjects showed that in both dwell and click modes the subjects tended to hold or pause the cursor over the target object while waiting for the dwell time, *and* while assessing when to click. This resulted in similar productive times for both dwell and click selection for the eye mouse.

The detailed breakdown of task time into differing elements for the devices suggested that if the control correction time (shown in blue on Figure 10.3) could be reduced for the eye mouse, then the eye mouse could perform as well as, and potentially higher, than the head mouse. By recording and analysing the detailed composition of device task times, the test method clearly identified key features or characteristics of the devices and highlighted areas where the difference between the devices could be reduced; findings that would be difficult to determine by simply recording overall task times for the devices, as is often done with simpler target acquisition tests.

In summary:

- *Monomodal dwell selection increases task time*
- *Eye mice have shorter productive times, due to higher pointing speed, than head mice but have longer total task times than head mice due to greater positional control corrections caused by lower pointing inaccuracy*
- *If the non-productive eye mice times could be reduced then an eye mouse could equal or exceed the performance of a head mouse*

10.8 Device task quality

With device efficiency also proportional to device task quality (efficiency = device task quality / device task time), an investigation into task quality ratings was also essential to understand the underlying factors influencing device efficiency. The total task time results for the devices were calculated (Appendices Table A10.9, summary Table 10.5) and any statistical significance of any difference determined (Wilcoxon matched pairs signed rank tests between pairs of devices only, Appendices Table A10.10).

Device task quality	
Device	*Task quality (1-5)*
Hand	4.90
Head Dwell	4.26 [a]
Eye Dwell	3.25
Head Click	4.23 [a]
Eye Click	3.42

Data with the same letter are not significantly different ($p > 0.05$) [1]

Table 10.5 Device task quality comparisons

From this (Table 10.5) all of the total task quality ratings were found to be statistically significantly different from each other except for the head monomodal and multimodal selection devices. This showed that the head mouse task quality was not affected by the mode of target selection. In contrast there was a difference between monomodal and multimodal eye mouse task quality, with the monomodal eye mouse having a lower quality rating than the multimodal eye mouse. Clearly, using dwell click had the penalty of reducing task quality when used with an inaccurate pointing device.

Investigation into the counts of incorrect commands, target misses and control correction movements (Chapter 5) that contributed to the quality scores was performed. Analysis of the factors revealed the causes of the differences in quality between the devices (Appendix 2 Table A10.9, Graphed Figure 10.4). Counting the mean number of quality errors per interaction, there was a noticeable reduction in error rates with an increase in the severity of the consequences of the error (as defined in Chapter 5). This showed that the test subjects took more care when the consequences of error were more severe (such as pressing an incorrect button and causing an incorrect command to be generated, as opposed to simply missing a target object with no commands generated). Examining the results for each device, the hand mouse exhibited very few errors of any type, leading to a high quality rating. The head mouse in both mono and multimodal configurations had a comparatively low error rate for incorrect commands and target misses, suggesting that the

[1] Note these results are not corrected for alpha, see Notes Discussion N2. Multiple comparisons in this work in the Appendices for a discussion on multiple comparisons. Overall mean times shown for task time, non-parametric statistical comparisons used raw task time data.

device pointing accuracy allowed accurate target selection when required irrespective of the selection method used, although the device did show a moderate rate of control corrections, indicating some difficulty in manoeuvring and positioning the cursor, possibly due to the mass of the head resulting in some loss of fine cursor control or cursor overshoot of the target.

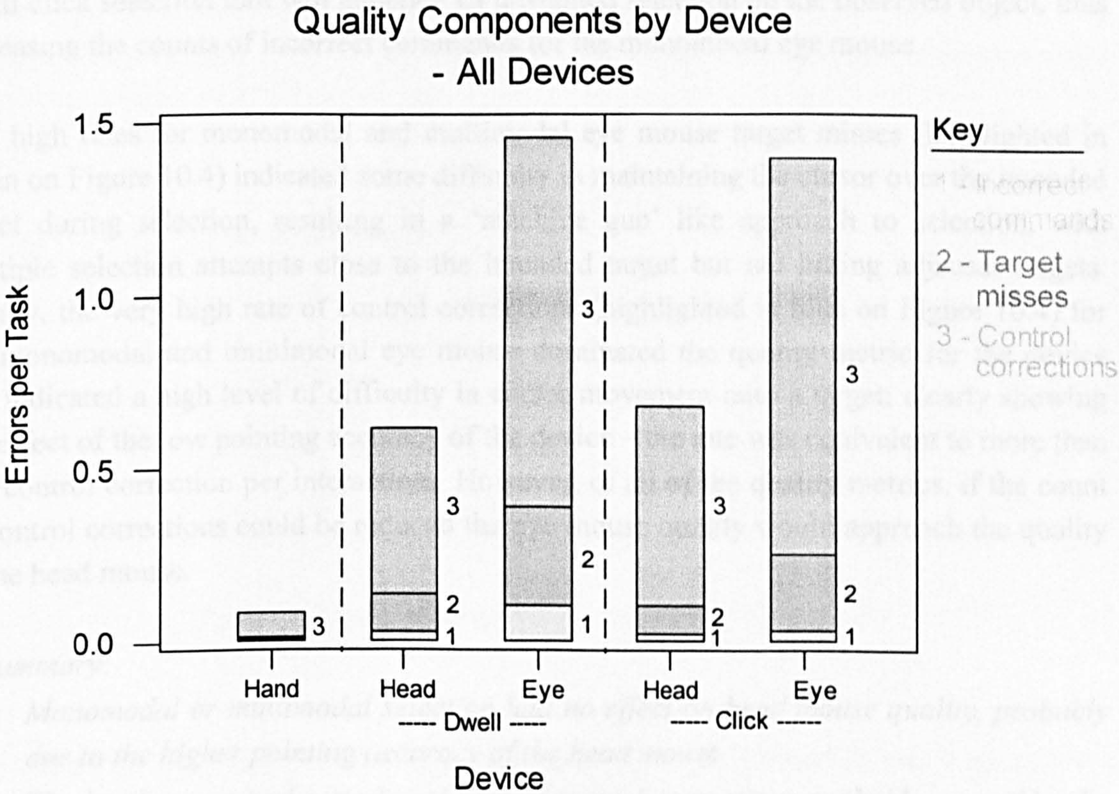


Figure 10.4 Composition of device task quality[1]

The analysis revealed that the eye mouse in multimodal configuration showed a low error rate for incorrect commands, suggesting target selection accuracy could be achieved when necessary. This was not so for the monomodal eye mouse which had a higher rate of incorrect commands than any other device (highlighted in orange on Figure 10.4). This suggested that the use of monomodal dwell click target selection greatly increased the chances of an inadvertent incorrect command being generated, probably as the device had low pointing accuracy and the test subject had less control over the exact timing of the target selection than with a multimodal switch operated selection device, and so could not

[1] The elements of task quality as shown on the graph are defined in Chapter 5.

prevent incorrect targets being accidentally selected. This selection of incorrect targets by the monomodal eye mouse can be characterised as a 'Midas touch', or perhaps 'Midas gaze' (as discussed in Chapter 2), where the gaze of the test subject may briefly alight on a target of interest on the interface, such as an object giving feedback or an object attracting the attention of the test subject. Whilst the gaze point is on the target, the cursor is also on the target as it is driven by the gaze position and this gives rise to the possibility that the dwell click selection tool will generate an unwanted selection on the observed object, thus increasing the counts of incorrect commands for the monomodal eye mouse.

The high rates for monomodal and multimodal eye mouse target misses (highlighted in green on Figure 10.4) indicated some difficulty in maintaining the cursor over the intended target during selection, resulting in a 'machine gun' like approach to selection, with multiple selection attempts close to the intended target but not hitting adjacent targets. Finally, the very high rate of control corrections (highlighted in blue on Figure 10.4) for the monomodal and multimodal eye mouse dominated the quality metric for the device and indicated a high level of difficulty in cursor movement onto a target; clearly showing the effect of the low pointing accuracy of the device – the rate was equivalent to more than one control correction per interaction. However, of all of the quality metrics, if the count of control corrections could be reduced the eye mouse quality would approach the quality of the head mouse.

In summary:

- *Monomodal or multimodal selection had no effect on head mouse quality, probably due to the higher pointing accuracy of the head mouse*
- *The head mouse had a moderate rate of control corrections probably caused by the mass of the head causing overshoot and subsequent compensation in pointing*
- *The monomodal eye mouse had the highest counts of incorrect commands, probably due to the 'Midas touch property of the device and the combination of lower pointing accuracy and uncertainty in monomodal selection timing*
- *The multimodal eye mouse had lower counts of incorrect commands and target misses than the monomodal eye mouse due to the selection timing accuracy of the multimodal switch device*
- *Both eye mice had high counts of control corrections, indicating poor pointing accuracy*
- *Of all metrics, if the number of eye mouse control corrections could be reduced it would approach the quality of the head mouse*

10.9 Device satisfaction

The overall perceived test subject satisfaction with each of the devices was assessed after each test session with a device, and calculated by pooling the individual section and category results (as described in Chapter 5) with the statistical significance of any differences determined (Wilcoxon matched pairs signed rank tests, Appendices Table A10.11), (Table 10.6). Ranking the satisfaction ratings for the devices showed that the hand mouse was judged the most satisfying to use with a high rating of 6.2/7.0, followed by the multimodal then monomodal head mouse, with the multimodal and monomodal eye mice being rated least satisfying to use.

Device satisfaction		
Device	*Satisfaction (1-7)*	*Device rank*
Hand	6.20	1
Head Dwell	4.36 [a]	3
Eye Dwell	2.93 [b]	5
Head Click	4.73	2
Eye Click	3.90 [a, b]	4

Data with the same letter are not significantly different ($p > 0.05$)

Table 10.6 Device satisfaction ratings and rankings[1]

The validity of the satisfaction-rating questionnaire (Chapters 5 and 8) was supported by the device efficiency results with the same device rankings for satisfaction and efficiency for the devices. Comparing the efficiency and satisfaction results further for range and ratios of results found that the satisfaction questionnaire preserved the range and ratios found in the efficiency results. Taking the highest rated device (hand mouse) at 6.2 satisfaction rating and 83.3% efficiency, and the lowest rated device (monomodal eye mouse) at 2.9 satisfaction rating and 42.9% efficiency and examining the differences showed a difference of 46% between the devices on the satisfaction questionnaire scale and a difference of 51.5% between the devices on the efficiency results. The similarity between the differences on the scales indicated that the satisfaction scale closely followed the efficiency results, and also had sufficient range to measure a wide range of device performances. These findings further validated the satisfaction questionnaire.

[1] Note these results are not corrected for alpha, see Notes Discussion N2. Multiple comparisons in this work in the Appendices for a discussion on multiple comparisons. Overall mean satisfaction ratings shown, non-parametric statistical comparisons used raw satisfaction rating data

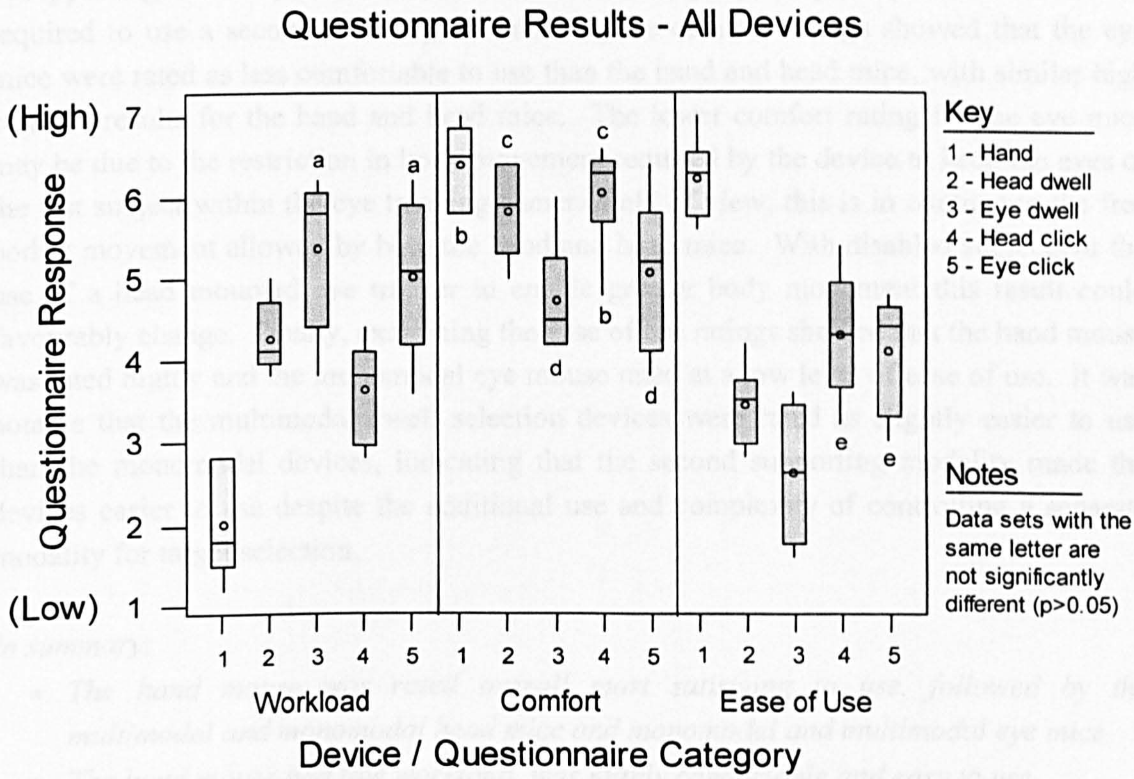


Figure 10.5 Device satisfaction questionnaire results[1]

Examining the individual satisfaction questionnaire sections for the devices (workload, comfort and ease of use, from Chapter 5) suggested differences and trends in perceived workload, comfort and ease of use for the devices (Figure 10.5). Wilcoxon matched pairs signed rank tests were used to determine the statistical significance of any differences between devices within each satisfaction category (Appendices Tables A10.12 to A10.14). Note that greater satisfaction is shown by lower workload ratings but higher comfort and ease of use ratings. The workload questionnaire results showed that the baseline hand mouse had, as expected, the lowest rated workload. The head mice had higher workload than the hand mouse with the eye mice having the highest workload, probably due to the poor pointing performance of the device. Such relatively poor subjective workload and comfort results may have indicated some of the reasons why the eye mouse has not been widely accepted as a viable device. The monomodal devices tended to have slightly higher workload than the multimodal devices when comparing like devices, indicating that

[1] Note that greater satisfaction is shown by lower workload ratings but higher comfort and ease of use ratings.

a supporting modality could decrease workload slightly despite the additional work required to use a second modality. Examining the comfort ratings showed that the eye mice were rated as less comfortable to use than the hand and head mice, with similar high comfort results for the hand and head mice. The lower comfort rating for the eye mice may be due to the restriction in body movement required by the device to keep the eyes of the test subject within the eye tracking camera field of view; this is in contrast to the free bodily movement allowed by both the hand and head mice. With disabled subjects or the use of a head mounted eye tracker to enable greater body movement this result could favourably change. Finally, examining the ease of use ratings showed that the hand mouse was rated highly and the monomodal eye mouse rated at a low level of ease of use. It was notable that the multimodal dwell selection devices were rated as slightly easier to use than the monomodal devices, indicating that the second supporting modality made the devices easier to use despite the additional use and complexity of controlling a separate modality for target selection.

In summary:

- *The hand mouse was rated overall most satisfying to use, followed by the multimodal and monomodal head mice and monomodal and multimodal eye mice*
- *The hand mouse had low workload, was highly comfortable and easy to use*
- *The head mouse had greater workload, was comfortable but difficult to use*
- *The eye mouse had very high workload, low comfort and was difficult to use*
- *The use of a supporting modality decreased workload within devices, increased ease of use but did not change comfort*
- *If the eye mouse workload could be reduced and comfort increased it could approach the head mouse satisfaction.*

10.10 Individual satisfaction factors

The results of the individual satisfaction factors that made up the total satisfaction categories and overall satisfaction ratings (Chapter 5) were calculated as averages of the six test subject ratings (Table 10.7). To aid identification of factors that were rated poorly and were particularly of interest, the two lowest satisfaction results for each satisfaction factor (such as 'physical workload' or 'eye discomfort' for example) are shown in bold type. Examining the individual factors for workload showed that the eye mouse exhibited consistently higher workload than the head mouse for all factors. This can be attributed to the requirement for the test subject to remain fairly immobile and within the field of view of the eye gaze-tracking camera. This difference was particularly noticeable for mental workload, indicating a high degree of concentration was required for the eye mouse, with the monomodal eye mouse having the highest mental workload probably due to the additional effort required in timing the dwell selection timing. Frustration was higher for the eye mouse than the head mouse, probably due to the lower pointing accuracy of the eye mouse, with the monomodal eye mouse having the highest frustration, probably due to the higher counts of errors found in the quality metrics for the device. However, it was notable that there was a considerably smaller difference between the devices for performance, indicating that the test subject felt that although the workload was high for the eye mouse, it approached the performance of the head mouse.

Individual satisfaction factors						
Factor / Device			*Monomodal*		*Multimodal*	
		Hand	*Head Mouse*	*Eye Mouse*	*Head Mouse*	*Eye Mouse*
Workload (low=good)	Physical	1.8	4.5	**5.3**	3.8	**5.5**
	Mental	2.3	4.5	**6.6**	3.8	**5.7**
	Temporal	2.1	3.5	**3.6**	2.7	**4.3**
	Frustration	1.8	4.5	**6.1**	3.7	**5.0**
	Performance	1.8	4.1	**5.1**	4.2	**4.7**
Comfort (high=good)	Headache	6.5	6.1	**5.0**	6.5	**5.5**
	Eye	6.1	6.0	**3.6**	6.2	**4.7**
	Facial	6.5	6.1	**5.0**	6.2	**5.0**
	Mouth	6.8	6.5	**6.0**	6.8	**6.3**
	Neck	6.0	4.3	**4.0**	4.6	**3.8**
Ease of Use (high=good)	Pointing Accuracy	6.3	3.5	**1.8**	3.8	**2.1**
	Pointing Speed	6.0	**3.1**	3.8	**3.5**	4.5
	Clicking Accuracy	6.3	**3.3**	**2.3**	4.5	4.5
	Clicking Speed	6.5	**2.8**	**2.3**	4.5	5.1
	System Control	6.0	4.3	**2.6**	5.2	**4.1**

Table 10.7 Individual factors of device satisfaction[1]

Examining the individual factors for comfort showed consistently poor ratings for the eye mouse, this was particularly notable for eye discomfort. There was a small reduction in difference between the devices for neck discomfort due to a low rating for the head mouse, indicating that the head mouse did cause some neck discomfort in operation. Taken together the individual workload and comfort ratings confirmed that the eye mouse caused considerable workload and discomfort during operation.

[1] The two lowest satisfaction results for each satisfaction factor are shown in **bold type.**

Finally examination of the ease of use ratings revealed the operational properties of the devices. The multimodal and monomodal head mice were rated as having superior pointing accuracy and overall control than the eye mice, probably due to the higher pointing accuracy of the head mouse. The monomodal and multimodal eye mice showed superior pointing speed perception than the head mouse probably due to the rapid movement of the eye, but were not perceived as rapid as the hand mouse. This was a curious result, but it is possible that subjects combined pointing accuracy and pointing speed perception. Both multimodal devices were rated highly for clicking speed and clicking accuracy, probably again due to the control afforded by the supporting modality over the selection timing compared to the dwell click software device.

In summary:

- *The eye mice had consistently high ratings for all workload factors*
- *The monomodal eye mouse had the highest mental workload and frustration, probably due to the combination of pointing inaccuracy and dwell selection*
- *The eye mouse was most uncomfortable to use, with a high rating for eye discomfort.*
- *The head mice were rated with higher pointing accuracy and control than the eye mouse*
- *The eye mice were rated as having higher pointing speed than the head mouse*
- *The multimodal devices had higher clicking speed and accuracy.*

10.11 A summary of the performance hand, head and eye mice

The performances of the head and eye mice were characterised as the head mouse having slower but more accurate pointing and the eye mouse more rapid but less accurate pointing. Pointing accuracy had a strong influence on eye mouse performance but little influence on head mouse performance. In contrast, pointing speed had a strong influence on head mouse performance but less influence on eye mouse performance. Multimodal selection outperformed monomodal selection for all devices. The hand mouse was rated most satisfying to use, followed by the monomodal and multimodal head mice and then the monomodal and multimodal eye mice. The results showed that interaction was feasible with all devices, although the performance of the eye-based devices was poor in comparison with the performance of a hand mouse. The main factors influencing the performance of the head and eye mice were found to be pointing speed and pointing accuracy. Neither the head nor eye based devices achieved the same performance as the hand mouse baseline, clearly showing the penalty users of these devices incur in comparison to hand mouse users.

Chapter 11

A Detailed Examination of Eye and Head Mouse Performance

The aim of this chapter is to continue the investigation of the performance of head and eye mice in comparison to a hand mouse baseline, with the aim of identifying factors that most influence the performance of head and eye mice, and from these determine how the performance of head and eye based pointing may be enhanced.

This chapter builds on the work of the previous chapter by using the experimental results from Chapter 10. The chapter first examines the effect of target size on the performance of the devices, and then examines how interaction technique influences performance, and finally how test subject experience with the devices influences performance. The chapter then examines how the effect of both target size and experience with the devices combine to influence, and maximise performance. Finally, the chapter ends with the prediction of a method of performance enhancement that can be applied to head, and particularly eye mice, that would greatly enhance the performance of these devices, and bring their performance closer to that of the hand mouse baseline.

11.1 Target size and device efficiency

From the results in Chapter 10, it was found that device pointing speed and pointing accuracy had a great effect on the performance of the head and eye mice respectively. It is clear that, in general, the effects of pointing speed will be influenced by the distance travelled to a target on the interface, and the effects of pointing accuracy will be influenced by the size of targets on that interface. Distance is determined by the nature of the tasks and the placement of the various target objects on the interface, and can be regarded as essentially random within the bounds of the interface. However, the target sizes on the interface are of certain defined sizes as determined by the survey of the interface objects in Chapter 4. In addition, as discussed in Chapter 3, both target distance and target size can influence pointing performance (as defined in Fitts' Law, Chapter 3). Since target distance is essentially random and cannot easily be defined, and target size may be clearly defined, an investigation into the detailed effects of target size on device pointing speed and pointing accuracy under 'real world' conditions was required.

The construction of the test method allowed this detailed examination of the performance of the devices with differing target sizes. Breaking down the data by the four target size categories in the test method (Chapter 4: 0.3°, 0.6°, 0.9°, 1.2° subtended visual arc at 60cm) suggested a relationship between device efficiency and target size (Appendices Table A11.1, Graphed Figure 11.1), with efficiency increasing with target size.

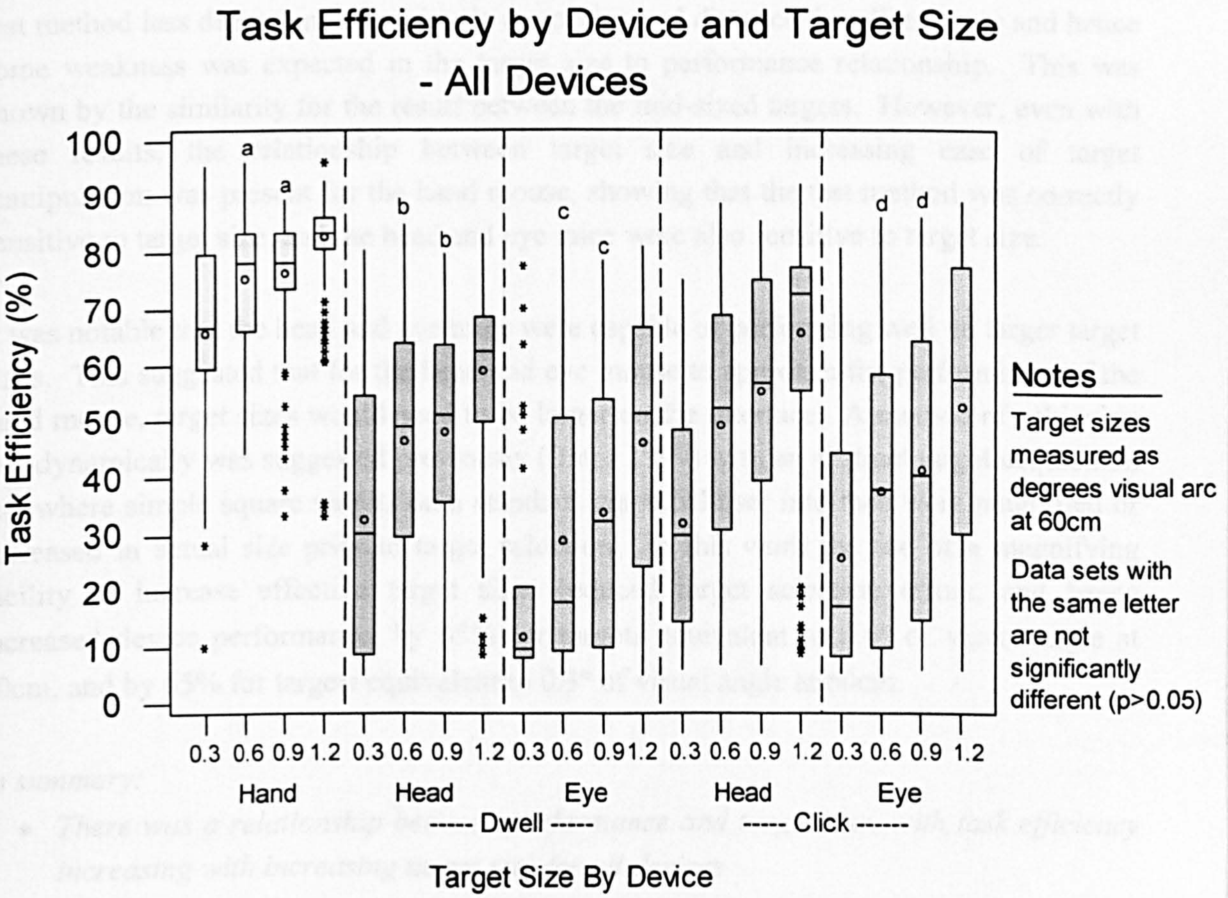


Figure 11.1 Hand, head and eye mouse device task efficiency by target size

The statistical significance of differences between the results within each device were calculated (Mann-Whitney two-sample rank test, Appendices Table A11.2). There was a progression of efficiency with increasing target size, from the smallest target size, S1 = 0.3°, having the lowest efficiency to the largest target size, S4 = 1.2°, having the greatest efficiency. Although the smallest and largest targets for all devices had efficiencies significantly different from each other, the two middle-sized targets, except the multimodal head mouse, did not. However, there was a progression of increasing efficiency from S1 through S2/3 to S4 (Figure 11.1). The relationship between target size and efficiency was expected, as the relationship between increasing target size and increasing ease of target acquisition has been shown previously with Fitts-Law (Fitts 1954) based work (for a review see Chapter 3). However, unlike Fitts Law that uses target size and distance travelled to the target with simple target acquisition tests of low cognitive load, the test method uses defined target sizes but with essentially random distances, and with tasks containing a moderate degree of cognitive load. This makes the test method less dependent upon simple target size and distance than Fitts' Law and hence some weakness was expected in the target size to performance relationship. This was shown by the similarity for the result between the mid-sized targets. However, even with these results, the relationship between target size and increasing ease of target manipulation was present for the hand mouse, showing that the test method was correctly sensitive to target size, and the head and eye mice were also sensitive to target size.

It was notable that the head and eye mice were capable of performing well on larger target sizes. This suggested that for the head and eye mouse to approach the performance of the hand mouse, target sizes would need to be larger on the interface. A method of achieving this dynamically was suggested previously (Bates 1999) with an abstract target acquisition test where simple square targets on a standard graphical user interface were magnified or increased in actual size prior to target selection. In this work the use of a magnifying facility to increase effective target sizes reduced target selection errors, and hence increased device performance, by 45% for targets equivalent to 0.6° of visual angle at 60cm, and by 65% for targets equivalent to 0.3° of visual angle at 60cm.

In summary:

- *There was a relationship between performance and target size, with task efficiency increasing with increasing target size for all devices*
- *Increasing target sizes by interface design or artificially via a soft device should increase device performance*

11.2 Target size and device quality

Further investigation into the effects of target size on the components of device task quality was required to understand the potential of increasing the task quality performance of the assistive technology devices by modifying or changing target size. The test method allowed the breakdown of the quality components of the method by the four target size categories (Appendices Table A11.3, Graphed Figure 11.2). Examining the task quality results showed a steady decrease in errors, and hence increases in quality, for increasing target size for all devices. It was notable that all devices, except for the monomodal eye mouse, had low counts of incorrect commands, even for the smallest target sizes, further supporting the suggestion that the devices could be accurate when the consequences of error, such as correcting the outcome of an incorrect command, were high. The higher incorrect command counts for the monomodal eye mouse again illustrated the lack of control of the device when inaccurate pointing was combined with inaccurate selection timing.

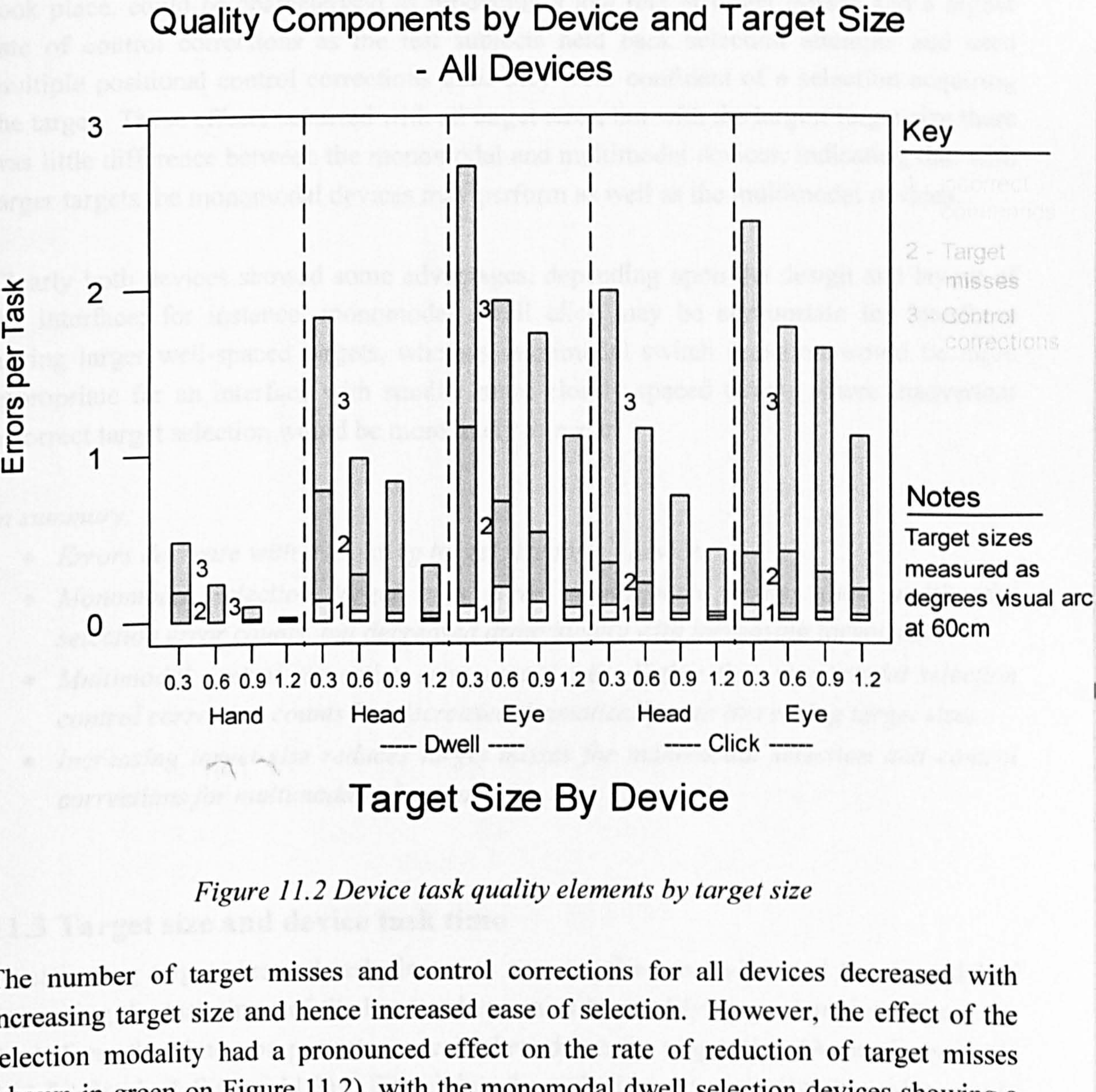


Figure 11.2 Device task quality elements by target size

The number of target misses and control corrections for all devices decreased with increasing target size and hence increased ease of selection. However, the effect of the selection modality had a pronounced effect on the rate of reduction of target misses (shown in green on Figure 11.2), with the monomodal dwell selection devices showing a far more pronounced reduction in the counts of target misses with increasing target size than the multimodal selection devices. In contrast, the counts of control corrections were higher with the multimodal selection devices than with the monomodal dwell selection devices. Taken together these results suggested that the monomodal devices could be characterised as producing more selection attempts earlier in the target selection process (a 'machine gun' effect caused by the dwell click soft device producing inadvertent clicks) that caused higher counts of target misses (shown in green on Figure 11.2) but also reduced the counts of control corrections (shown in blue on Figure 11.2) as the target tended to be selected by one of these many dwell clicks early in the interaction. However, the multimodal devices, where the test subject had full control over when the selection took place, could be characterised as producing a low rate of target misses and a higher rate of control corrections as the test subjects held back selection attempts and used multiple positional control corrections until they were confident of a selection acquiring the target. These effects occurred with all target sizes, but with the largest target size there was little difference between the monomodal and multimodal devices, indicating that with larger targets the monomodal devices may perform as well as the multimodal devices.

Clearly both devices showed some advantages, depending upon the design and layout of the interface; for instance, monomodal dwell click may be appropriate for interfaces having larger well-spaced targets, whereas multimodal switch selection would be more appropriate for an interface with smaller more closely spaced targets where inadvertent incorrect target selection would be more likely to occur.

In summary:

- *Errors decrease with increasing target size for all devices*
- *Monomodal selection target miss error counts were higher than multimodal selection error counts, but decreased dramatically with increasing target sizes*
- *Multimodal control correction error counts were higher than monomodal selection control correction counts but decreased dramatically with increasing target sizes*
- *Increasing target size reduces target misses for monomodal selection and control corrections for multimodal selection*

11.3 Target size and device task time

Investigating target size and task time was now required to understand the potential of decreasing the task times of the head and eye mice by modifying or changing target size. As before, the time components were broken down by target size (Appendices Table A11.5, Graphed Figure 11.3). Examining the task time components showed a steady decrease in the time used by non-productive actions, and hence decreases in task time, with increasing target size for all devices. Here, as with the quality components, the proportion of task time used by generating incorrect commands was low in proportion to the total task time, even for the smallest target size, confirming that the devices could be accurate when the consequences of error were high. However, unlike the quality components (Figure 11.2), the monomodal eye mouse results for incorrect commands did not differ substantially from the results for the other devices. Although the monomodal eye mouse did generate more incorrect command quality counts than the other devices, it did so with little additional task time in comparison to the other devices, confirming that the device tended to generate a 'machine gun' approach to selection, generating excessive incorrect commands but in a short space of time.

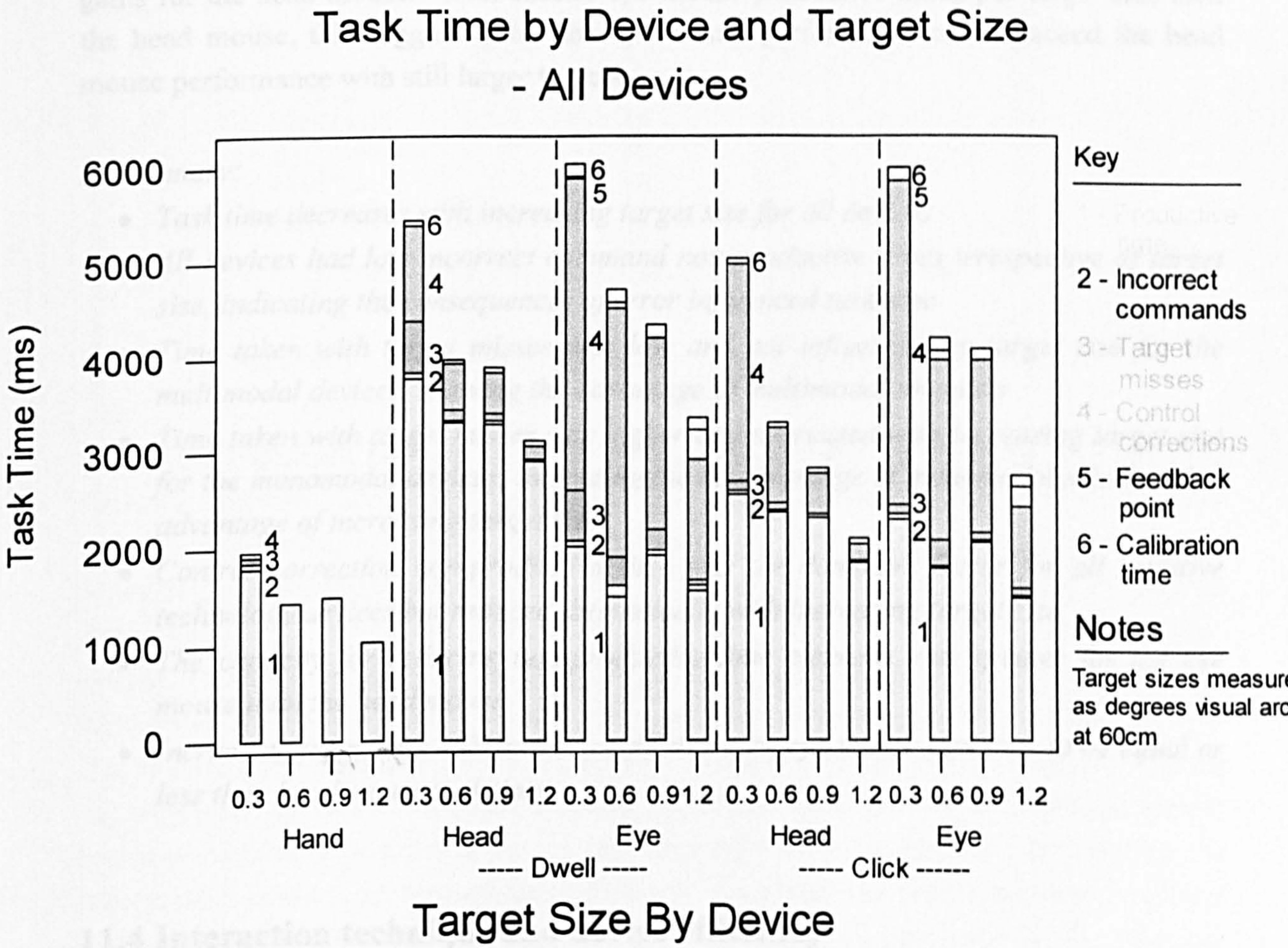


Figure 11.3 Device task time elements by target size

The time lost in positional control corrections (shown in blue on Figure 11.2) reduced dramatically with increasing target size for all devices and was the dominant non-productive time factor. This was particularly evident for the eye mouse with control corrections exceeding the productive time, clearly characterising the device as having inaccurate pointing. It was notable that the eye mouse had longer control correction times and shorter productive times for each target size than the head mouse. This confirmed that if the non-productive times, and in particular the control correction times of the eye mouse could be reduced it could equal or exceed the performance of the head mouse, irrespective of target size. Examining the non-productive task time elements for the larger target sizes showed that the head mouse had few non-productive task time elements that could be reduced whereas the eye mouse still had appreciable non-productive times that could be reduced. This suggested that increased target size would give scope for further improvement of the non-productive time elements of the eye mouse but would give few gains for the head mouse. With shorter eye mouse productive times per target size than the head mouse, this suggested that the eye mouse performance could exceed the head mouse performance with still larger targets.

In summary:

- *Task time decreases with increasing target size for all devices*
- *All devices had low incorrect command non-productive times irrespective of target size, indicating the consequences of error influenced task time*
- *Time taken with target misses was low and not influenced by target size for the multimodal devices, showing the advantage of multimodal selection*
- *Time taken with target misses was higher and decreased with increasing target size for the monomodal devices, indicating the disadvantage of monomodal selection but advantage of increasing target size*
- *Control correction non-productive time was the dominant factor for all assistive technology devices but reduced dramatically with increasing target size*
- *The capacity for reducing non-productive time elements was greater for the eye mouse than the head mouse*
- *Increasing target size still further could decrease eye mouse task time to be equal or less than head mouse task time*

11.4 Interaction technique and device efficiency

The effects of the differing interaction techniques[1] (Chapter 4) present on the interface on the level of device efficiency were investigated to determine if the type of object manipulation had an effect on device efficiency. Task efficiency was broken down by interaction technique for each device (Appendices Table A11.5, summarised in Table 11.1) and the statistical significance of differences between all interaction types within each device were calculated (Mann-Whitney two-sample rank tests, Appendices Table A11.6). Examining the individual interaction techniques across the devices showed that neither the head or eye mouse devices approached the efficiency of the baseline hand mouse for any given interaction technique, although the general progression of efficiency reducing with the increased complexity of interaction technique followed the pattern of the hand mouse, with single clicks showing the highest efficiency, through double clicks to dragging showing the lowest efficiency.

Efficiency and interaction technique					
Device	*Interaction Technique Efficiency (%)*				
	-------- Restricted --------			*---- Unrestricted ----*	
	Single	*Double*	*Drag*	*Single*	*Drag*
Hand	81	76	65	83	61
Head Dwell	**48**	49	36	**60**	34
Eye Dwell	***47***	42	12	***45***	9
Head Click	**57**	50	31	**69**	40
Eye Click	**31**	30	27	**53**	17

(Larger differences within devices are highlighted in **bold**).

Table 11.1 Effect of interaction technique on device efficiency[1]

One main difference between the hand mouse baseline and the other devices emerged from the analysis. Unlike the hand mouse, the monomodal and multimodal head mouse and the multimodal eye mouse results showed statistically significant increases in the efficiency of unrestricted single clicks over restricted single clicks (highlighted in **bold** on Table 11.1), with increases in efficiency of 12.1% (p = 0.003) for the monomodal head mouse, 11.7% ($p < 0.001$) for the multimodal head mouse and 21.9% (p = 0.005) for the multimodal eye mouse (Mann-Whitney two-sample rank tests, Appendices Table A11.6). These differences suggested that the increased levels of freedom to move the cursor from the target during interaction (without losing control of the target) with unrestricted targets had a considerable positive effect on the efficiency of these devices. In contrast, the monomodal eye mouse showed no statistically significant difference between restricted and unrestricted clicks (p = 0.370) (highlighted in ***bold italic*** on Table 11.3). The low pointing accuracy of the device coupled with dwell click target selection may explain this.

[1] As described in Chapter 4, there are three basic forms of pointing device object manipulation on a Windows interface. These are a single click on an object, a double click on an object, and a drag of an object. These actions can be either *restricted*, where cursor movement is confined within the area of the object to retain control of that object, or *unrestricted* where cursor movement may move from the object and then return to the object and still retain control of that object. For example, a button object on the Windows interface has *unrestricted* manipulation as, provided the mouse button is held down, the cursor may be moved away from the button and returned without losing permanent control of the button. In contrast, a hypertext link has *restricted* manipulation, as any movement of the cursor away from the object with the mouse button held down will lose control of the object.

The low overall efficiency results for the monomodal eye mouse, for all of the interaction techniques, suggested that target selection was difficult. This difficulty of locating the cursor on the target was compounded by difficulty in timing a dwell click when on the target. Any benefit that could be gained from freedom of movement from and returning to unrestricted targets was probably lost in the overall difficulty of overcoming pointing accuracy and timing the target selection, with the restricted or unrestricted nature of the target having little influence on the interaction. The results showed that the for devices with low pointing accuracy but high selection timing accuracy, device task efficiency could be improved by using targets on an interface that allowed unrestricted manipulation.

In summary:

- *Task efficiency reduces with increasing complexity of object manipulation requirements*
- *The hand mouse task efficiency was unaffected by the restricted or unrestricted manipulation freedoms of interface objects*
- *The monomodal and multimodal head mouse and the multimodal eye mouse had higher performance on unrestricted manipulation objects than restricted objects due to the additional freedom of movement from the objects*
- *Interfaces designed for head and eye mice manipulation should use unrestricted in preference to restricted object manipulation characteristics*

11.5 Subject experience

Breaking down the test results by the three test subject experience ratings (Chapter 10) showed that there was a relationship between the experience rating of test participants with the devices (Table 10.1, Chapter 10, shows the numbers of hours of device usage accumulated by participants in each of the three experience categories) and their performance, with increasing experience resulting in increasing task performance (Appendices Table A11.7, Graphed Figure 11.4). The statistical significance of the results (Appendices Table A11.8) showed that the baseline hand mouse outperformed all of the other devices, even for the high experience (shown in pink on Figure 11.4) head and eye mouse test subjects. The eye mouse was inferior to the head mouse for the low (green) and medium (blue) experienced participants in both monomodal and multimodal configurations. However, the two devices achieved parity of performance within a selection modality; at 63.9% for the monomodal head mouse and 61.1% for the monomodal eye mouse, and for the multimodal devices at 73.0% for the multimodal head mouse and 73.5% for the multimodal eye mouse in the high experience (pink) participants.

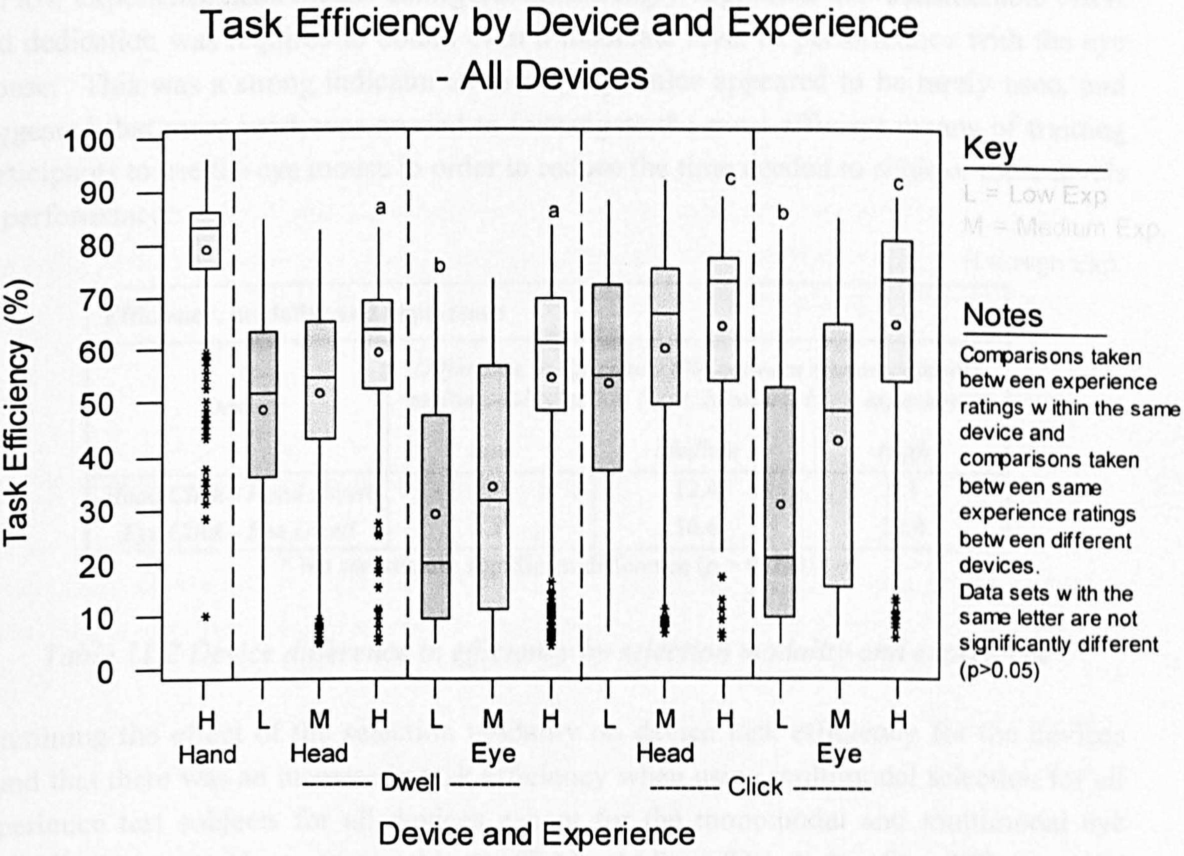


Figure 11.4 Device task efficiency by subject experience[1]

While the number of hours of device experience for the high experience participants is very different between head and eye mouse, the data suggested that the eye mouse in either monomodal or multimodal configuration could approach the performance levels of the head mouse for the same selection modality if participants were sufficiently practiced. (It is acknowledged that the performance of the head mouse is not known for experience levels extended to those of the high experience eye mouse users). The performance of the low experience group of head mouse users after 0.25 hours experience was far higher than the low experience eye mouse group with 1 to 2 hours experience and the medium experience eye mouse group with 6 to 8 hours experience. These long learning times coupled with the poor performance results for low and medium experienced participants for the eye mouse in both monomodal and multimodal configurations when compared to

[1] A little care is required with these results since there are only 2 test subjects in each of the head and eye experience categories, although there are 2 subjects * 150 test tasks = 300 test task samples in each group.

the low experience head mouse configurations strongly suggested that considerable effort and dedication was required to obtain even a moderate level of performance with the eye mouse. This was a strong indicator as to why eye mice appeared to be rarely used, and suggested that more work was needed to investigate the most efficient means of training participants to use the eye mouse in order to reduce the time needed to achieve these levels of performance.

Efficiency, modality and experience			
Device	*Difference in Efficiency (%) between monomodal and multimodal selection (Low, Medium, High experience)*		
	Low	*Medium*	*High*
Head Click - Head Dwell	3.4	12.4	9.1
Eye Click - Eye Dwell	1.5*	16.6	12.4

* No statistically significant difference ($p > 0.050$)

Table 11.2 Device difference in efficiency by selection modality and experience

Examining the effect of the selection modality on device task efficiency for the devices found that there was an increase in task efficiency when using multimodal selection for all experience test subjects for all devices except for the monomodal and multimodal eye mice for low experience test subjects (Table 11.2). The lack of a difference, or improvement, between monomodal and multimodal task efficiency for the low experience eye mouse test subjects, coupled with the very low performance of the eye mouse for low experience test subjects indicated that at this level of experience any advantage gained by an additional supporting selection modality was lost in the overall difficulty of using the device. It was notable that even for the high experience test subjects there was a significant difference (12.4%) between monomodal and multimodal performance, indicating that even when experienced with the monomodal dwell click software, it could not perform as well as using an additional multimodal supporting modality for selection. In order to further explain the effects of experience on device performance further investigation into the task time and task quality factors was required.

In summary:

- *Efficiency increased with increasing test subject experience*
- *The hand mouse outperformed all other devices even for high experience test subjects*
- *The head mouse outperformed the eye mouse in both monomodal and multimodal configurations for low and medium experience test subjects*

- *The head and eye mice achieved parity of performance for the high experience test subjects*
- *To achieve parity of performance, the eye mouse required considerably more experience than the head mouse*
- *Multimodal devices, except the low experience eye mouse, outperformed monomodal devices, indicating that an additional supporting modality improved performance*
- *The low experience eye mouse had exceptionally poor performance even after using the device for longer than medium experienced head mouse subjects, indicating a reason why eye mice are often rejected when first used*

11.6 Subject experience and device quality

Further investigation into the effects of test subject experience on the components of device task quality was required to understand the potential of increasing the task quality performance of the head, and more particularly eye mouse, devices by training or experience. The test results were broken down by the quality components of the test method (Chapter 5) and then grouped by the three test subject experience ratings (Chapter 10). Examining the results (Appendices Table A11.9, Graphed Figure 11.5) showed that the hand mouse had fewer errors than the other devices at any experience level. The breakdown of the individual quality components for increasing test subject experience showed steady decreases in error counts per task, and hence increases in quality, with increasing experience for both the head and eye mice.

The head mouse in both monomodal and multimodal configurations had fewer errors than the eye mouse for all experience group participants, except for the multimodal eye mouse. It was notable that the high experience eye mouse test subjects achieved a marked reduction in control corrections when compared to the eye mouse low and medium experience test subjects, indicating that the high experience test subjects had achieved good control over the eye mouse pointing inaccuracy. Further examination of the overall rate of reduction in total error counts per task showed a much more marked reduction in total error counts from low to high experience for the multimodal eye mouse (75.0%) than the other devices (Table 11.3). This rapid rate of reduction brought the multimodal eye mouse error count close to (within 10%) the multimodal head mouse.

It was notable that the overall proportions of each of the quality elements remained approximately constant as the overall count decreased with increasing participant experience, with control corrections dominating the quality metric irrespective of experience.

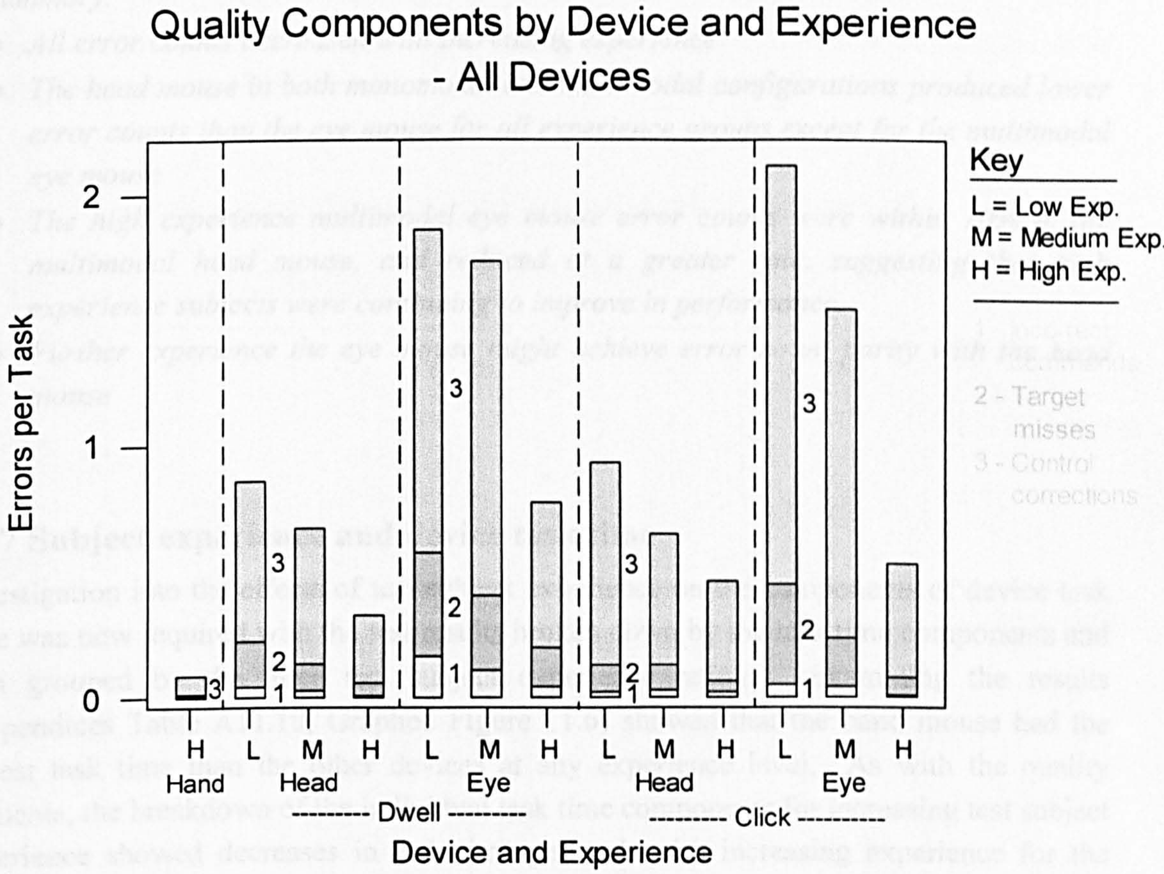


Figure 11.5 Device quality elements by subject experience

Experience and task quality	
Reduction in total task error counts (%) between Low and High test subject experience	
Device	*Reduction (%)*
Head Dwell: High - Low	62.4
Eye Dwell: High – Low	58.2
Head Click: High – Low	50.2
Eye Click: High - Low	**75.0**

(Larger differences within devices are highlighted in bold).

Table 11.3 Device subject experience and difference between task quality elements

In summary:

- *All error counts decreased with increasing experience*
- *The head mouse in both monomodal and multimodal configurations produced lower error counts than the eye mouse for all experience groups except for the multimodal eye mouse*
- *The high experience multimodal eye mouse error counts were within 10% of the multimodal head mouse, and reduced at a greater rate, suggesting that high experience subjects were continuing to improve in performance*
- *Further experience the eye mouse might achieve error count parity with the head mouse*

11.7 Subject experience and device task time

Investigation into the effects of test subject experience on the components of device task time was now required with the test results broken down by the task time components and then grouped by the three test subject experience ratings. Examining the results (Appendices Table A11.10, Graphed Figure 11.6) showed that the hand mouse had the lowest task time than the other devices at any experience level. As with the quality elements, the breakdown of the individual task time components for increasing test subject experience showed decreases in task time per task with increasing experience for the multimodal head and eye mice. In contrast with the quality results (Figure 11.5) there was no reduction in task time between low and medium experience test subjects for the monomodal devices, with a reduction only occurring for the high experience monomodal devices (Figure 11.6). In addition, the productive time (shown in orange on Figure 11.5) for the monomodal devices did not tend to reduce with increasing experience but remained approximately constant for the monomodal head mouse and actually increased for the monomodal eye mouse. These results suggested that the operation of the monomodal dwell click software was influencing the task times.

The long productive times for the monomodal head mouse in contrast to the multimodal head mouse suggested that test subjects used the higher pointing accuracy of the device to slowly move the cursor onto the target while waiting for the dwell click to occur. This care in pointing was shown in the low non-productive times for the monomodal head mouse. A similar effect was shown by the monomodal eye mouse, but here the lower pointing accuracy of the device was evident. Low and medium experience test subjects showed low productive times and high non-productive times, particularly control correction times, indicating a rapid movement to the target followed by many positional control corrections waiting for the dwell click to occur. In contrast high experience monomodal eye mouse test subjects showed a change in behaviour with a slower productive movement to the target and shorter non-productive control correction times waiting for the dwell click to occur, and indicating more control over the device.

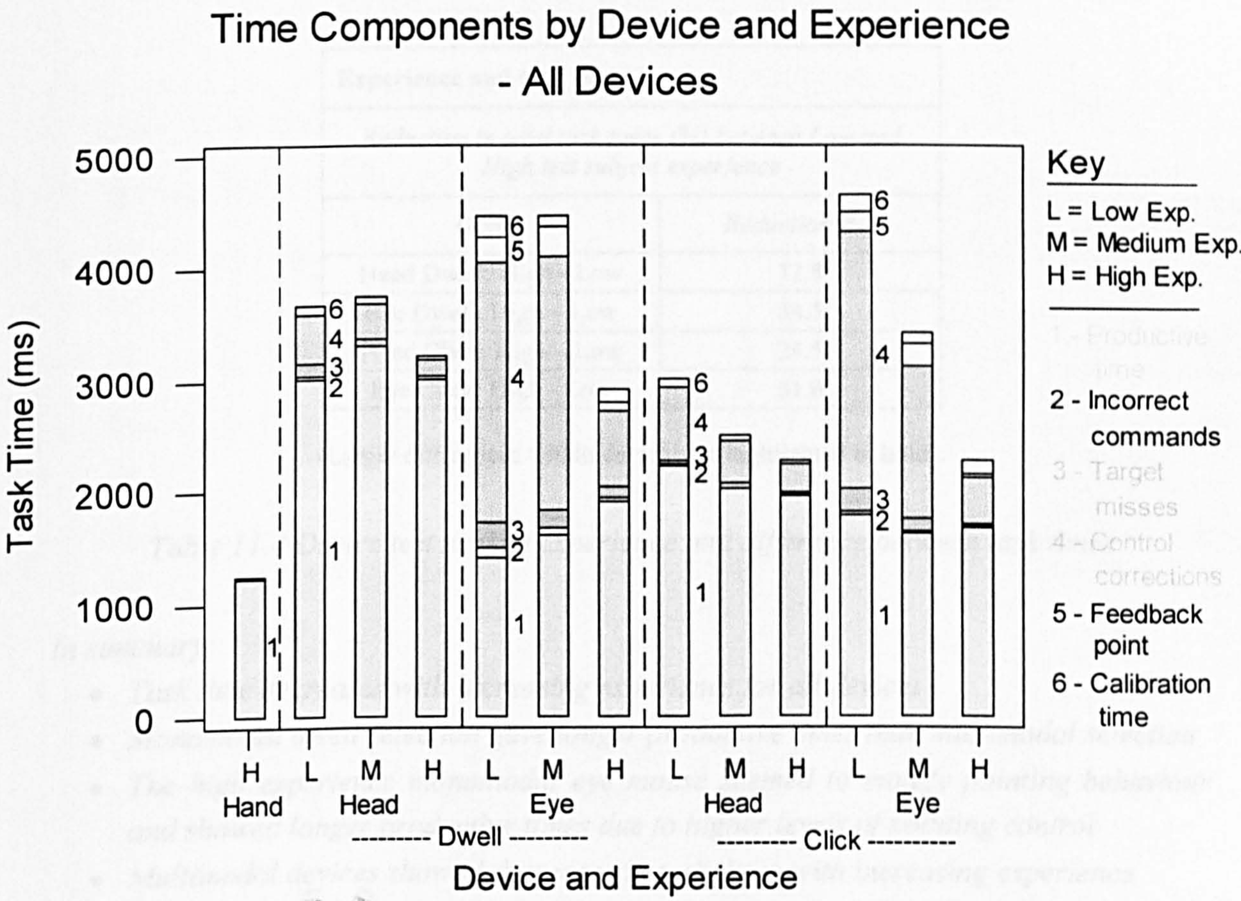


Figure 11.6 Device task time elements by subject experience

Examining the multimodal devices there were dramatic reductions in the non-productive elements of the eye mouse task time for highly experienced participants. Calibration time and time spent at the feedback point has been reduced significantly and the time lost in cursor control corrections reduced to near parity with the head mouse. Both sets of data showed a levelling-out of the productive time component with increasing experience, with the eye mouse having a lower productive time than the head mouse for all experience groups. Further examination of the overall rate of reduction in task times showed a marked reduction from low to high experience for the multimodal eye mouse (51.0%) than the other devices (Table 11.4). This rapid rate of reduction brought the multimodal eye mouse task time to parity (2273 ms vs. 2271 ms respectively, Appendices Table A11.11) with the multimodal head mouse. If the non-productive time elements of the eye mouse could be further reduced by increased test subject experience or other means then it would gain parity or have lower task times than the head mouse.

Experience and task time	
Reduction in total task times (%) between Low and High test subject experience	
Device	*Reduction (%)*
Head Dwell: High - Low	12.8
Eye Dwell: High – Low	34.5
Head Click: High – Low	24.5
Eye Click: High - Low	**51.0**

(Larger differences within devices are highlighted in **bold**).

Table 11.4 Device test subject experience and difference between task times

In summary:

- *Task time decreases with increasing experience for all devices*
- *Monomodal dwell selection gave longer productive times than multimodal selection*
- *The high experience monomodal eye mouse seemed to modify pointing behaviour and showed longer productive times due to higher levels of pointing control*
- *Multimodal devices showed decreases in task times with increasing experience*
- *The multimodal eye mouse had shorter productive times than the multimodal head mouse indicating that with further increases in experience the device might gain parity with the head mouse*

11.8 Target size and subject experience

The maximum achievable performance of the head and eye mice, based on the previous results, was investigated. Examining previous sections of the results showed that performance was related strongly to the target size of objects on the interface and to the experience of the test subjects with the devices, and to a lesser extent the number of modalities used to support target selection. Hence, in order to investigate the highest achievable performance of the head and eye mice within the data collected, the performance of the highly experienced participants only was separated from the data for all of the devices, including the baseline hand mouse, and broken down by selection modality and the target sizes of the objects on the interface (Appendices Table A11.12, Graphed Figure 11.7).

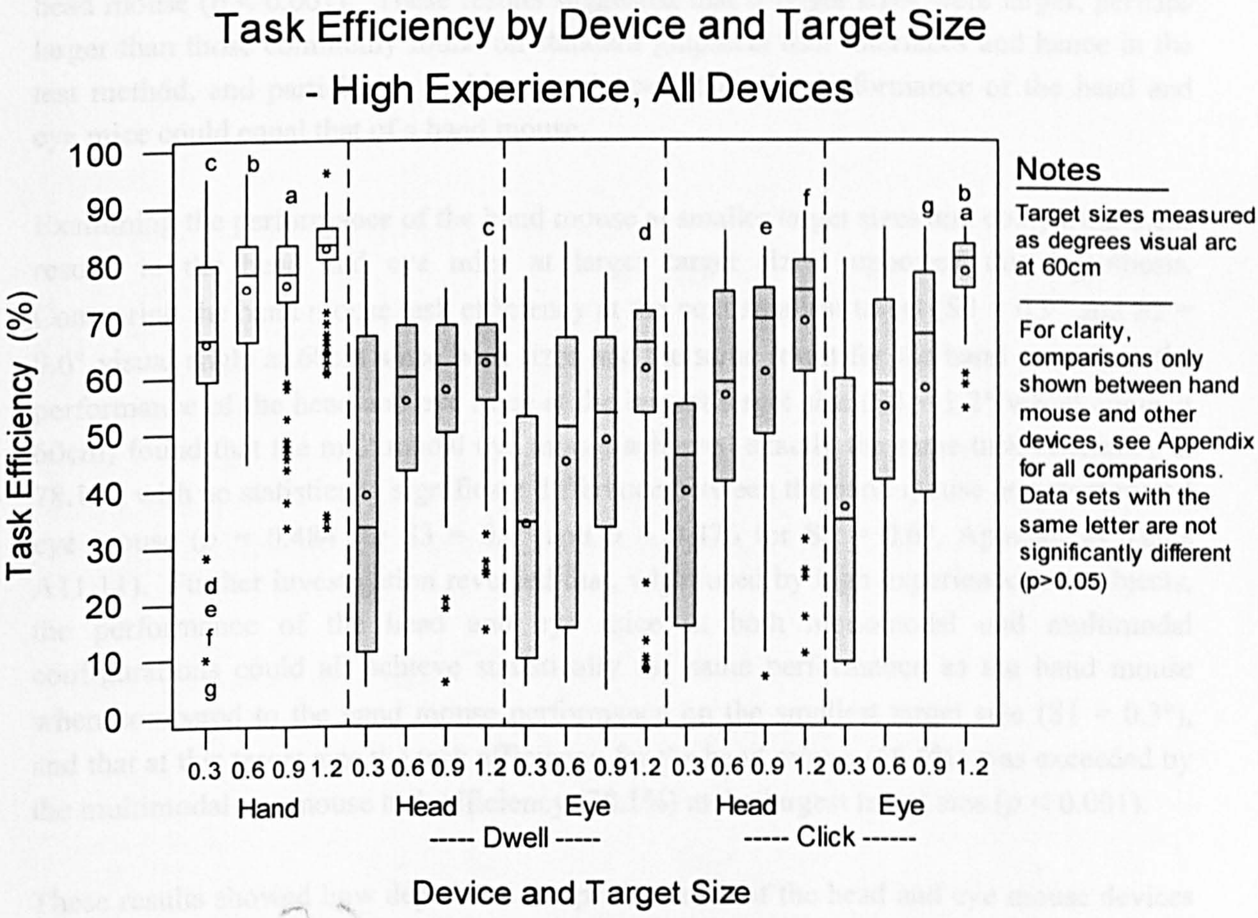


Figure 11.7 Device task efficiency for high experience subjects by target size

The statistical comparisons (Appendices Table A11.12) showed that the hand mouse again out-performed the head and eye mice, even when experienced participants used these devices. However, with the high experience test subjects, as the target size increased so the performance of the head mouse and eye mouse approached that of the baseline hand mouse. This was an important finding, and clearly indicated that target size coupled with user experience were the dominant factors influencing both head and eye mouse performance. Examining the largest target size (S4 = 1.2° visual angle at 60cm) the hand mouse had a task efficiency of 83.3% compared to the multimodal eye mouse at 78.1% and the multimodal head mouse at 75.2%. At this point the multimodal eye mouse exceeded the performance of the multimodal head mouse, this difference was statistically significant ($p < 0.001$, Table 2.17) and reached to within a few percentage points (5.2%) of the hand mouse at this target size. In addition, at this largest target size the monomodal eye mouse performance (65.9%) exceeded the performance (63.9%) of the monomodal head mouse ($p < 0.001$). These results suggested that if target sizes were larger, perhaps larger than those commonly found on standard graphical user interfaces and hence in the test method, and participants highly experienced, then the performance of the head and eye mice could equal that of a hand mouse.

Examining the performance of the hand mouse at smaller target sizes and comparing these results to the head and eye mice at larger target sizes supported this hypothesis. Comparing the hand mouse task efficiency at the next smallest target (S3 = 0.9° and S2 = 0.6° visual angle at 60cm since both sizes had the same result for the hand mouse) to the performance of the head and eye mice at the largest target size (S4 = 1.2° visual angle at 60cm) found that the multimodal eye mouse achieved exactly the same task efficiency at 78.1%, with no statistically significant difference between the hand mouse and multimodal eye mouse ($p = 0.484$ for S3 = 0.9° and $p = 0.476$ for S2 = 0.6°, Appendices Table A11.11). Further investigation revealed that, when used by high experience test subjects, the performance of the head and eye mice in both monomodal and multimodal configurations could all achieve statistically the same performance as the hand mouse when compared to the hand mouse performance on the smallest target size (S1 = 0.3°), and that at this target size the task efficiency for the hand mouse (66.6%) was exceeded by the multimodal eye mouse task efficiency (78.1%) at the largest target size ($p < 0.001$).

These results showed how dependent the performance of the head and eye mouse devices was on the size of the targets present on the interface, even for highly experienced test subjects, and showed that increasing target sizes on the interface coupled with sufficient experience could significantly increase the performance of the multimodal eye mouse to exceed that of the head mouse and approach the performance of the hand mouse. Since one of the aims of this work was to study and enhance manipulation on a standard graphical user interface, a custom interface with overly large object sizes would not be in the bounds of this work, nor necessarily desirable. However, as suggested earlier in this chapter, a method of temporarily magnifying target sizes could in principle be applied to a standard graphical interface.

In summary:

- *For high experience and the largest target size, the performance of the monomodal eye mouse exceeded the performance of the monomodal head mouse; and the* *performance of the multimodal eye mouse exceeded the performance of the multimodal head mouse*

- *For high experience and the largest target size the multimodal eye mouse approached the performance of the baseline hand mouse*
- *If target sizes were sufficiently large and test subjects sufficiently experienced, and a supporting modality available, then the head and eye mice can approach the performance of the hand mouse*

11.9 Satisfaction and subject experience

As the questionnaires were administered after a complete test had been completed, rather than after each task had been completed (which would be impractical) it is not possible to separate satisfaction by individual task properties, such as target size or interaction technique. However, the effect of test subject experience with the devices on the satisfaction questionnaire categories could be investigated, by comparing the ratings and rating differences for all subjects to the ratings of the high experience category subjects (Table 11.5). Note that the largest differences are shown in bold type.

High experience and satisfaction									
Device	*Satisfaction Questionnaire response (1-7)*								
	Workload (low=good)			*Comfort (high=good)*			*Ease of Use (high=good)*		
Experience	*All*	*High*	*Diff.*	*All*	*High*	*Diff.*	*All*	*High*	*Diff.*
Hand	1.8	1.8	-	6.5	6.5	-	6.3	6.3	-
Head Dwell	4.1	4.1	-	5.9	6.2	+0.3	3.5	3.5	-
Eye Dwell	5.8	4.2	**-1.0**	4.5	5.5	**+1.0**	2.6	3.5	**+0.9**
Head Click	3.8	3.7	-0.1	6.3	6.4	+0.1	4.3	4.6	+0.1
Eye Click	5.1	4.0	**-1.1**	5.2	5.9	**+0.7**	4.4	4.7	**+0.3**

(Larger differences within devices are highlighted in **bold**).

Table 11.5 Satisfaction factors for all subjects against high experience subjects

Comparing only the results for the head mouse in both monomodal and multimodal configurations revealed none or only small increases in satisfaction ratings between all subjects and just the experienced subjects, indicating that the device was very rapid to learn, and satisfaction with the head mouse did not improve appreciably with experience. This result supported the head mouse hours of experience groupings (Table 10.1, Chapter

10) by showing that subjects reached a steady state of satisfaction with the head mouse after only a very short time using the device, and that further experience with the device would be unlikely to greatly improve satisfaction or performance with the device. Examining the results for the eye mouse showed appreciable increases in satisfaction between all subjects and just the experienced subjects (highlighted in bold on Table 11.5), indicating that test subjects had still not achieved steady state ratings of satisfaction (or performance, from previous sections) with the eye mouse even after many hours of experience with the device (Table 10.1, Chapter 10). This suggested that further improvements in satisfaction (and almost certainly performance) could be gained by additional experience with the device.

Examining the *workload* satisfaction ratings in detail showed an improvement in rating for high experience subjects in workload for both monomodal and multimodal configurations of the eye mouse by approximately 1 one complete rating point (out of seven). This improvement brought the monomodal eye mouse to near parity in workload rating to the monomodal head mouse (0.1 rating point difference), and the multimodal eye mouse to near parity in workload rating to the multimodal head mouse (0.3 rating point difference) for high experience subjects. Clearly experience reduced the perception of workload for the eye mouse. The monomodal devices exhibited higher workload than the multimodal devices, suggesting that the use of the dwell click device produced more workload than a switch click device, even for experienced subjects.

There were also improvements in *comfort* ratings for high experience subjects using the eye mouse in both configurations although the eye mouse was still rated less comfortable to use than the head mouse. Clearly the eye mouse was uncomfortable to use, even with considerable subject experience. This was probably due to the restriction on test subject movement required by the device. It should be noted again that this rating could change with motor-disabled subjects.

Finally, the *ease of use* rating for the eye mouse showed an improvement in rating for both configurations of the eye mouse, with the monomodal eye mouse achieving parity in ease of use rating to the monomodal head mouse, and the multimodal eye mouse achieving near parity to the multimodal head mouse (0.1 rating point difference). The greater increase in monomodal eye mouse ease of use rating, in comparison to the multimodal eye mouse ease of use rating, may have been due to the low initial rating of the monomodal eye mouse. This was perhaps due to the initial unfamiliarity and difficulty of use of the dwell click device in comparison to the perhaps more familiar operation of the switch click device. Except for the comfort rating of the head mouse, none of the devices approached the satisfaction ratings of the hand mouse, even for experienced subjects. Some care should be taken with these results due to the small sample sizes for experienced participants; however they did offer indications into the effect of participant experience on the performance and satisfaction of the devices.

In summary:

- *High experience test subjects showed little improvement in head mouse satisfaction, indicating that the device had reached steady-state satisfaction*
- *High experience eye mouse test subjects showed increases in satisfaction in all categories and achieved near parity with the head mouse in workload and ease of use*
- *The eye mouse was still rated less comfortable to use than the head and hand mice, indicating that the device tended to be uncomfortable to use, although that this rating could change with motor-disabled subjects.*

11.10 A summary of the examination of eye and head mouse performance

Four clear factors that influenced device performance, and hence offered methods of improving and enhancing the performance of head and eye based pointing direct interaction on a standard graphical interface, were found. The first factor was the effective *target size* of objects on the interface. Here larger targets resulted in higher performance when using an inaccurate pointing device, but what was of note was the dramatic performance increase, and reduction in tasks times and error rates, that was evident with increasing target size.

The second factor was the use of a *supporting modality* or multimodal operation, when possible, for controlling object selection. This enhancement alone gave clear performance benefits in comparison to head and eye mouse monomodal operation by giving users control over the timing of selection attempts and target misses.

The third factor was the *type of interaction* required by objects on the interface. Due probably to the imprecise pointing of head, and particularly, eye mice (this inaccuracy being inherent in eye mice), unrestricted objects (that allowed the cursor to momentarily leave and re-enter the object during manipulation) had higher performance than restricted objects. However, the unrestricted or restricted nature of objects on the interface is a property of the interface, and allows little possibility for enhancement without deviating away from a standard interface.

The final factor was the performance improvements given by *user experience* with the devices, most notably the eye mice. Here experienced users showed dramatic improvements in eye mouse performance, with reduced error counts, reduced task times and increased satisfaction. Of note was the length of time taken to become proficient with the eye mice. To date, and to the knowledge of the author, there have been no other long-term studies undertaken where eye mouse users have been allowed to accumulate extensive experience with the device. This perhaps explains why eye mice have been anecdotally regarded as very difficult to use, quite simply, they require very long learning times that are not achieved by users who may abandon the device before they become proficient. Hence, a second method of enhancement would be to instigate some form of training or encouragement to persist with the devices until users become proficient.

Of all of the potential factors that could be enhanced, increasing target size was the area that could be exploited to increase efficiency by temporarily magnifying target sizes on the interface.

Chapter 12

Enhancing Eye and Head Mouse Performance

This chapter follows on from the results in Chapter 11 that hypothesised a target magnification enhancement for head and eye-based interaction. This chapter shows the construction of such a target magnification enhancement to the eye and head mouse systems detailed in Chapter 9. The chapter starts with a survey of previous methods of target object magnification and evaluates the advantages and disadvantages of these methods. The chapter then goes on to briefly show how a 'zoom screen' facility was developed for target magnification. Finally, the chapter illustrates the zoom screen in operation.

12.1 Methods of target magnification

A survey of previous work found two methods of 'zooming-in' or increasing the effective sizes of targets to aid object selection on a user interface. The first method was indirect zooming (Istance et. al 1996b, Lankford 2000), and the second direct zooming (Bolt 1981, Bates 1999, Bates and Istance 2002, Albinsson 2003). Indirect zooming typically involved the user pointing at, or close to, the object of interest with the region around the cursor being captured and then displayed in a magnified form in a separate window, with the interface remaining static. In contrast direct zooming involved the user again pointing at, or close to, the object of interest but in this case the complete interface, including the object of interest, was magnified as a whole dynamically at the location of the cursor.

Of the two instances of indirect zoom, the simplest (Lankford 2000) involved a basic 'zoom' window parked toward the lower right of the screen showing a fixed level of magnification of the interface around the region of the cursor (Figure 12.1). This magnified view was updated each time the user allowed a dwell 'zoom' selection, in the same manner as a dwell click would be generated (Chapter 2). A more complex approach (Istance et al. 1996b) involved the user first selecting a zoom option on an onscreen keyboard, with the next selection on the interface producing a magnified view, embedded within the keyboard, of the interface around the region of the cursor (Figure 12.2). There were a number of disadvantages with these systems. The first was that they both occupied and obscured space on the interface for the magnified views of the interface. Secondly, in both of these cases only a fixed level of zoom-in on the interface was allowed, with the user unable to generate subsequent magnification on smaller objects. In theory higher magnification could be achieved by the user pre-selecting a high magnification level on the on-screen keyboard (Figure 12.2) prior to the zoom selection, although, due to high magnification and inaccurate pointing, this could result in the target being missed altogether from the magnified view. Finally, they did not allow full direct manipulation of the interface, but only allow limited *indirect* object manipulation by translating operations on the magnified view to the original cursor location on the unmagnified interface. So, for example, dragging a magnified object was not possible outside of the boundaries of the somewhat confined magnified image.

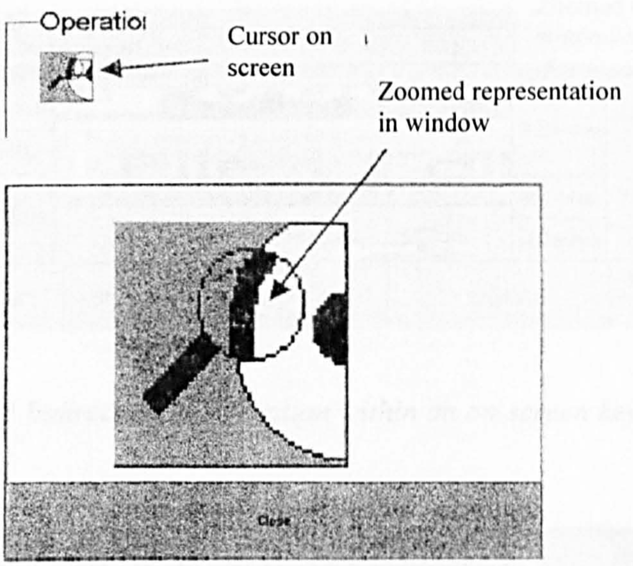


Figure 12.1 Indirect zoom operation[1]

Of the instances of direct zoom, all used basically the same approach of simply magnifying the image of the interface, centred at the location of the cursor. Of these, perhaps the first use of a 'zoom' facility in conjunction with eye-gaze tracking and any form of graphical or image-based interface was the 'Wall of Wonder' (Bolt 1981). Here a video wall of differing images (Figure 12.3) was presented to a viewer who would be automatically zoomed-in on any video stream, and hence could then watch that stream at full-screen, by gazing at a stream for greater than 5 seconds. The strength of this system was its use of direct interaction with the images rather than the more usual indirect interaction via a console of buttons for selection of video streams.

[1] From Lankford 2000.

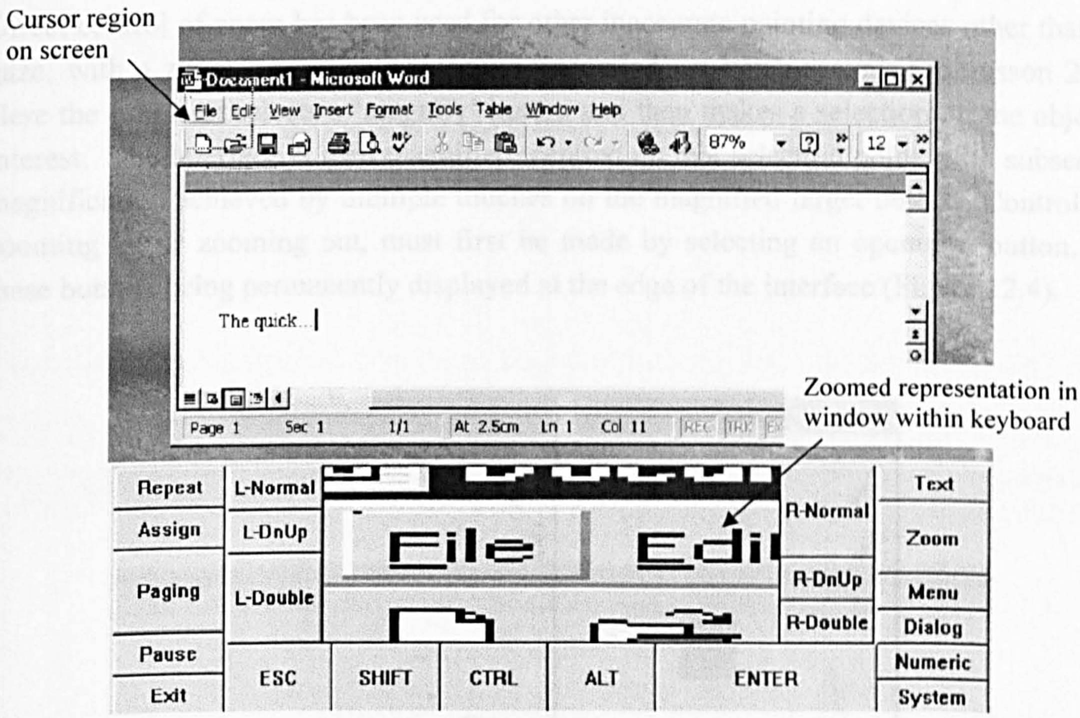


Figure 12.2 Indirect zoom operation within an on-screen keyboard[1]

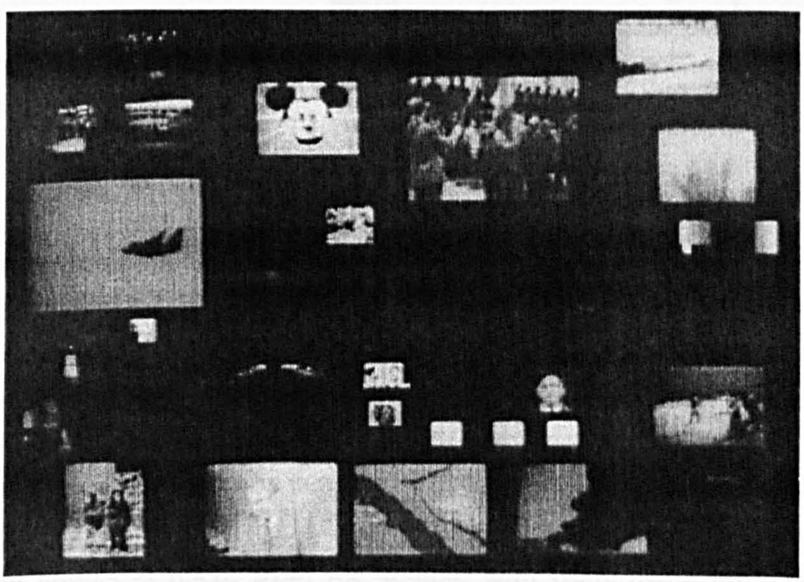


Figure 12.3 The 'Wall of Wonder' Zooming Interface[2]

[1] 'ECKey' keyboard, described in Istance et al 1996b.
[2] 'Wall of Wonder' from Bolt 1981.

Direct control of zoom has been used for other inaccurate pointing devices other than eye gaze, with a zoom magnification feature suggested for touch screens (Albinsson 2003). Here the user first selects a 'magnify' button and then makes a selection on the object of interest. The interface is then magnified, centred on that selection point, with subsequent magnification achieved by multiple touches on the magnified target object. Control over zooming in, or zooming out, must first be made by selecting an operation button, with these buttons being permanently displayed at the edge of the interface (Figure 12.4).

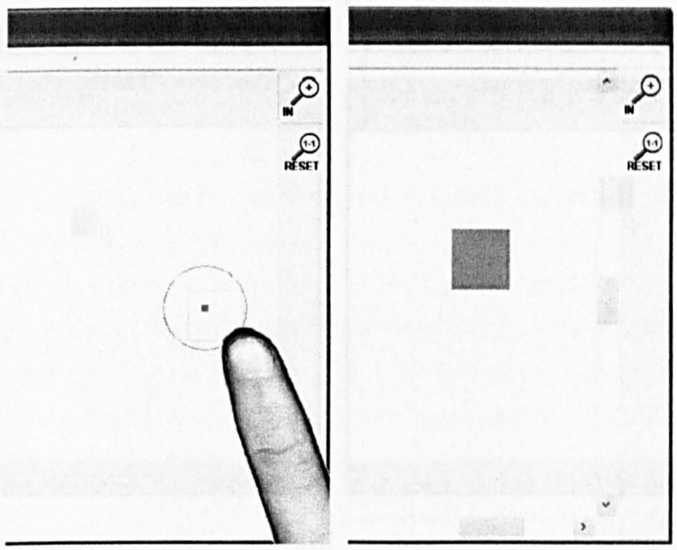
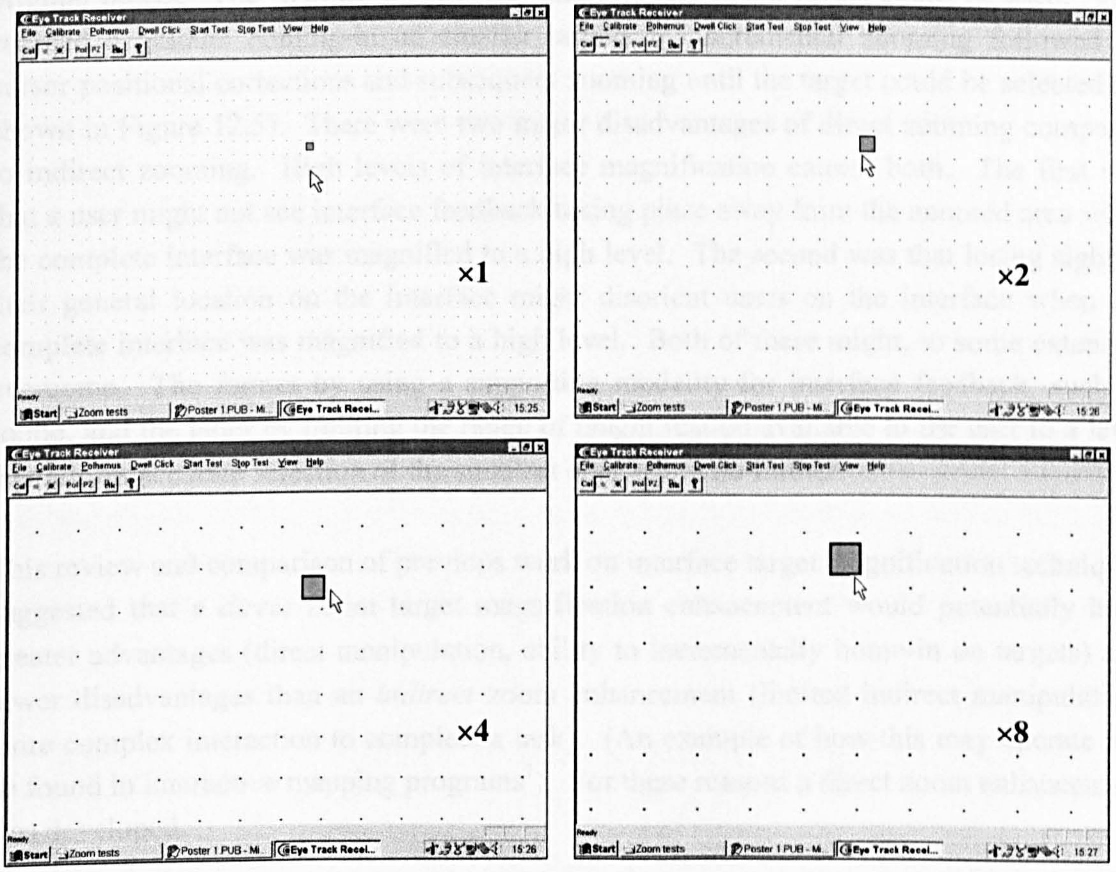


Figure 12.4 Direct zoom operation on a touch screen[1]

Perhaps the most in-depth investigation into eye-gaze interaction and direct control of interface magnification has been by the author (Bates 1999). Here a target acquisition test was performed with and without a zoom feature to aid object selection (Figure 12.5). Zoom level was controlled by raising or lowering shoulder position (a shoulder 'shrug') as an alternate means of control for users with high-level motor disabilities who may not be able to use a hand for magnification control, with target selection controlled by dwell click (in a similar manner to that described in Chapter 2). Here target sizes were 1.2°, 0.9°, 0.6° and 0.4° visual angles at 60cm from the screen, with trials using no zoom, and direct zoom. The results showed error rate reductions of 45% and 65% for the two smaller target sizes respectively for the direct zoom condition over the no zoom condition.

[1] From Albinsson 2003.



Note: Upper left illustration shows screen before zoom, upper right shows screen at x2 zoom, lower left at x4, and lower right at x8. Note the cursor remains the same size during zoom to maintain unobstructed interaction.

Figure 12.5 Direct zoom operation and eye-gaze target selection[1]

From the previous survey, comparing direct zooming to indirect zooming found three major advantages of direct zooming over indirect zooming. Firstly, the user could maintain direct interaction with the interface at all times. Thus they could carry out the same complex manipulations, such as dragging, that could be carried out when the interface was not magnified. Secondly, interaction required fewer steps, with direct zooming requiring the user to simply point at the object of interest and select or wait for a zoom, compared to indirect zoom where the user must first select the object area to be zoomed, wait for a zoom, and then interact with the magnified representation of the

[1] From Bates 1999.

original object. The final advantage was that multiple zoom levels could be used. This enabled a gradual homing-in on smaller targets by incremental zooming followed by cursor positional corrections and subsequent zooming until the target could be selected (as shown in Figure 12.5). There were two major disadvantages of direct zooming compared to indirect zooming. High levels of interface magnification caused both. The first was that a user might not see interface feedback taking place away from the zoomed area when the complete interface was magnified to a high level. The second was that losing sight of their general location on the interface might disorient users on the interface when the complete interface was magnified to a high level. Both of these might, to some extent be overcome. The former by using a supporting modality for interface feedback, such as sound, and the latter by limiting the range of magnification available to the user to a level that allows accurate selection of the smallest objects but no further.

This review and comparison of previous work on interface target magnification techniques suggested that a *direct* zoom target magnification enhancement would potentially have greater advantages (direct manipulation, ability to incrementally home-in on targets) and fewer disadvantages than an *indirect* zoom enhancement (limited indirect manipulation, more complex interaction to complete a task). (An example of how this may operate can be found in interactive mapping programs[1]). For these reasons a direct zoom enhancement was developed.

12.2 Developing a zoom screen

The requirements for the zoom screen enhancement for this work were that the enhancement might be used on any standard graphical user interface, that it would not require modification to that interface, and that it would allow direct interaction with that interface.

Many simple magnification tools operate by taking a 'snap-shot' of the full interface and then displaying that 'snap-shot' at varying levels of magnification. The user then performs an object manipulation on the 'snap-shot', with that interaction then translated to the interface. This approach has the inherent disadvantage that the user is interacting with a still picture of the interface, rather than the interface itself. Thus any feedback from the interface is not shown in the 'snap-shot' until the user zooms-out to see the full, actual interface. This approach was not acceptable for this work, as the user may manipulate objects and receive feedback from the interface whilst maintaining a zoomed-in view on

[1] www.multimap.com for example allows zooming at the cursor into an online map.

the interface. Thus a zoom tool was required that would be 'live' and actually allow direct interaction with, and feedback from, the interface even at high levels of magnification. Developing such a magnification tool was not trivial, and hence the tool was developed from an existing screen magnifier designed for assisting users with low vision[1].

The screen magnification software was modified to magnify the complete interface, rather than a portion of the interface as the software was originally designed. In addition the software was modified to automatically centre magnification at the cursor location during zooming, to also allow full software control over zoom level, and to automatically allow panning of the screen when zoomed. Automatic panning was added to allow more natural interaction so that objects could be dragged to any part of the interface whilst still zoomed-in on the interface, rather than cause the user to first pick up an object, zoom-out, drag the object, and then zoom-in again to drop the object. Zoom level was controlled by using two additional buttons, one for zoom-in, one for zoom-out, placed on the hand held switch box used for multimodal target selection (Figure 12.6). Zoom levels were set to ×1, ×2, ×4, ×8, ×16, with each press of the zoom control buttons zooming either in or out by one step. This gave the ability for the user to magnify the smallest target size in the test method (0.3° visual angle at 60cm) up to a size of 4.8° visual angle, equivalent to 4 times the size of the largest target (1.2° visual angle at 60cm) in the test method. Greater levels of magnification were not permitted by software control to avoid users being disoriented on the interface when the interface was magnified to a very high level.

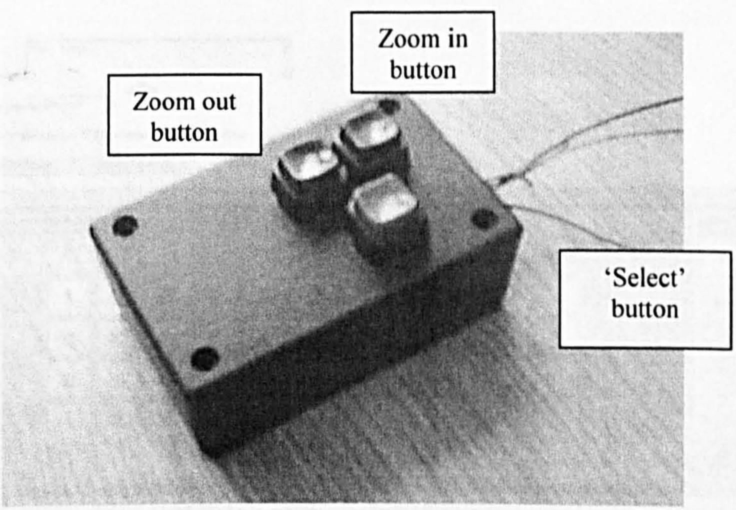


Figure 12.6 Switch click and zoom level tool

---

[1] The author would like to express his thanks to Dolphin Computer Access Ltd. for developing a modified version of their 'Supernova' screen magnifier specifically for this work. www.dolphinuk.co.uk

One important factor when controlling the operation of the interface during zooming was the placement of the monomodal operation dwell click tool. As detailed in Chapter 9, the tool was specifically written for this work, rather than using commercially available software, to allow customisation of the tool operation. This allowed the placement of the dwell click tool to be manipulated to keep the tool visible and at the correct size during zoom operation. If replacement and resizing of the dwell tool were not implemented then during a zoom the tool would both increase unnecessarily in size and also potentially disappear from the visible magnified screen area. Hence the dwell tool software was amended to monitor the zoom level of the interface and correctly resize and reposition the tool on the interface.

The operation of the zoom software and dwell click tool during a target selection is shown in a sequence of screen captures (Figures 12.7, 12.8 and 12.9). Note the 'live' operation of the screen zoom software where manipulation and feedback from the interface are shown during a zoom on the magnified interface (Figure 12.9).

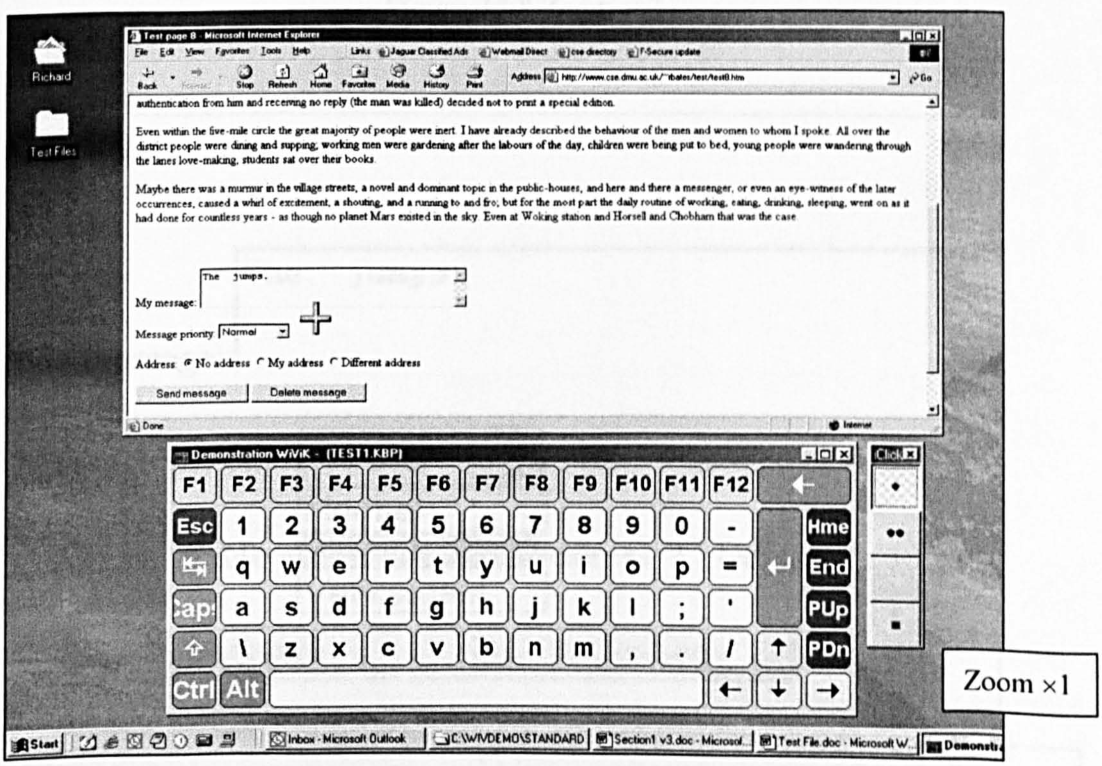


Figure 12.7 Zoom × 1

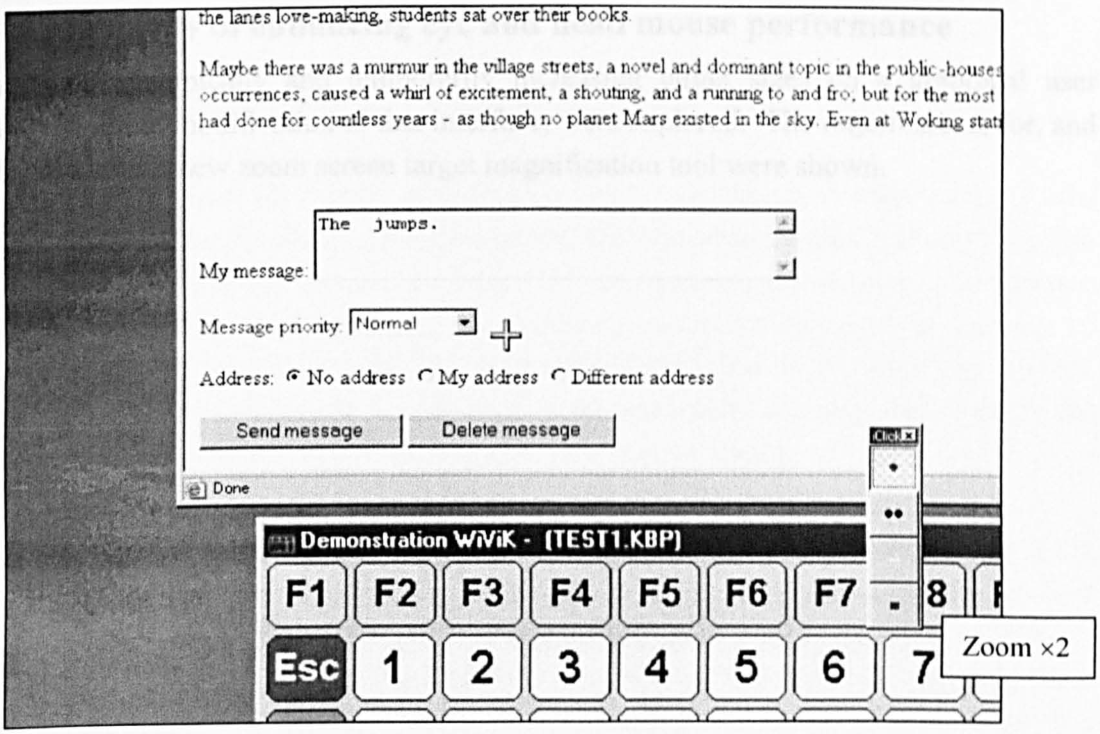


Figure 12.8 Zoom × 2

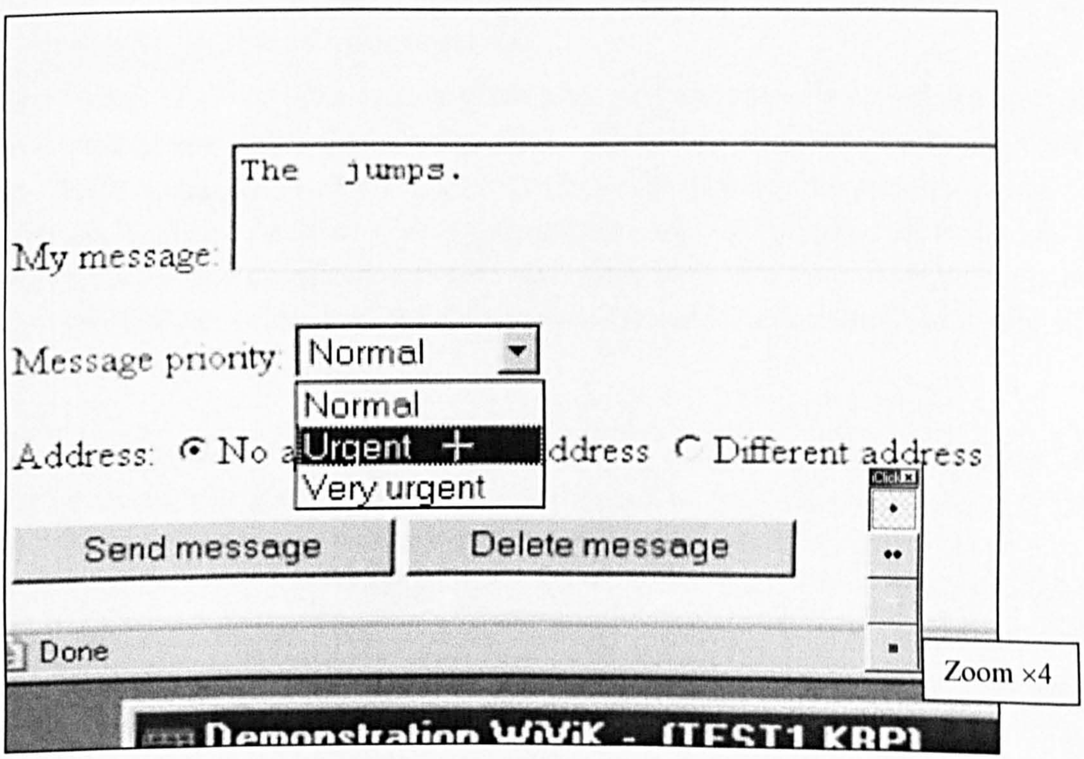


Figure 12.9 Zoom × 4

12.3 A summary of enhancing eye and head mouse performance

Methods of dynamically and temporarily increasing target sizes on a graphical user interface, without modification to that interface, were explored. The requirements for, and development of, a new zoom screen target magnification tool were shown.

Chapter 13

The Effect of Enhanced Eye and Head Mice

This chapter determines the effects of adding the object magnification enhancement to the head and eye mice in this work. This chapter and the following chapter build on all of the previous work contained in this thesis, and are the final product of this work. This chapter first predicts, based on the results from the previous experiment examined in Chapters 10 and 11, where and to what extent the enhancement described in the previous chapter, Chapter 12, will benefit the performance of head and eye mice, and also whether the enhancement will have any associated costs for any benefits found.

The chapter goes on to describe a new experiment with the enhanced head and eye mice, using the verified assessment method produced in the earlier chapters of this work. This experiment is used to determine the validity of the performance predictions for the enhancement. This chapter gives a high-level analysis of the objective and subjective results of the experiment and the effects of the enhancement, and finally leads on to a final, more in-depth, analysis in the following chapter, Chapter 14, of this work.

13.1 Predicted effects of enhancement

Chapters 10 and 11 of this work examined the basic performance of head and eye mice, in both monomodal and multimodal configurations, against the baseline of a desktop hand mouse. These chapters showed that neither device could approach the performance of the hand mouse, and that the head mouse outperformed the eye mouse. However, these chapters showed that if the effective target object sizes on the interface could be increased then the performance of the both head and eye mice could be increased to potentially equal hand mouse performance.

From the basic results of Chapter 10 the following predictions were made about the effect of adding a target magnification enhancement to the head and eye mice in this work, with this chapter then addressing each of these predictions in numerical order:

1. *No difference* in performance *between word and internet task domains* for the enhanced head and eye mice as the tasks within each domain were unchanged (from Figure 10.1 and Chapter 10.6).

2. *The benefit* of *increased overall* enhanced head and eye mouse *performance* due to the effective target sizes of objects in the test tasks being increased by the enhancement thus reducing the effects of the low pointing accuracy of the head and eye mice (from Figure 10.2 and Chapter 10.6).

3. *The benefit* of a *reduction* in *control correction times* above other task time elements due to the effective target sizes of objects in the test tasks being increased by the enhancement (from Figure 10.3 and Chapter 10.7).

4. *The cost* of an *increase in overall task time* due to the addition of zoom control times when increasing target object sizes with the enhancement (from Figure 10.3 and Chapter 10.7).

5. *The benefit* of an *increase* in interaction *task quality*, with a *marked reduction in control corrections* above other errors due to the effective target sizes of objects in the test tasks being increased by the enhancement (from Figure 10.4 and Chapter 10.8).

6. *The cost* of a *decrease* in interaction *task quality* with the addition of *zoom level corrections* when controlling the zoom enhancement for the enhanced devices (from Figure 10.4 and Chapter 10.8).

7. *The benefit* of an *increase* in overall device *satisfaction* for the *enhanced devices* as the effective target sizes of objects in the test tasks are increased by the enhancement and are hence easier to select (from Table 10.6 and Chapter 10.9).

8. *The cost* of an *increase* in overall *workload* due to the additional complexity of controlling the zoom enhancement for the enhanced devices (from Figure 10.5 and Chapter 10.9).

9. *No change* in device overall *comfort* for the enhanced devices as the devices themselves have not substantially changed by the addition of the zoom enhancement (from Figure 10.5 and Chapter 10.9).

10. *The benefit* of an increase in *ease of use* for the enhanced devices as the target sizes of objects in the test tasks are increased by the enhancement and are hence easier to select (from Figure 10.5 and Chapter 10.9).

13.2 Experiment 6: The effect of enhancing head and eye mice

Four devices were to be assessed; the eye mouse and the head mouse in both monomodal and multimodal form, as described in Chapter 9, with the zoom enhancement, as described in Chapter 12, by using the 'real world' assessment method devised in this work[1].

A within-subjects test design was adopted, with the same subjects from the previous assessment of non-enhanced head and eye mice used, as described in Chapter 10. Exactly the same test method was used as before, described in Chapter 10, to enable comparison between non-enhanced and enhanced devices and also to the baseline hand mouse. As before, there was no penalty for non-participation, all subjects signed a consent form (Appendices Figure A13.1) and all data was anonymous. To compensate for order effects in the testing, the presentation order of the devices was prescribed with an incomplete Latin Square design (Appendices Tables A13.2 and A13.3).

The test procedure was the same as before, Chapter 10, except that test subjects now had the opportunity to control screen zoom magnification level using the enhancement detailed in Chapter 12. One addition to the experimental procedure shown previously (Chapter 10, Figure 10.2) was that during the practice session before each test, subjects were also allowed to practice using the zoom enhancement until they became comfortable with its operation.

Data analysis was as before, Chapter 10, with all data analysed by stepping through captured video files noting task quality and task time, with the addition of also noting how the magnification enhancement was manipulated. This was characterised by recording the zoom magnification level used when object manipulations occurred, the time taken to manipulate the zoom enhancement during a task, and finally the number of zoom level corrections generated. Here a zoom level correction was counted when a subject reversed zoom level during a task due to poor control of the magnification enhancement, for instance a sequence of ×1, ×2, ×4, ×2 would generate a zoom level correction count due to the inadvertent use of ×4.

A new test marking sheet was used to include these additional zoom level observations (Appendices Chapter 13, Figure A13.1), with these additions to the original device assessment (Chapter 10) shown with a dark bullet (●):

[1] The hand mouse with zoom was not assessed as an informal pilot study (not reported here) showed that zoom was not used in any interaction with the hand mouse

Independent variables were:

- ● Device (Monomodal head and eye mouse with zoom, multimodal head and eye mouse with zoom) (Chapter 12).
- ○ User experience (Low, Medium, High) (Chapter 10).
- ○ Task target size (S1, S2, S3, S4 in pixel/mm/visual angle) (Chapter 4).
- ○ Task interaction type (Single click, Double click, Drag, Restricted/Unrestricted) (Chapter 4).

Dependent variables were:

- ○ Task Efficiency (%) (Chapter 5).
- ○ Task Quality (1-5) (Chapter 5).
- ○ Task Time (mS) (Chapter 5).
- ○ Task time taken by non-productive actions (mS) (Chapter 5).
- ● Task Zoom magnification level (×1, ×2, ×4, ×8, ×16) (Chapter 12).
- ● Task time taken for controlling the zoom enhancement (mS) (Chapters 5 and 12).
- ● Task Zoom level corrections (count / task) (Chapter 13).
- ○ Device pointing accuracy (pixel/mm/visual angle from test targets) (Chapter 9).
- ○ Device assessment questionnaire (workload, comfort, ease of use, 1-7) (Chapter 5).

13.3 Enhanced eye and head mouse efficiency and task domain

This section extends the results from the standard devices (Chapter 10.5) to include the results from the enhanced devices. As before, the efficiency results for all of the devices on the assessment method (Appendices Table A13.3, Graphed in Figure 13.1) showed efficiencies, including the enhanced devices, which were all still lower than the baseline hand mouse for all domains.

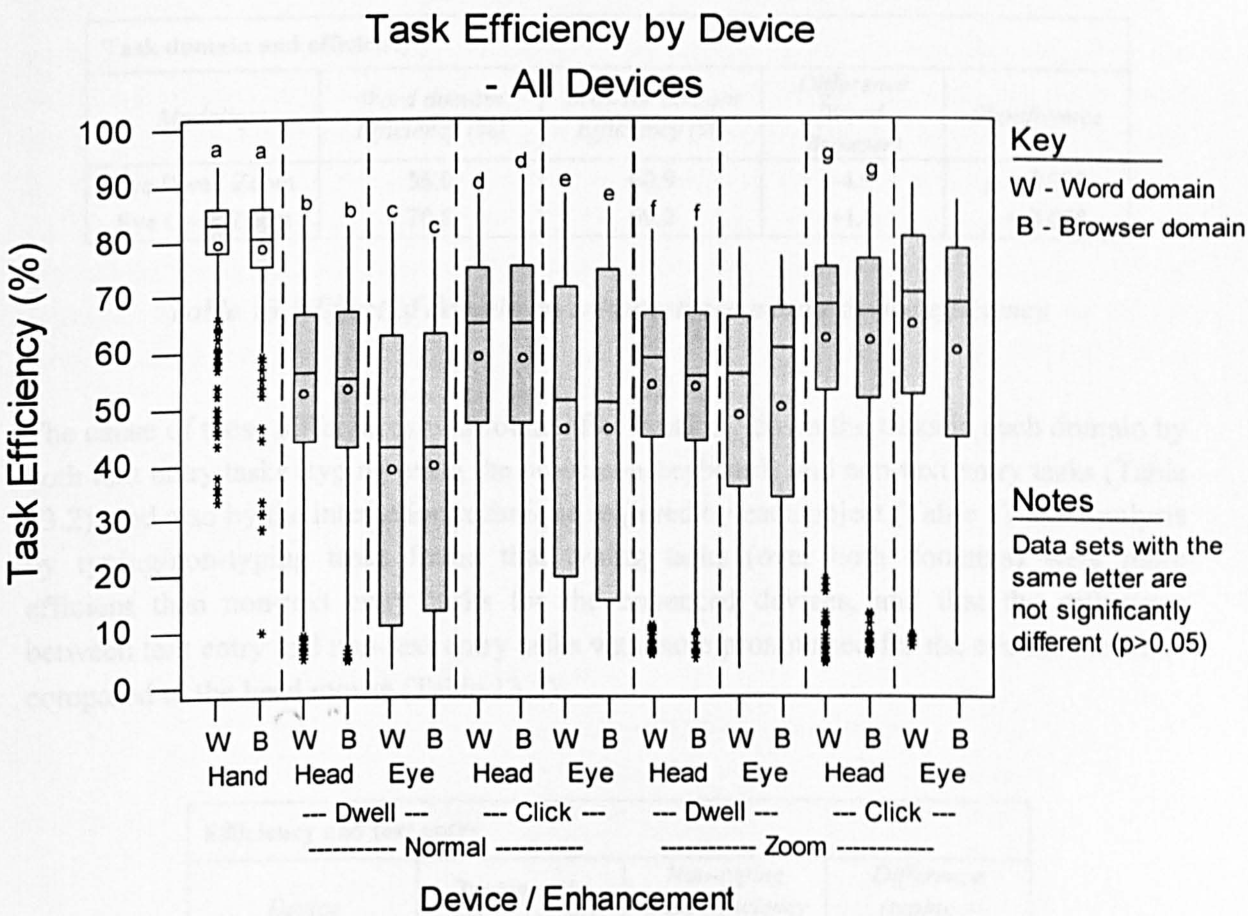


Figure13.1 Enhanced and standard device task efficiency by domain

As with the non-enhanced devices (Figure 10.1, repeated within Figure 13.1), the performance of the enhanced devices appeared to show little difference between the two test domains, Mann-Whitney two-sample rank tests (Comparisons shown in Appendices Table A13.4). This similarity between domains confirmed that the context or nature of the tasks had little effect on the performance of the enhanced head mouse devices, but found a significant difference between the domains for both the monomodal (p = 0.037) and multimodal (p = 0.008) enhanced eye mice (Table 13.1). This difference from the prediction that enhancement would not affect the similarities in domain performance (*Prediction 1*) was investigated for the enhanced eye mice. Of interest was that the differences between the domains were different between the monomodal dwell and multimodal click enhanced eye mice, with the Word domain being less efficient than the Browser domain for monomodal operation, and the Word domain being more efficient than the Browser domain for multimodal operation (Table 13.1).

Task domain and efficiency				
Modality	*Word domain Efficiency (%)*	*Browser domain Efficiency (%)*	*Difference (Word-Browser)*	*Significance*
Eye Dwell Zoom	56.0	60.9	**-4.9**	$p = 0.037$
Eye Click Zoom	70.8	69.2	**+1.6**	$p = 0.008$

Table 13.1 Effect of domain on enhanced eye mouse device efficiency

The cause of these differences was found after breaking down the tasks in each domain by both text entry tasks (typing using the on-screen keyboard) and non-text entry tasks (Table 13.2), and also by the interaction technique required by each object (Table 13.3). Analysis by typing/non-typing tasks found that typing tasks (over both domains) were more efficient than non-text entry tasks for the enhanced devices, and that the difference between text entry and non-text entry tasks was more pronounced for the eye mouse when compared to the head mouse (Table 13.2).

Efficiency and text entry			
Device	*Typing tasks Efficiency (%)*	*Non-typing tasks Efficiency (%)*	*Difference (typing – non typing)*
Head Dwell Zoom	62.0	52.7	***9.3***
Eye Dwell Zoom	61.0	53.6	**7.6**
Head Click Zoom	71.2	62.0	***9.2***
Eye Click Zoom	76.4	62.7	**14.7**

(Similarities between devices are highlighted in ***bold italic***)
(Larger differences between devices are highlighted in **bold**)

Table 13.2 Effect of text entry on enhanced eye and head mouse device efficiency

Examining these found very little difference between the monomodal and multimodal head mouse for text and non-text entry tasks (9.3 and 9.2 percentage points) (Table 13.2, column 4, in ***bold italic***), but a considerable difference between the monomodal and multimodal eye mouse (7.6 and 14.7 percentage points) (Table 13.2, column 4, in **bold**).

The increased performance of the enhanced multimodal eye mouse was explained by examination of the video of the test tasks. Participants were found to use the zoom enhancement to magnify the on-screen keyboard and then type a sequence of letters before zooming out and returning the interface to its normal state. In the zoomed state the keyboard offered large easily selected targets that were rapidly selected using the hand held click button with the multimodal eye mouse, but were relatively more slowly selected by the dwell click device, where users were required to wait until a dwell click elapsed (1000ms) before each keyboard key could be selected. This resulted in a performance advantage for typing for the enhanced multimodal eye mouse. This advantage was not repeated for the multimodal head mouse due to the slower pointing speed of the device between keyboard keys, resulting in the similarity of results between the head mice.

The effects of interaction technique were then investigated. Here there were differences between the monomodal and multimodal eye mouse between single restricted and unrestricted clicks (Table 13.3).

Efficiency and interaction technique					
Device	*Interaction Technique Efficiency (%)*				
	-------- Restricted --------			*---- Unrestricted ----*	
	Single	*Double*	*Drag*	*Single*	*Drag*
Head Dwell Zoom	53.8	48.0	33.7	60.5	32.9
Eye Dwell Zoom	*59.1*	52.0	28.6	*59.8*	36.1
Head Click Zoom	66.4	49.4	34.7	70.7	39.9
Eye Click Zoom	**51.9**	54.5	45.3	**72.1**	50.9

(Similarities within devices are shown in *italic*)
(Larger differences within devices are highlighted in **bold**)

Table 13.3 Effect of interaction technique on device efficiency

The monomodal eye mouse maintained similar performance between the two interaction techniques (59.1% and 59.8%) as the effectiveness of the dwell click device was unaffected by the restriction of manipulation of the target objects. However, the multimodal eye mouse was affected by the two interaction techniques (51.9% and 72.1%) with a lower performance for restricted clicks, where any movement off the target during selection will result in the object not being selected, and a higher performance for unrestricted clicks where the cursor is free to move during selection. The difference between the task efficiencies between the Word and Browser domains for the enhanced eye mouse was thus accounted for by a composite of typing tasks being more efficient with click than dwell (14.7 % points difference vs. 7.7 % points difference) and restricted single clicks being more efficient with dwell than click (59.1 % points difference vs. 51.9 % points difference) and click being more efficient with unrestricted clicks (72.1 % points difference vs. 59.8 % points difference).

In summary:

- *Prediction 1 – "that there would be no difference in performance between domains for the enhanced head and eye mice as the tasks within the domains were unchanged" - did not hold for the eye mouse*
- *The test task domain influences eye mouse performance once the effects of target size are reduced through the use of a magnification enhancement*
- *Typing tasks are more efficient for the multimodal eye mouse due to speed of selection*
- *Restricted object manipulation is more efficient for the monomodal eye mouse due to automated dwell selection*

13.4 Enhanced eye and head mouse efficiency

The domain results were pooled for each device to give greater clarity and ease of comparison of the performances of the devices to each other (Appendices Table 13.5, showed graphically in Figure 13.2). Wilcoxon matched pairs signed rank tests (Sprent 1993) were used to investigate the significance of any differences between the pooled domain efficiencies of each device (Appendices Table A13.5). The comparisons showed that the pooled performances of the enhanced monomodal head and monomodal eye devices were not statistically significantly different from each other (p = 0.180) and that the pooled performances of the enhanced multimodal head and multimodal eye devices were also not statistically significantly different from each other (p = 0.145). In both cases this showed that the zoom enhancement disproportionately *increased* eye mouse performance in comparison to head mouse performance, resulting in *equality of performance* between head and eye mice for either monomodal or multimodal operation. This result differed from the prediction that enhancement would increase performance for *all* devices due to increased target sizes (*Prediction 2*). The prediction was correct for the eye mice, but less so for the head mice. The lack of any appreciable performance increase for the enhanced head mice over the non-enhanced head mice can probably be explained by the pointing speed of the head mice being more of a limiting factor than target size due to the higher pointing accuracy of the head mice (Chapter 10.6). Hence increased target sizes offered little advantage for the head mice. In contrast, the prediction was true for the eye mice as target size was a greater limiting factor for the eye mice due to their inherent pointing inaccuracy. Hence increased target sizes offered advantages for the eye mice in comparison to the head mice.

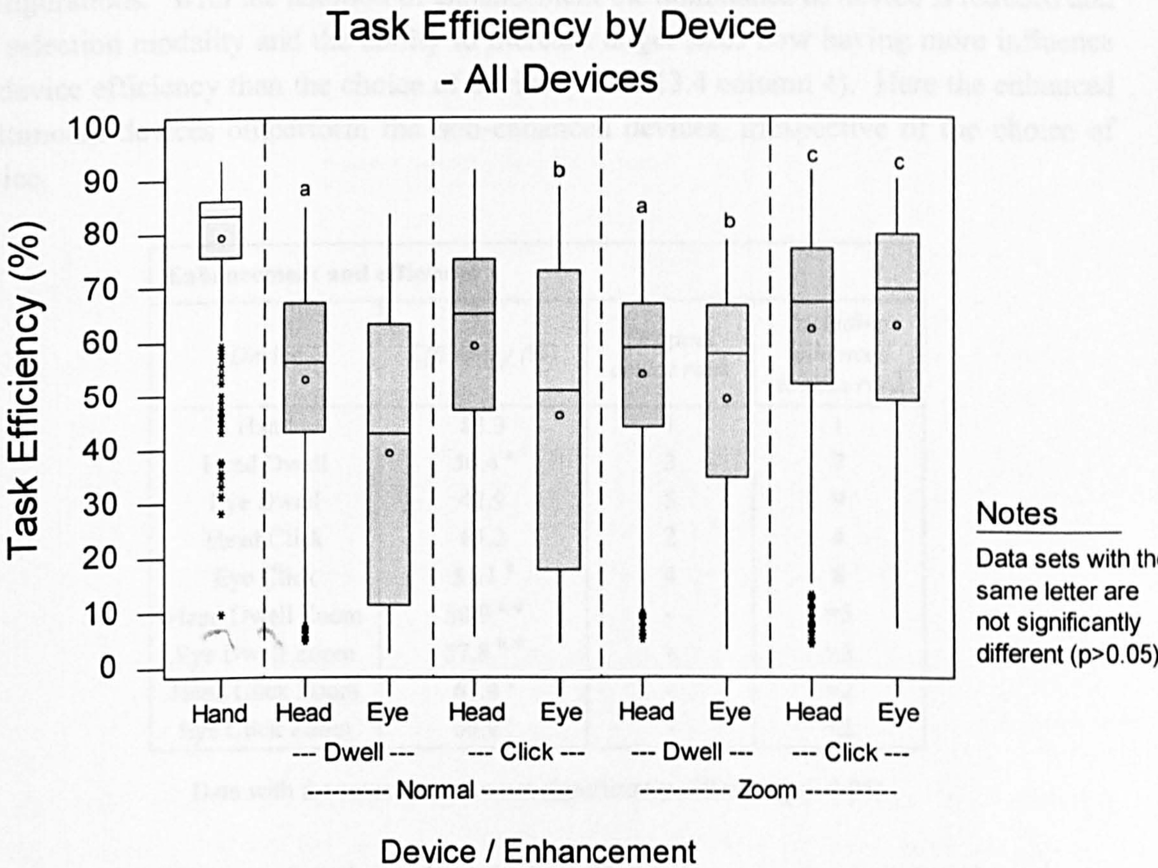


Figure13.2 Standard and enhanced device overall task efficiency

In summary:

- *Adding control of target object size appreciably increases eye mouse performance but does not appreciably increase head mouse performance*
- *Eye mouse performance was raised disproportionately to be in parity with head mouse performance*

- Prediction 2 – *"That there would be a benefit of increased overall enhanced head and eye mouse performance due to the effective target sizes of objects in the test tasks being increased" - held for the eye mouse but not for the head mouse.*

Ranking the devices by efficiency (Table 13.4) showed that the hand mouse still outperformed all devices, despite the addition of the zoom enhancement. In the original device ranks (Chapter 10 Table 10.3, repeated in Table 13.4 column 3) the choice of device had more influence on measured efficiency than the choice of selection modality, with the head mouse outperforming the eye mouse in both monomodal and multimodal configurations. With the addition of enhancement the dominance of device is reduced and the selection modality and the ability to increase target sizes now having more influence on device efficiency than the choice of device (Table 13.4 column 4). Here the enhanced multimodal devices outperform the non-enhanced devices, irrespective of the choice of device.

Enhancement and efficiency			
Device	*Efficiency (%)*	*Original device rank*	*Including enhanced devices rank*
Hand	83.3	1	1
Head Dwell	56.4 [a]	3	7
Eye Dwell	42.9	5	9
Head Click	65.2	2	4
Eye Click	51.1 [b]	4	8
Head Dwell Zoom	58.9 [a, d]	-	=5
Eye Dwell Zoom	57.8 [b, d]	-	=5
Head Click Zoom	67.4 [c]	-	=2
Eye Click Zoom	69.9 [c]	-	=2

Data with the same letter are not significantly different ($p > 0.05$)

Table 13.4 Enhanced and standard device efficiency comparisons and rankings[1]

In summary:

- *Multimodal target selection and the zoom enhancement remove the dominance of device choice (between either head or eye based pointing) for the highest assistive technology device task efficiency*

[1] Note these and following tabled results are not corrected for alpha, see Notes Discussion N2. 'Multiple comparisons in this work' in the Appendices for a discussion on multiple comparisons. Overall mean values shown, non-parametric statistical comparisons used raw satisfaction rating data.

- *Provided the subjective satisfaction of the devices was similar, the enhancement allows users to choose either head or eye devices for similar interaction performance*

13.5 Enhancement and device task time

This section extends the results from the standard devices (Chapter 10.7) with the total task time results for all of the devices now calculated (Appendices Table A13.7, summary Table 13.5) and any statistical significance of any difference determined (Wilcoxon matched pairs signed rank tests, Appendices Table A13.8).

Enhancement and task time	
Device	*Task time (ms)*
Hand	1246
Head Dwell	3489 [a b]
Eye Dwell	3668 [a c]
Head Click	2537 [d]
Eye Click	3289 [e]
Head Dwell Zoom	3480 [b c]
Eye Dwell Zoom	3646
Head Click Zoom	2900 [d e]
Eye Click Zoom	2225

Data with the same letter are not significantly different ($p > 0.05$)

Table 13.5 Enhanced and standard device task time comparison

For the head mice, comparison of the standard head dwell device task time with the enhanced head dwell device task time found no significant difference between the two times for the monomodal devices ($p = 0.459$), and comparison between the standard head click device task time with the enhanced head click device task time also found no significant difference between the two times ($p = 0.121$). These results indicated that any time used controlling the enhancement or gained from the benefit of the enhancement did not affect overall task time for the head mice. For the eye mice, there were significant differences between the standard eye dwell device task time and the enhanced eye dwell task time ($p = 0.012$) and between the standard eye click device task time and the enhanced eye click task time ($p < 0.001$), with the enhanced devices exhibiting lower task times. Clearly the enhancement reduced overall task times for the eye mice.

The overall task times were then broken down into their composite time elements (Appendices Table A13.7, Graphed in Figure 13.3). It was predicted that the addition of enhancement would reduce the task time taken with unproductive *control corrections*[1] (shown in blue on Figure 13.3) above other time elements based on the analysis of task time in Chapter 10.7 due to the effective target sizes of objects in the test tasks being increased by the enhancement, and so requiring less pointing accuracy to manipulate (*Prediction 3*). It was also predicted that this benefit would also incur the cost of the time taken controlling the zoom enhancement (shown in pink on Figure 13.3) (*Prediction 4*). In order to determine which predictions were correct, investigation into the device tasks times was required.

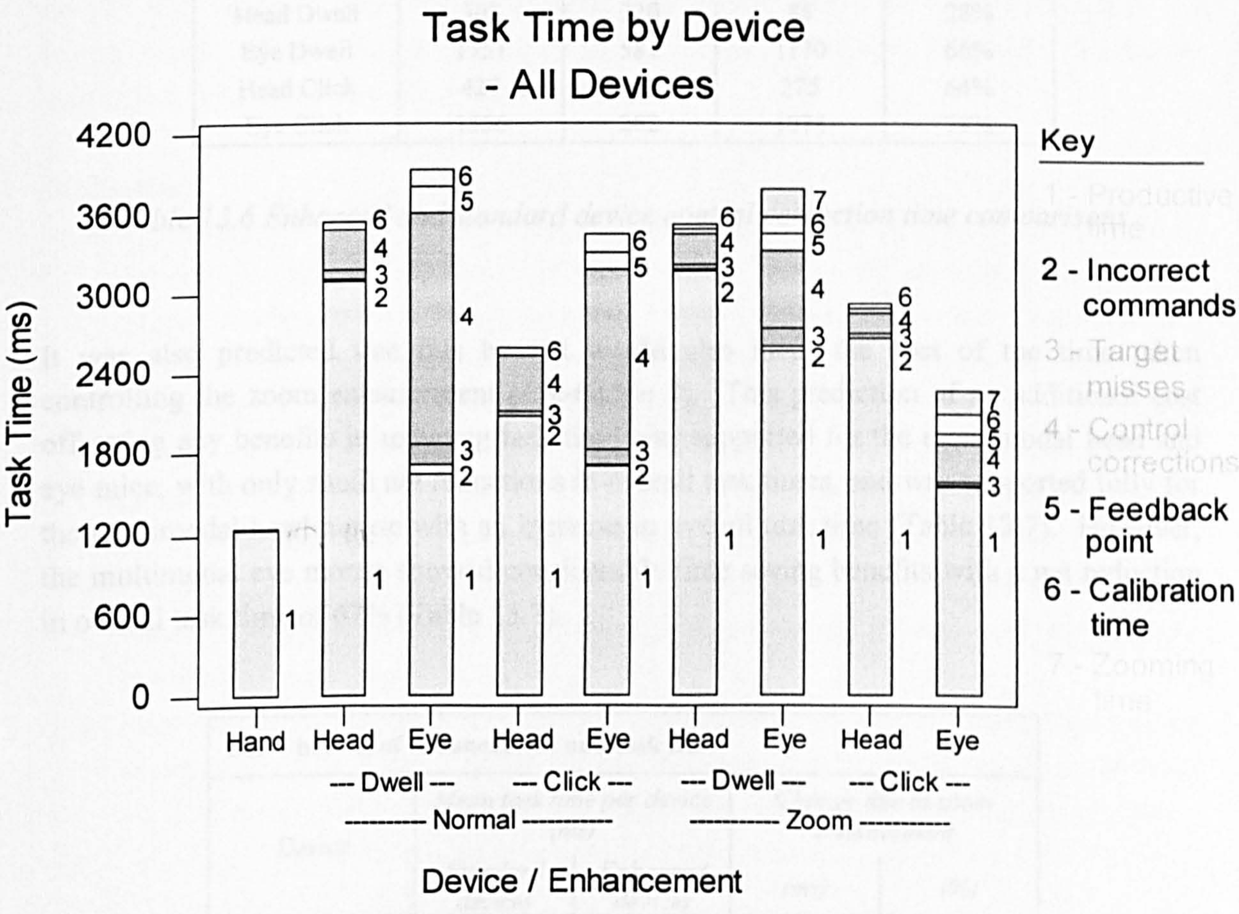


Figure13.3 Composition of enhanced and standard device task time

[1] Control corrections are cursor path corrections or pauses of cursor movement away from an idealised 'perfect' cursor movement during a task (MacKenzie 2001). These variations and pauses, described in Chapter 5.4, indicate a lack of control when compared to such an idealised 'perfect' cursor movement as they generate output that was not asked for, and hence give an unwanted increase in task time and also an unwanted reduction in task quality.

The analysis by individual time elements (Figure 13.3) clearly showed considerable reductions in time spent with control corrections. This supported the prediction (*Prediction 3*) that increasing target object sizes would reduce the time spent in unproductive cursor control corrections, with reductions for all devices (Table 13.6).

Enhancement and control correction time				
Device	*Mean control correction time per task (ms)*		*Reduction due to enhancement*	
	Standard devices	*Enhanced devices*	*(ms)*	*(%)*
Hand	4	-	-	-
Head Dwell	305	220	85	28%
Eye Dwell	1751	581	1170	66%
Head Click	427	152	275	64%
Eye Click	1356	282	1074	79%

Table 13.6 Enhanced and standard device control correction time comparisons

It was also predicted that this benefit would also incur the cost of the time taken controlling the zoom enhancement (*Prediction 4*). This prediction of an additional cost offsetting any benefits in reducing task time was supported for the monomodal head and eye mice, with only small net reductions in overall task times, and was supported fully for the multimodal head mouse with an increase in overall task time (Table 13.7). However, the multimodal eye mouse showed considerable time saving benefits with a net reduction in overall task time of 67% (Table 13.7).

Cost / benefit of enhancement and task time				
Device	*Mean task time per device (ms)*		*Change due to zoom enhancement*	
	Standard devices	*Enhanced devices*	*(ms)*	*(%)*
Hand	1246	-	-	-
Head Dwell	3489	3480	-9	-0.2%
Eye Dwell	3668	3646	-22	-0.5%
Head Click	2537	2900	+363	+14%
Eye Click	3289	2225	-1064	-67%

Table 13.7 Enhanced and standard device task time cost/benefit of enhancement

The results for the monomodal head and eye mice and the multimodal head mouse (Figure 13.3) showed that although the zoom enhancement reduced time spent in unproductive control corrections, the addition of the enhancement also tended to increase productive times so there was little net time benefit. This increase in productive times can be explained for the monomodal devices, as subjects appeared to take more care and time positioning the cursor before a dwell click was generated, and can also be explained for the monomodal and multimodal head mice by subjects again taking more care and time positioning the cursor due to the slower cursor movement of the head mouse. The benefit of the enhancement was most obvious when multimodal selection was combined with the rapid cursor movement of the eye mouse (compare columns 5 and 9 on Figure 13.3 and the reduction due to enhancement for the eye click mouse on Table 13.6). Here the net overall task time was reduced as the cursor was rapidly placed on a now magnified and hence easily selected target object, with a resultant reduction in productive cursor movement time and also control correction time.

In summary:

- <u>*Prediction 3*</u> – *"that enhancement would give the benefit of a reduction in control correction times above other task time elements due to the effective target sizes of objects in the test tasks being increased by the enhancement" – was supported with the zoom enhancement reducing time spent in control corrections for all devices*
- <u>*Prediction 4*</u> – *"that enhancement would give the cost of a potential increase in overall task time due to the addition of zoom control times when increasing target object sizes with the enhancement" – was not supported for the head mice and the monomodal eye mouse, but was supported for the multimodal eye mouse*
- *The addition of the zoom enhancement did not reduce overall task times for the head mouse and the monomodal eye mouse, but did reduce overall task time for the multimodal eye mouse*
- *Note that these findings may change if the nature of the test tasks are changed, for example if a significantly greater or smaller number of large or small targets were introduced into the test resulting in significantly reduced or increased use of zoom.*

13.6 Enhancement and device quality

This section extends the results from the standard devices (Chapter 10.8) with the total task quality results for all of the devices now calculated (Appendices Table A13.9, summary Table 13.8) and any statistical significance of any difference determined (Wilcoxon matched pairs signed rank tests, Appendices Table A13.10).

Enhancement and task quality	
Device	*Task quality (1-5)*
Hand	4.90
Head Dwell	4.26 [abc]
Eye Dwell	3.25 [e]
Head Click	4.23 [ad]
Eye Click	3.42 [e]
Head Dwell Zoom	4.38 [bf]
Eye Dwell Zoom	4.10 [cd]
Head Click Zoom	4.57 [g]
Eye Click Zoom	4.47 [fg]

Data with the same letter are not significantly different ($p > 0.050$)

Table 13.8 Enhanced and standard device task quality comparisons

For the head mice, comparison of the standard monomodal head mouse task quality with the enhanced monomodal head mouse task quality found no significant difference between the two quality ratings for the monomodal devices ($p = 0.060$), this result indicated that any time used controlling the enhancement or gained from the benefit of the enhancement did not affect overall task time for the monomodal head mouse. In contrast, the addition of the enhancement did benefit the other devices, with the enhanced multimodal head mouse having a significantly higher task quality than the standard multimodal head mouse ($p < 0.001$), the enhanced monomodal eye mouse having a significantly higher task quality than the standard monomodal eye mouse ($p < 0.001$), and finally the enhanced multimodal eye mouse having a significantly higher task quality than the standard multimodal eye mouse ($p < 0.001$). Notably, the enhancement also brought the task quality of the enhanced multimodal eye mouse (quality = 4.47) into parity with the enhanced multimodal head mouse (quality = 4.57) ($p = 0.012$).

The overall task quality ratings were then broken down into their composite task quality elements (Appendices Table A13.11, Graphed in Figure 13.4). It was predicted that the addition of enhancement would give an increase in interaction task quality with a marked reduction in *control corrections* above other error elements, based on the analysis of task quality in Chapter 10.8, due to the effective target sizes of objects in the test tasks being increased by the enhancement and so requiring less pointing accuracy to manipulate (*Prediction 5*). It was also predicted that this benefit would also incur the cost of a potential slight decrease in overall interaction task quality with the addition of zoom level corrections when controlling the zoom enhancement for the enhanced devices (*Prediction*

6). As with task time, in order to determine which predictions were correct, investigation into the device task quality was required.

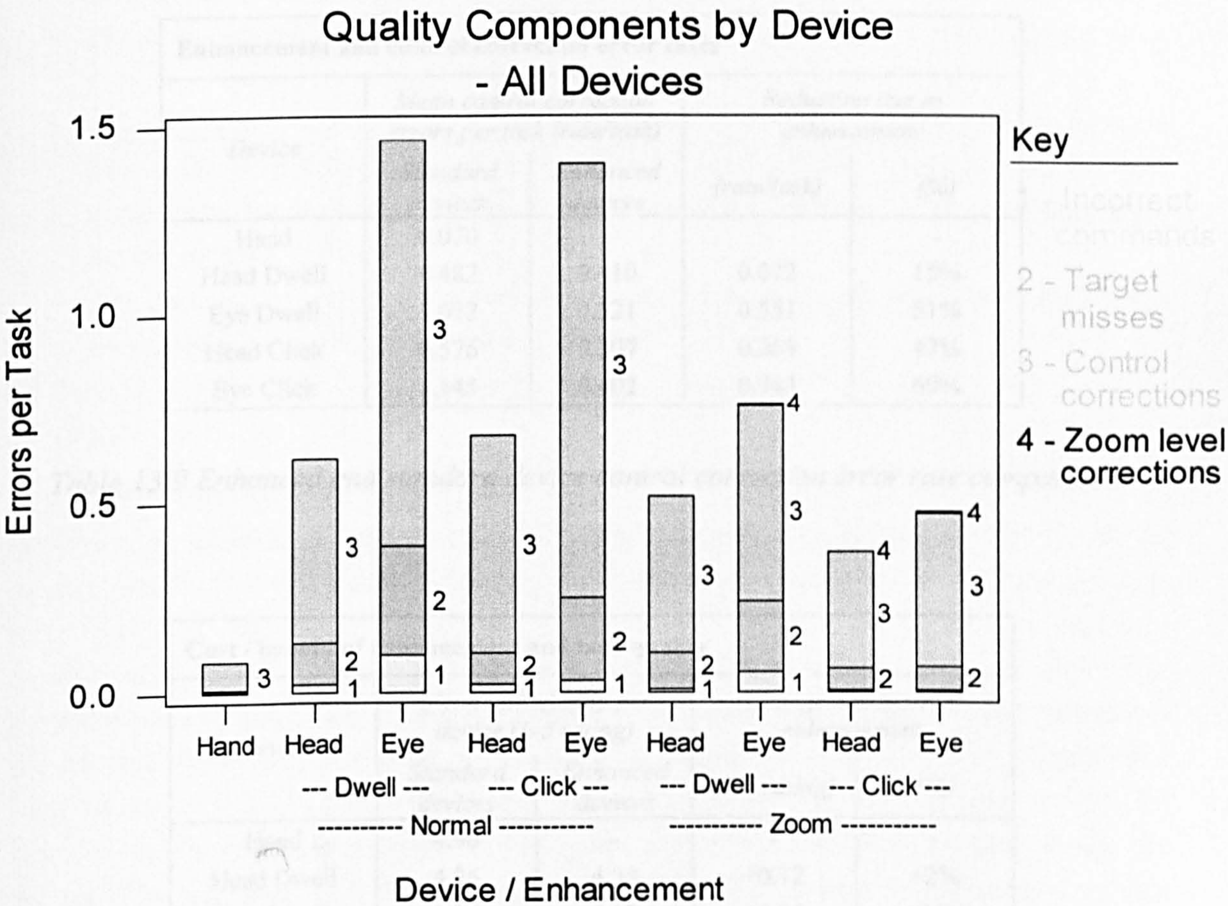


Figure13.4 Composition of enhanced and standard device task quality

The analysis by individual task quality elements (Figure 13.4) clearly showed considerable reductions in the rate of errors per task due to control corrections (shown in blue on Figure 13.4). This supported the prediction (*Prediction 5)* that increasing target object sizes would reduce the number of errors due to unproductive cursor control corrections, with reductions for all devices (Table 13.9).

It was also predicted that this benefit would also incur the cost of the errors produced in controlling the zoom enhancement (*Prediction 6*). This prediction of an additional cost offsetting any benefits in task quality was not supported, with net increases in overall task quality for all enhanced devices (Table 13.10), and further, examining other error types it was also notable that there was also a reduction in the rates of target misses (shown in green on Figure 13.4) for the enhanced eye mice.

Enhancement and control correction error rates				
Device	*Mean control correction errors per task (rate/task)*		*Reduction due to enhancement*	
	Standard devices	*Enhanced devices*	*(rate/task)*	*(%)*
Hand	0.070	-	-	-
Head Dwell	0.482	0.410	0.072	15%
Eye Dwell	1.072	0.521	0.551	51%
Head Click	0.576	0.307	0.269	47%
Eye Click	1.145	0.402	0.743	69%

Table 13.9 Enhanced and standard device control correction error rate comparisons

Cost / benefit of enhancement and task quality				
Device	*Mean task quality per device (1-5 rating)*		*Change due to zoom enhancement*	
	Standard devices	*Enhanced devices*	*(1-5 rating)*	*(%)*
Hand	4.90	-	-	-
Head Dwell	4.26	4.38	+0.12	+2%
Eye Dwell	3.25	4.10	+0.85	+26%
Head Click	4.23	4.57	+0.34	+8%
Eye Click	3.42	4.47	+1.05	+31%

Table 13.10 Enhanced and standard device task quality cost/benefit of enhancement

The benefit of the enhancement was most obvious when multimodal selection was combined with the rapid cursor movement of the eye mouse. Here the net overall task quality was increased, as with decreases in task time for this device, the cursor was rapidly placed on a now magnified and hence easily selected target object, with a resultant reduction in cursor control corrections and target misses, and giving a disproportionate increase in task quality to parity with the head mouse.

In summary:

- <u>*Prediction 5*</u> – *"that the enhancement would give the benefit of an increase in interaction task quality, with a marked reduction in control corrections above other errors due to the effective target sizes of objects in the test tasks being increased by the enhancement" – was supported with the zoom enhancement reducing the error rate for control corrections for all devices*
- <u>*Prediction 6*</u> – *"that enhancement would give the cost of a potential slight decrease in overall interaction task quality with the addition of zoom level corrections when controlling the zoom enhancement for the enhanced devices" – was not supported for the devices*
- *The addition of the zoom enhancement reduced overall task error rates and hence increased quality for the head and eye mice.*

13.7 Enhancement and satisfaction

This section extends the results from the standard devices (see Chapter 10.9) with the overall perceived test subject satisfaction with each of the enhanced devices now assessed after each test session with a device (as described in Chapter 5), with the statistical significance of any differences determined (Wilcoxon matched pairs signed rank tests, Appendices Table A13.11, Summary Table 13.11). Ranking the satisfaction ratings for the devices showed that the hand mouse was still judged the most satisfying to use with a high rating of 6.2/7.0, followed by the enhanced multimodal and standard multimodal head mice, the enhanced monomodal head mouse, and then the enhanced multimodal eye mouse. The rating of this highest performing of the eye mouse configurations as less satisfying than the enhanced head mouse in both multimodal and monomodal configurations, and also less satisfying than the standard multimodal head mouse was perhaps a little surprising given that this eye mouse configuration had achieved parity of efficiency with the highest performing of the head mice. This indicated that although eye mouse efficiency could be dramatically improved with enhancement coupled with multimodal selection, it nevertheless remained less satisfying to use in comparison to the head mice and hand mouse.

It was predicted that the addition of enhancement would give an increase in subjective satisfaction ratings for the enhanced devices as the effective target sizes of objects in the test tasks are increased by the enhancement and are hence easier to select, and so being more satisfying to use (*Prediction 7*). This prediction was supported, although marginally for the head mice, with an increase in subjective overall satisfaction for the enhanced devices over the standard devices (Table 13.12).

Enhancement and subjective satisfaction		
Device	*Satisfaction (1-7)*	*Device rank*
Hand	6.20	1
Head Dwell	4.36 [a]	6
Eye Dwell	2.93 [b]	9
Head Click	4.73 [d]	3
Eye Click	3.90 [abcd]	7
Head Dwell Zoom	4.70 [abd]	4
Eye Dwell Zoom	3.47 [c]	8
Head Click Zoom	4.83 [bd]	2
Eye Click Zoom	4.50 [cd]	5

Data with the same letter are not significantly different ($p > 0.05$)

Table 13.11 Enhanced and standard device satisfaction ratings and rankings

Cost / benefit of enhancement and subjective device satisfaction				
Device	*Satisfaction per device (1-7 rating)*		*Change due to zoom enhancement*	
	Standard devices	*Enhanced devices*	*(1-7 rating)*	*(%)*
Hand	6.20	-	-	-
Head Dwell	4.36	4.70	+0.34	+8%
Eye Dwell	2.93	3.47	+0.54	+18%
Head Click	4.73	4.83	+0.10	+2%
Eye Click	3.90	4.50	+0.60	+15%

Table 13.12 Enhanced and standard device satisfaction cost/benefit of enhancement

In summary:

- <u>*Prediction 7*</u> – *"that there would be the benefit of an increase in overall device satisfaction for the enhanced devices as the effective target sizes of objects in the test tasks are increased by the enhancement and are hence easier to select" – did hold for the devices*

The overall subjective satisfaction ratings were broken down by their individual section results (workload, comfort and ease of use, from Chapter 5) to determine the validity of the satisfaction predictions (Predictions 8, 9 and 10 made at the start of this chapter) (Figure 13.5). Wilcoxon matched pairs signed rank tests were then used to determine the statistical significance of any differences between devices within each satisfaction category (Appendices Tables A13.12 to A13.14).

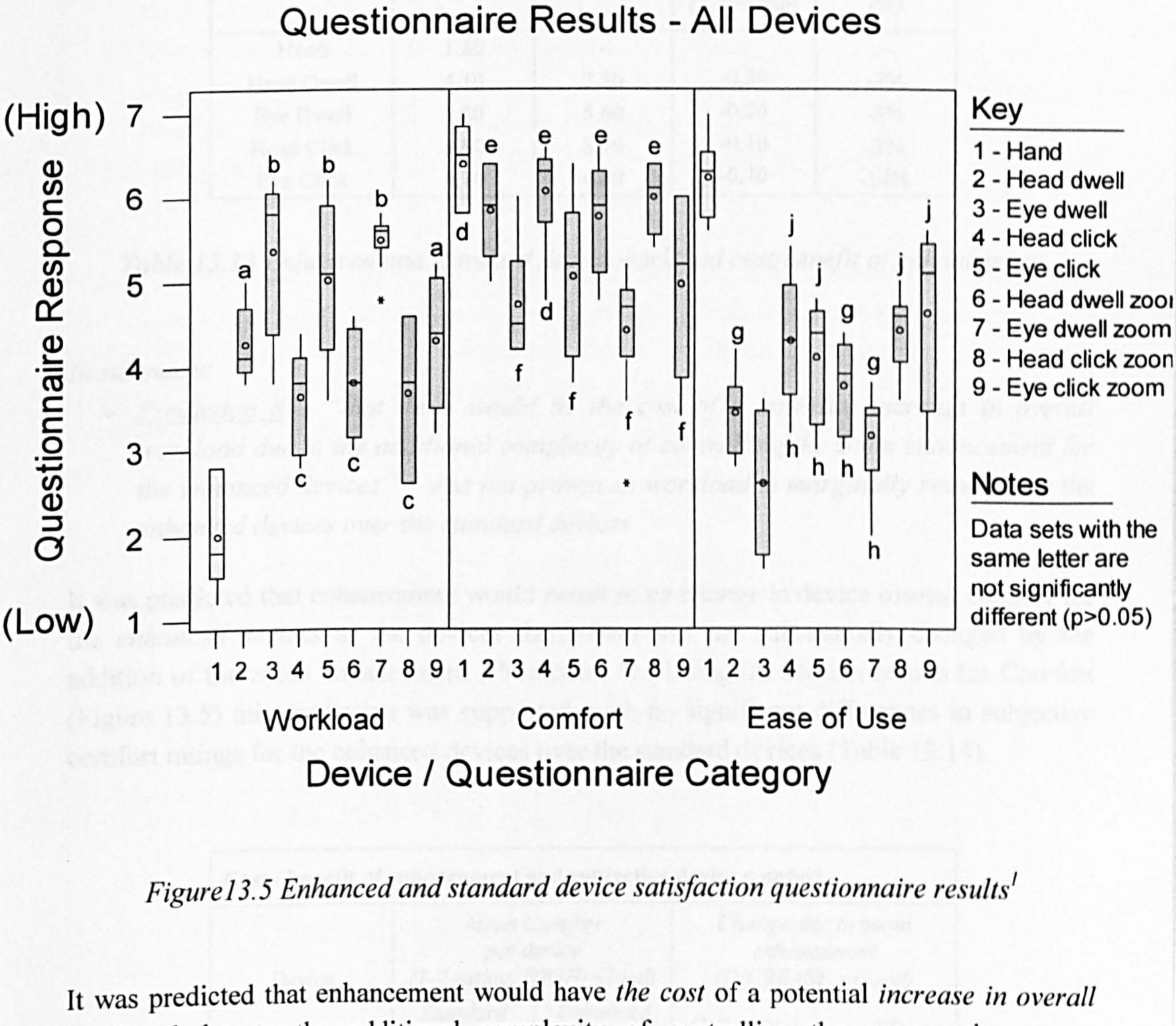


Figure13.5 Enhanced and standard device satisfaction questionnaire results[1]

It was predicted that enhancement would have *the cost* of a potential *increase in overall workload* due to the additional complexity of controlling the zoom enhancement *(Prediction 8)*. Using the section results for Workload (Figure 13.5) this prediction was not supported, with an overall reduction in mean subjective workload ratings for the enhanced devices over the standard devices (Table 13.13).

[1] Note that greater satisfaction is shown by lower workload ratings but higher comfort and ease of use ratings.

Cost / benefit of enhancement and subjective device workload				
Device	*Mean Workload per device (1-7 rating, LOW=Good)*		*Change due to zoom enhancement (REDUCTION =Good)*	
	Standard devices	*Enhanced devices*	*(1-7 rating)*	*(%)*
Hand	1.80	-	-	-
Head Dwell	4.10	3.80	-0.30	-7%
Eye Dwell	5.80	5.60	-0.20	-3%
Head Click	3.80	3.70	-0.10	-3%
Eye Click	5.10	4.40	-0.70	-14%

Table 13.13 Enhanced and standard device workload cost/benefit of enhancement

In summary:

- <u>*Prediction 8*</u> – *"that there would be the cost of a potential increase in overall workload due to the additional complexity of controlling the zoom enhancement for the enhanced devices" – was not proven as workload is marginally reduced for the enhanced devices over the standard devices*

It was predicted that enhancement would *result in no change* in device *overall comfort* for the *enhanced devices* as the devices themselves had not substantially changed by the addition of the zoom enhancement *(Prediction 9).* Using the section results for Comfort (Figure 13.5) this prediction was supported, with no significant differences in subjective comfort ratings for the enhanced devices over the standard devices (Table 13.14).

Cost / benefit of enhancement and subjective device comfort				
Device	*Mean Comfort per device (1-7 rating, HIGH=Good)*		*Change due to zoom enhancement (INCREASE =Good)*	
	Standard devices	*Enhanced devices*	*(1-7 rating)*	*(%)*
Hand	6.50	-	-	-
Head Dwell	5.90	5.90	0.00	0%
Eye Dwell	4.50	4.70	+0.20	+4%
Head Click	6.30	6.10	+0.20	+3%
Eye Click	5.20	5.20	0.00	0%

Table 13.14 Enhanced and standard device comfort cost/benefit of enhancement

In summary:

- <u>*Prediction 9*</u> – *"that there would be no change in device overall comfort for the enhanced devices as the physical requirements of the devices themselves had not substantially changed by the addition of the zoom enhancement" – was proven for these devices only*
- *It is possible that increased use of zoom to magnify targets to larger sizes, or a test sequence with only very large targets, would increase comfort for the devices as less pointing accuracy would be required resulting in less need for the subject to hold either their head still to a high degree of accuracy for the head mouse, or to fixate accurately with their eyes with the eye mouse*

Finally, it was predicted that enhancement would have *the benefit* of an increase in *ease of use* for the *enhanced devices (Prediction 10)* as the effective target sizes of objects in the test tasks are increased by the enhancement and are hence easier to select for the enhanced devices. Using the section results for Ease of Use (Figure 13.5) this prediction was supported, with overall increases in mean subjective ease of use ratings for the enhanced devices over the standard devices (Table 13.15).

Cost / benefit of enhancement and subjective device ease of use				
Device	*Mean Ease of Use per device (1-7 rating, HIGH=Good)*		*Change due to zoom enhancement (INCREASE =Good)*	
	Standard devices	*Enhanced devices*	*(1-7 rating)*	*(%)*
Hand	6.30	-	-	-
Head Dwell	3.50	3.90	+0.40	+11%
Eye Dwell	2.60	3.40	+0.80	+31%
Head Click	4.30	4.60	+0.30	+7%
Eye Click	4.40	5.10	+0.70	+16%

Table 13.15 Enhanced and standard device ease of use cost/benefit of enhancement

In summary:

- <u>*Prediction 10*</u> – *"that there would be the benefit of an increase in ease of use for the enhanced devices as the target sizes of objects in the test tasks are increased by the enhancement, and are hence easier to select by the addition of the zoom enhancement" – was proven for all devices*

13.8 Enhancement and individual satisfaction factors

The final part of this chapter extends the results from the standard devices (Chapter 10.10) and examines the detailed individual factors that comprise both the overall and the sectional satisfaction factors of the enhanced devices so far examined (Table 13.16).

Individual subjective device satisfaction factors

Factor / Device		Hand	*Standard Devices*				*Enhanced Devices*			
			Dwell		*Click*		*Dwell Zoom*		*Click Zoom*	
			Head Mouse	*Eye Mouse*	*Head Mouse*	*Eye Mouse*	*Head Mouse*	*Eye Mouse*	*Head Mouse*	*Eye Mouse*
Workload (low=good)	Physical	1.8	4.5	**5.3**	3.8	**5.5**	4.0	**5.5**	3.6	**4.5**
	Mental	2.3	4.5	**6.6**	3.8	**5.7**	4.1	**6.3**	4.2	**5.3**
	Temporal	2.1	3.5	**3.6**	2.7	**4.3**	3.0	**4.5**	3.3	**3.6**
	Frustration	1.8	4.5	**6.1**	3.7	**5.0**	3.7	**5.7**	3.5	**4.1**
	Performance	1.8	4.1	**5.1**	4.2	**4.7**	**4.2**	**5.5**	3.7	*3.8*
Comfort (high=good)	Headache	6.5	6.1	**5.0**	6.5	**5.5**	5.8	**4.3**	6.1	**5.5**
	Eye	6.1	6.0	**3.6**	6.2	**4.7**	5.7	**4.0**	6.0	**4.3**
	Facial	6.5	6.1	**4.7**	6.2	**5.0**	6.0	**4.5**	6.5	**5.0**
	Mouth	6.8	6.5	**6.0**	6.8	**6.3**	6.5	**6.1**	6.7	**6.3**
	Neck	6.0	4.3	**4.0**	4.6	**3.8**	4.8	**3.1**	4.7	**3.7**
Ease of Use (high=good)	Point Acc.	6.3	3.5	**1.8**	3.8	**2.1**	**4.1**	**2.8**	4.7	*4.7*
	Point Speed	6.0	**3.1**	3.8	**3.5**	4.5	**2.8**	4.0	**3.1**	4.5
	Click Acc.	6.3	**3.3**	**2.3**	4.5	4.5	**4.3**	**2.5**	5.0	4.8
	Click Speed	6.5	**2.8**	**2.3**	4.5	5.1	**2.7**	**3.0**	4.7	4.2
	Sys. Control	6.0	4.3	**2.6**	5.2	**4.1**	4.8	**3.5**	**4.7**	*5.0*

Table 13.16 Individual factors of enhanced and standard device satisfaction

To aid identification of factors that were rated poorly and were particularly of interest, the *two lowest satisfaction results* for each satisfaction factor (such as 'physical workload' or 'eye discomfort' for example) are shown in **bold** type separately for both the standard and enhanced devices. For example, the two poorest factors for the standard devices for the Performance Workload factor are the monomodal eye mouse rated at 5.1 and multimodal eye mouse rated at 4.7 - hence these are shown in **bold**. Separately for the enhanced devices, the two poorest factors for the enhanced devices for the Performance Workload factor are the monomodal head mouse rated at 4.2 and monomodal eye mouse rated at 5.5 – hence these are also shown in **bold**.

The results for the enhanced device were very similar to the standard devices, as discussed previously (Chapter 10.10), with the ratings for the head mice remaining very similar between standard and enhanced devices. Also of note were that there were no changes for the two poorest ranking results (highlighted in bold) between devices for the monomodal eye mouse in standard or enhanced operation, indicating that the enhancement to the monomodal eye mouse did not improve satisfaction.

The reductions for the enhanced multimodal eye mouse over the standard multimodal eye mouse in Physical, Temporal and Frustration factors and the increase in Performance (inverted) Workload factor, together with increases in Pointing Accuracy and System Control ratings for the Ease of Use factors indicated that the addition of the enhancement was subjectively effective for the multimodal eye mouse.

This improvement in subjective satisfaction led to the enhanced multimodal eye mouse producing three changes for the two poorest rankings caused by the addition of the zoom enhancement (highlighted in italic on Table 13.16). Here, in comparison with the enhanced multimodal head mouse, the larger effective object sizes generated by the enhancement lifted the subjective Performance and System Control of the multimodal eye mouse to near parity with the head mouse, and also increased subjective Pointing Accuracy to parity with the head mouse.

In summary:

- *The enhancement did not substantially change the individual satisfaction ratings between the standard and enhanced head mice*
- *The enhancement had little effect on the ratings of the monomodal eye mouse*
- *Dissatisfaction with monomodal dwell selection could not be overcome by the enhancement*
- *Enhancement had the greatest subjective effect on the multimodal eye mouse, with the eye mouse reaching the level of satisfaction of the head mouse for Performance, System Control and Pointing Accuracy*

13.9 A summary of the performance of enhanced eye and head mice

Reviewing the success of the predictions outlined at the start of this chapter showed that six of the ten were proven. These results are summarised:

<u>**Results of Predictions:**</u>

1. *No difference* in performance *between word and internet task domains* for the enhanced head and eye mice as the tasks in each domain were unchanged – **FALSE**

2. *The benefit* of *increased overall* enhanced head and eye mouse *performance* due to the effective target sizes of objects in the test tasks being increased by the enhancement thus reducing the effects of the low pointing accuracy of the head and eye mice – **TRUE**

3. *The benefit* of a *reduction* in *control correction times* above other task time elements due to the effective target sizes of objects in the test tasks being increased by the enhancement – **TRUE**

4. *The cost* of an *increase in overall task time* due to the addition of zoom control times when increasing target object sizes with the enhancement – **FALSE**

5. *The benefit* of an *increase* in interaction *task quality*, with a *marked reduction in control corrections* above other errors due to the effective target sizes of objects in the test tasks being increased by the enhancement – **TRUE**

6. *The cost* of a *decrease* in interaction *task quality* with the addition of *zoom level corrections* when controlling the zoom enhancement for the enhanced devices – **FALSE**

7. *The benefit* of an *increase* in overall device *satisfaction* for the *enhanced devices* as the effective target sizes of objects in the test tasks are increased by the enhancement and are hence easier to select – **TRUE**

8. *The cost* of an *increase* in overall *workload* due to the additional complexity of controlling the zoom enhancement for the enhanced devices – **FALSE**

9. *No change* in device *comfort* for the enhanced devices as the devices themselves have not substantially changed by the addition of the zoom enhancement – **TRUE**

10. *The benefit* of an increase in *ease of use* for the enhanced devices as the effective target sizes of objects in the test tasks are increased by the enhancement and are hence easier to select – **TRUE**

Prediction 1 did not hold as unlike the standard devices there were differences in task domain performance for the enhanced eye mouse, with typing tasks being more efficient for the multimodal eye mouse due to speed of selection, and restricted object manipulation being more efficient for the monomodal eye mouse due to automate dwell selection. *Prediction 4* did not hold as the time taken to control the zoom was not significant and the benefits of the zoom compensated for this small additional time. *Prediction 6* did not hold as there were extremely few zoom level corrections and the task quality benefit of the enhancement outweighed any corrections. Finally, *Prediction 8* did not hold as controlling the enhancement did not produce any additional workload as perhaps the control was found to be simple.

Overall, the enhancement increased device satisfaction but the enhanced eye mice were still less satisfying to use than the head mice, and both the head and eye mice were still less satisfying to use than the baseline hand mouse. Examining efficiency found that the enhancement disproportionately increased eye mouse performance in comparison to head mouse performance, resulting in equality of performance between the enhanced monomodal head and eye mice, and also equality of performance between the enhanced multimodal head and eye mice. This equality of performance between head and eye showed that the enhancement now removed the device differences, allowing users to potentially choose either device for similar performance. However, neither enhanced head nor eye devices approached the performance of the baseline hand mouse.

Chapter 14

A Detailed Examination of Enhanced Eye and Head Mouse Performance

The aim of this chapter is to continue, and conclude, the investigation into the effects of adding the target magnification enhancement to head and eye mice. The previous chapter, Chapter 13, followed the analysis of Chapter 10 and investigated the overall performance of the head and eye mice but with the addition of enhancement. Chapter 13 concluded that the enhancement was beneficial but did not sufficiently increase overall head and eye mouse performance to the level of the baseline hand mouse. This chapter will build on the previous detailed examination of head and eye mouse performance by following the analysis of Chapter 11, and investigate how enhancement interacts with the major factors that influence and maximise head and eye mouse performance – target size, interaction technique and subject experience. This chapter first predicts the effects of the enhancement on performance in terms of target size, interaction technique, and subject experience, and finally on the combination of all three. The chapter then goes on to determine if each of these predictions is in turn valid. Finally, this chapter goes on to determine the highest performance of the enhanced head and eye mice against the performance of the baseline hand mouse, addressing, within the bounds of this work, the main theme of this work by determining to what extent can a eye and head mouse, coupled with the enhancement produced within this work, achieve the same performance as the benchmark hand mouse on an unmodified graphical user interface.

14.1 Predicted effects of enhancement

Chapter 11 of this work examined the detailed performance of head and eye mice, in both monomodal and multimodal configurations, against the baseline of a desktop hand mouse. The chapter showed that neither device could approach the performance of the hand mouse, but this chapter did show that if the effective target object sizes on the interface could be increased (by employing the magnification enhancement) then the performance of the both head and eye mice could be increased substantially. The chapter also showed that subject experience had a great influence on head and eye mouse performance. Finally, Chapter 11 showed that a combination of these factors of larger target sizes and high subject experience could potentially raise the performance of the head and eye mice to equal the hand mouse performance. Hence, from the results of Chapter 11 it was predicted that adding a magnification enhancement would result in the following overall effects, with this chapter addressing each of these predictions in numerical order and with each prediction building on the results of the previous prediction:

11. *The benefit* of increased *task efficiency* for *smaller target sizes* (from Chapter 11.1).

12. Magnification levels used to give *effective target sizes sufficiently large* to achieve *parity of performance* with the baseline *hand mouse* (from Chapter 11.1).

13. That the magnification enhancement will be used such that *effective target sizes* are *constant*, regardless of the original target size (from Chapter 11.1).

14. The effect of a *supporting modality* would be maintained *irrespective of effective target size due to enhancement*, with monomodal selection being less efficient that multimodal selection (from Chapter 11.1).

15. The effect of *differing interaction techniques* would be maintained *irrespective of effective target size*, with task efficiency reducing with increasing complexity of manipulation (from Chapter 11.4).

16. The progression of *increasing performance* with *increasing subject experience* would be *maintained* and not affected by the enhancement (from Chapter 11.5).

17. The effect of a *supporting modality* would be maintained *irrespective of subject experience*, with monomodal selection being less efficient that multimodal selection (from Chapter 11.5).

18. The addition of the *maximum available number of supporting modalities* controlling any device would result in the *highest performance* for any given level of user experience (from Chapter 11.1).

19. Finally, for *high experience subjects* with the *maximum number of supporting modalities* the performance of the eye mouse would potentially be increased sufficiently that for large target sizes requiring less magnification, the eye mouse would outperform the head mouse and approach the baseline hand mouse performance (from Chapter 11.8).

14.1 Enhancement, target size and device efficiency

This section examines the effect enhancement had on the relationship between target size and efficiency, and extends the previous analysis on the standard devices (Chapter 11.1). Breaking down the data by the four target size categories in the test method (Chapter 4: S1=0.3°, S2=0.6°, S3=0.9°, S4=1.2° subtended visual arc at 60cm) suggested a relationship between standard device efficiency and target size (Chapter 11, Figure 11.1 repeated in Figures 14.1 and 14.2), with efficiency increasing with target size. The same analysis for the enhanced head and eye mice showed the same relationship (Appendices Table A14.1, Graphed in Figures 14.1 and 14.2).

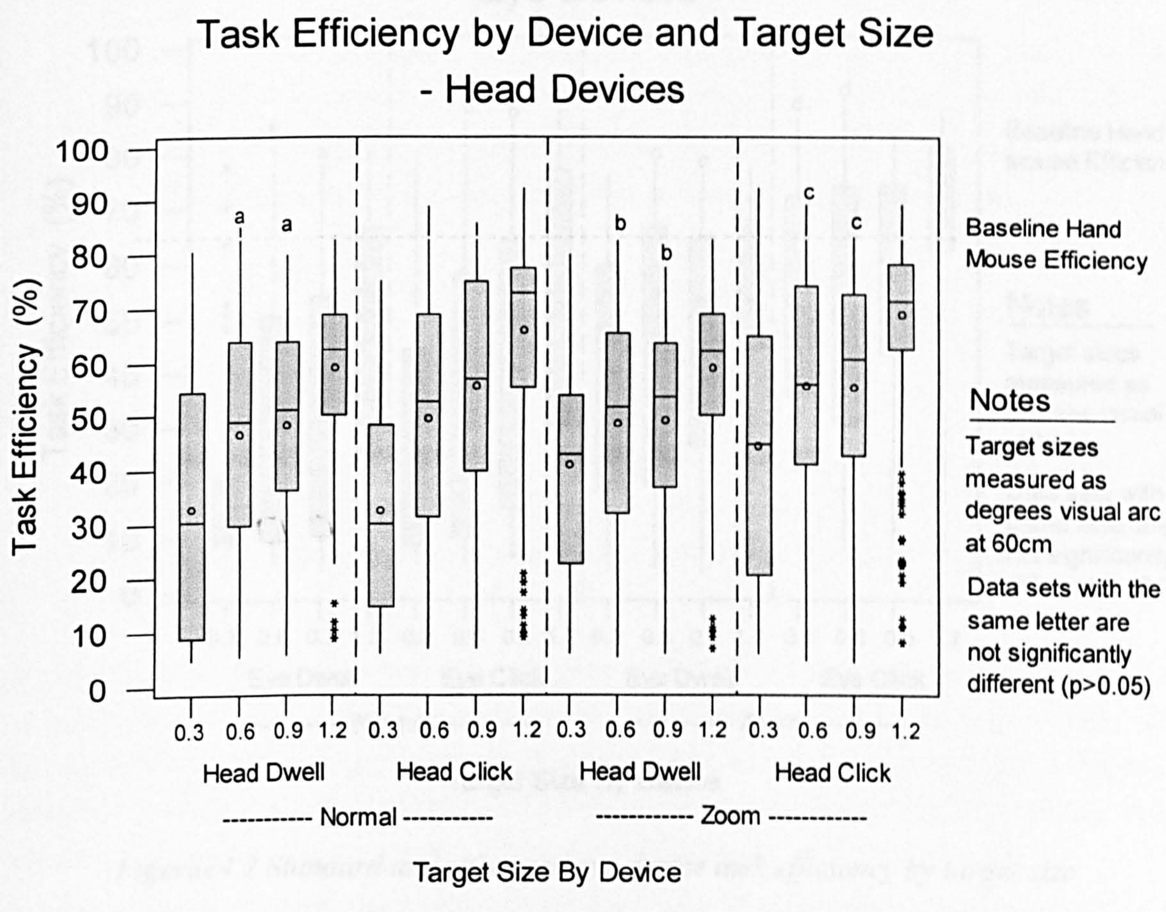


Figure14.1 Standard and enhanced head device task efficiency by target size

Here the statistical significance of differences between the results within each device were calculated (Mann-Whitney two-sample rank tests, Appendices Table A14.2). This confirmed that there was a progression of efficiency with increasing target size, from the smallest target size, S1 = 0.3°, having the lowest efficiency to the largest target size, S4 = 1.2°, having the greatest efficiency. Although the smallest and largest targets for all enhanced devices had efficiencies significantly different from each other, the two middle-sized targets, except the multimodal eye mouse, did not. However, as with the standard devices, there was a progression of increasing efficiency from S1 through S2/3 to S4 for the enhanced devices.

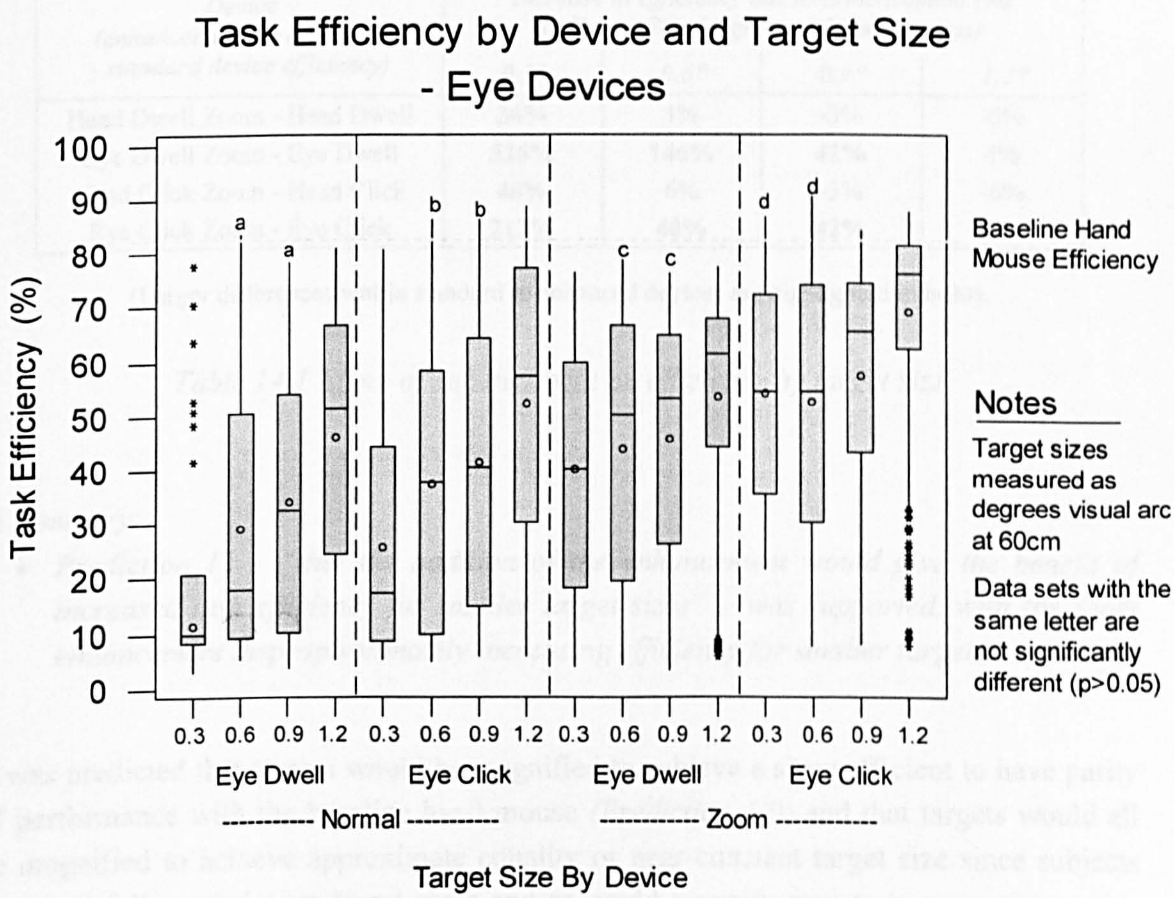


Figure14.2 Standard and enhanced eye device task efficiency by target size

It was predicted that the addition of the enhancement would have a greater benefit of increased task efficiency for smaller target sizes *(Prediction 11)* as these smaller targets could now be magnified to aid selection and manipulation. Analysing task efficiency by target size and calculating the percentage change in efficiency between standard and enhanced devices for each target size category showed that smaller target sizes did benefit disproportionably, with considerable increases in efficiency due to the enhancement

(Table 14.1). In addition, the effect of the enhancement was greater for the eye mice than the head mice. This was due to the low pointing accuracy of the eye mouse in comparison to the head mouse, with the eye mice gaining greater benefit from the enhancement for the smaller target sizes than the head mice. This benefit for the eye mice is highlighted in bold (Table 14.1).

Enhancement, Target Size and Efficiency				
Device (enhanced device efficiency - standard device efficiency)	*Increase in Efficiency due to Enhancement (%) at Target Size (degrees at 60cm distance)*			
	0.3°	*0.6°*	*0.9°*	*1.2°*
Head Dwell Zoom - Head Dwell	**36%**	1%	-3%	-5%
Eye Dwell Zoom - Eye Dwell	**326%**	**146%**	**42%**	4%
Head Click Zoom - Head Click	**48%**	6%	-3%	-6%
Eye Click Zoom - Eye Click	**212%**	**40%**	**42%**	**20%**

(Larger differences within standard to enhanced devices are highlighted in **bold**).

Table 14.1 Effect of enhancement on efficiency by target size

In summary:

- *Prediction 11 – "that the addition of the enhancement would give the benefit of increased task efficiency for smaller target sizes" – was supported, with the zoom enhancement disproportionately increasing efficiency for smaller targets.*

It was predicted that targets would be magnified to achieve a size sufficient to have parity of performance with the baseline hand mouse *(Prediction 12)*, and that targets would all be magnified to achieve approximate equality or near constant target size since subjects now had full control over target sizes and so could magnify targets to a consistent size suitable for reliable selection and manipulation with that device *(Prediction 13)*.

To determine if these predictions were valid, the effective target sizes after magnification were extracted from the test data and the increases in target sizes calculated (Table 14.2). The results showed that targets were magnified to a near constant size (Table 14.2) but that this size was not sufficiently large to allow performance equal the hand mouse baseline performance (Figure 14.2). Clearly there was a reluctance to use the enhancement to its full extent and magnify targets sufficiently to achieve a high level of performance.

After Magnification Target Sizes				
Original target size	*Actual effective zoomed target size used by the enhanced devices (Degrees at 60cm)*			
	Head Dwell	*Eye Dwell*	*Head Click*	*Eye Click*
0.3°	0.95°	1.57°	0.73°	1.61°
0.6°	0.88°	1.81°	0.88°	1.67°
0.9°	0.97°	1.88°	1.02°	1.82°
1.2°	1.20°	1.81°	1.22°	1.72°
Mean:	*1.00°*	*1.77°*	*0.96°*	*1.71°*
S.D.:	*0.14°*	*0.14°*	*0.21°*	*0.09°*

Table 14.2 After-magnification target sizes

This suggested that subjects, perhaps due to unfamiliarity with the enhancement, used the enhancement to achieve an *acceptable* level of performance with either the head or eye mice, but were reluctant to progress further.

In summary:

- *<u>Prediction 12</u> – "that the addition of the enhancement would allow subjects to magnify targets sufficiently to equal baseline hand mouse performance" – was not supported, with subjects not using sufficient magnification to achieve parity*

In summary:

- *<u>Prediction 13</u> – "that the addition of the enhancement would allow subjects to magnify targets to achieve approximate equality or near constant target size" – was supported, with subjects showing a consistent after magnification target size*

The effect of the enhancement on the differences between the monomodal and multimodal devices was investigated. Previous analysis (Chapter 11.1) had showed that the selection method had an effect on efficiency, with within device monomodal operation achieving lower efficiency than multimodal operation for any given single target size. It was predicted *(Prediction 14)* that this effect would be shown and hold for the enhanced devices. The results (Figure 14.2) showed that this prediction held for the enhanced devices, with the enhanced monomodal devices achieving lower efficiency for any given target size than the enhanced multimodal devices.

In summary:

- <u>*Prediction 14*</u> – *"that the effect of a supporting modality would be maintained irrespective of effective target size due to enhancement, with monomodal selection being less efficient that multimodal selection" – was supported*

14.2 Enhancement, interaction technique and device efficiency

The effects of the enhancement on the differing interaction techniques (Chapter 4) present within the test tasks on the level of device efficiency were investigated to determine if the enhancement and type of object manipulation had an effect on device efficiency. It was predicted *(Prediction 15)* that the effect of differing interaction techniques would be maintained irrespective of effective target size, with task efficiency reducing with increasing complexity of manipulation.

Device	*Interaction Technique Efficiency (%)*				
	--------- Restricted ---------			*---- Unrestricted ----*	
	Single	*Double*	*Drag*	*Single*	*Drag*
Hand	81	76	65	83	61
Head Dwell	**48**	49	36	**60**	34
Eye Dwell	47	42	12	45	9
Head Click	**57**	50	31	**69**	40
Eye Click	**31**	30	27	**53**	17
Head Dwell Zoom	54	48	34	60	33
Eye Dwell Zoom	59	52	28	60	36
Head Click Zoom	66	49	34	71	40
Eye Click Zoom	**52**	54	45	**72**	51

(Big differences within devices and equivalent techniques are highlighted in **bold**).

Table 14.3 Effect of interaction technique and enhancement on device efficiency[1]

[1] As described in Chapter 4, there are three basic forms of pointing device object manipulation on a Windows interface. These are a single click on an object, a double click on an object, and a drag of an object. These actions can be either *restricted*, where cursor movement is confined within the area of the object to retain control of that object, or *unrestricted* where cursor movement may move from the object and then return to the object and still retain control of that object.

For example, a button object on the Windows interface has *unrestricted* manipulation as, provided the mouse button is held down, the cursor may be moved away from the button and returned without losing permanent control of the button. In contrast, a hypertext link has *restricted* manipulation, as any movement of the cursor away from the object with the mouse button held down will lose control of the object.

As before (Chapter 11.4) task efficiency was broken down by interaction technique for each device (Appendices Table A14.3, summarised in Table 14.3) and the statistical significance of differences between interaction types within each device were calculated (Mann-Whitney two-sample rank tests, Appendices Table A14.4).

Examining the individual interaction techniques across the enhanced devices showed that the general progression of efficiency reducing with the increased complexity of interaction technique followed the same pattern as the standard devices, with single clicks showing the highest efficiency, through double clicks to dragging showing the lowest efficiency. As determined previously for the standard devices (Chapter 11.4) unrestricted single click manipulation was also more efficient than restricted single click manipulation for the enhanced devices. This indicated that objects that required restricted single click manipulation still exhibited lower interaction efficiency than objects that gave unrestricted single click manipulation. This was even with the ability to magnify these target objects to any size, and hence potentially reduce the manipulation difficulties of restricted objects by increasing their size sufficiently that it would be difficult to accidentally move the cursor from those objects. Hence, the finding for standard devices (Chapter 11.4) that interfaces designed for head and eye manipulation should use unrestricted in preference to restricted object manipulation characteristics, also held for the enhanced devices.

In summary:

- *<u>Prediction 15</u> – "that the effect of differing interaction techniques would be maintained irrespective of effective target size, with task efficiency reducing with increasing complexity of manipulation" – was supported*
- *Unrestricted manipulation is more efficient than restricted manipulation irrespective of enhancement*

14.3 Enhancement and subject experience

Breaking down the test results by the three subject experience ratings (Chapter 10) for the enhanced devices showed the relationship between increasing subject experience and increasing performance found for the standard devices (Chapter 11.5) was also maintained for the enhanced devices (Appendices Table A14.5, Graphed in Figures 14.1 and 14.2). Examining the statistical significance of the results (Appendices Table A14.6, Summarised in Table 14.4) showed that, even with the benefit of the enhancement of the head and eye mice and high experience subjects, the baseline hand mouse outperformed all devices. However, the disproportionate benefit of the enhancement for the eye mouse coupled with high subject experience brought the high experience enhanced eye mouse to within 4.4 percentage points of the baseline hand mouse. Of note was that the enhanced eye mouse now, for the first time, exceeded the performance of the enhanced head mouse when subjects were either medium experience subjects or high experience subjects.

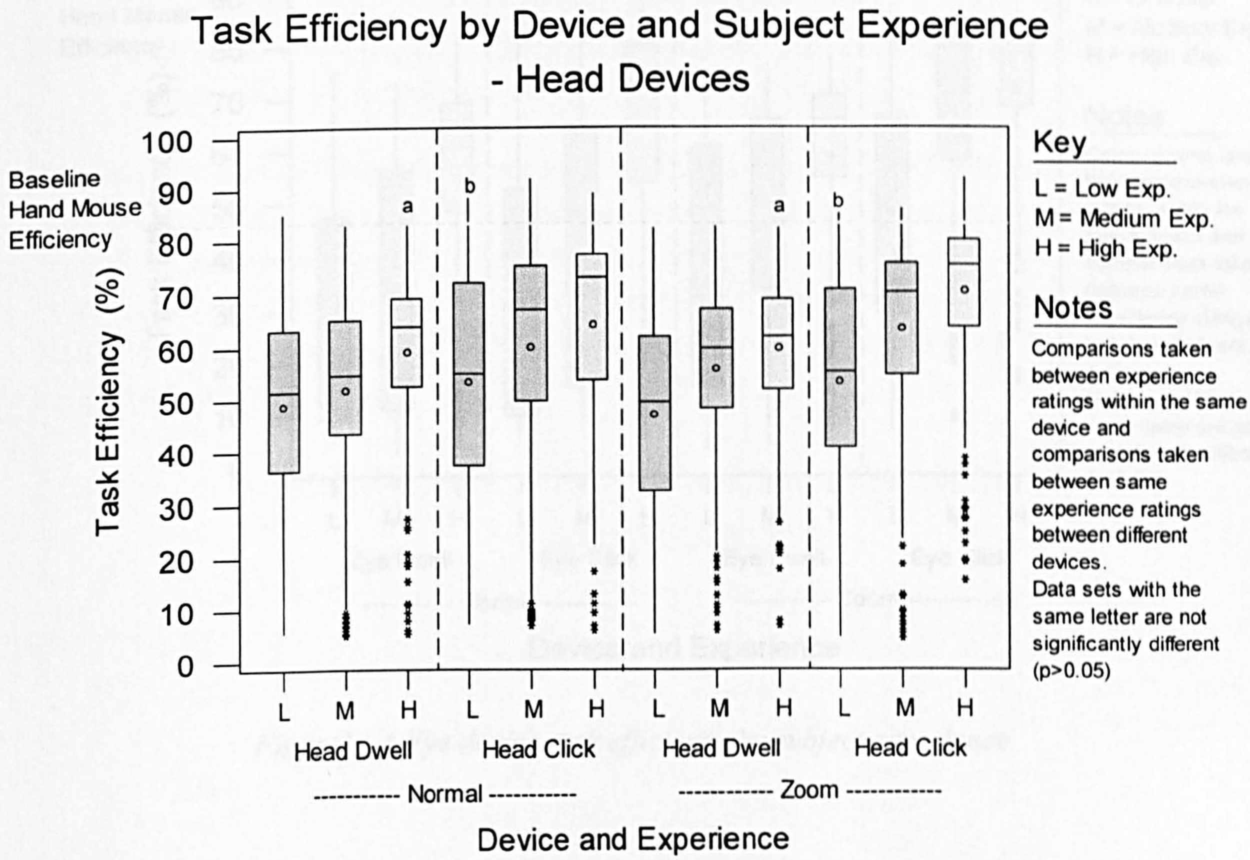


Figure14.3 Head device task efficiency by subject experience

The enhancement particularly benefited low and medium experience eye mouse subjects with near constant gains of 24.4 (monomodal) and 27.4 (multimodal) percentage points between the standard and enhanced low experience eye mouse subjects, and 25.3 (monomodal) and 25.7 (multimodal) percentage points between the standard and enhanced medium experience eye mouse subjects. In comparison, the enhancement only resulted in a benefit of 5.1 (monomodal) and 5.4 (multimodal) percentage points for the high experience eye mouse subjects. These results showed that the enhancement could reduce the amount of subject experience with the eye mouse that was required to achieve an acceptable level of performance with the device, and hence the enhancement could be used to reduce training times and increase user uptake of eye mice.

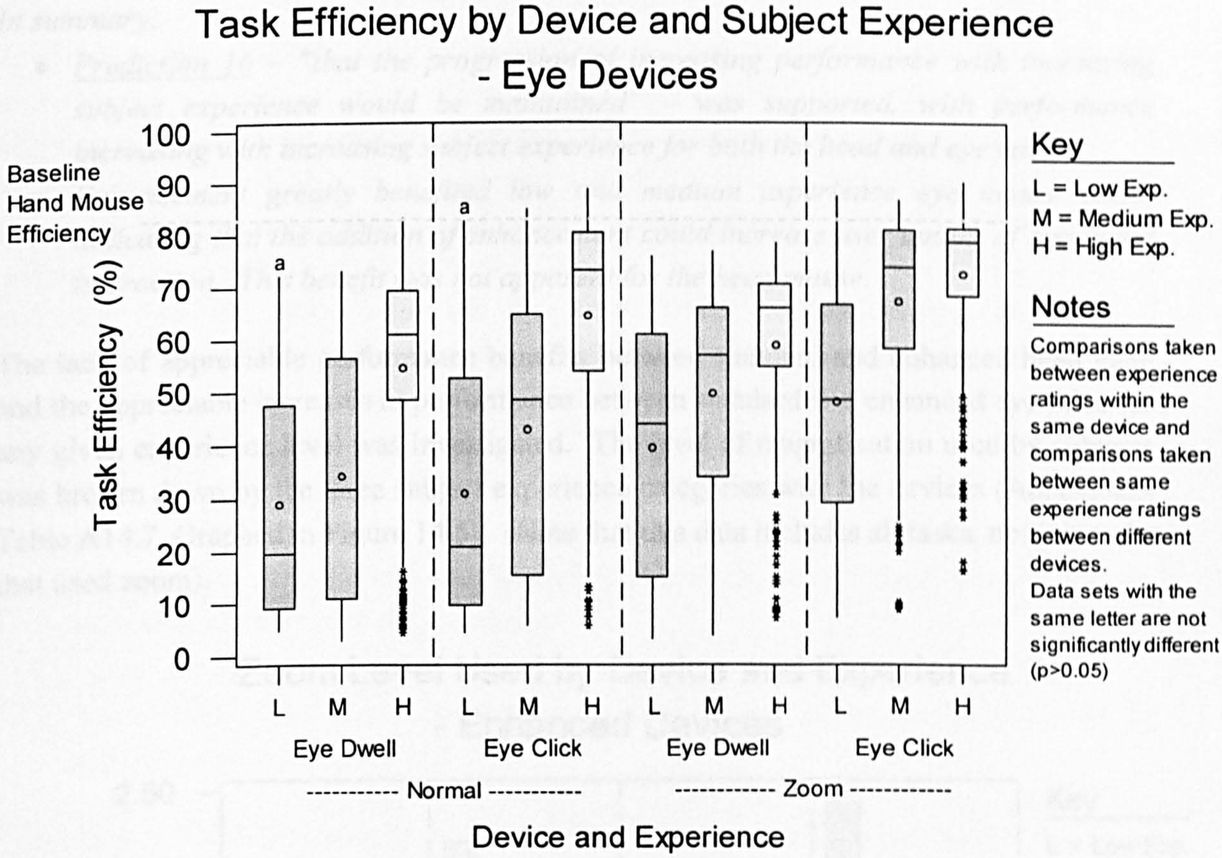


Figure14.4 Eye device task efficiency by subject experience

Subject Experience and Device Efficiency			
Device	*Efficiency (%) by subject experience (Low, Medium, High)*		
	L	*M*	*H*
Hand	-	-	83.3
Head Dwell	51.6	54.5	63.9[a]
Eye Dwell	19.7[c]	31.8	61.1[a]
Head Click	55.0[e]	66.9	73.0[b]
Eye Click	21.2[c]	48.4	73.5[b]
Head Dwell Zoom	49.4[f]	59.7	62.0[d]
Eye Dwell Zoom	44.1	57.1	66.2[d]
Head Click Zoom	55.0[e]	70.2	75.3
Eye Click Zoom	48.6[f]	74.1	78.9

Data with the same letter are not significantly different ($p > 0.05$)

Table 14.4 Effect of subject experience and enhancement on device efficiency

In summary:

- *<u>Prediction 16</u> – "that the progression of increasing performance with increasing subject experience would be maintained" – was supported, with performance increasing with increasing subject experience for both the head and eye mice*
- *Enhancement greatly benefited low and medium experience eye mouse users, indicating that the addition of enhancement could increase user uptake of eye-based interaction. This benefit was not apparent for the head mouse.*

The lack of appreciable performance benefits between standard and enhanced head mice and the appreciable increases in performance between standard and enhanced eye mice for any given experience level was investigated. The level of magnification used by subjects was broken down by the three subject experience categories with the devices (Appendices Table A14.7, Graphed in Figure 14.5). (Note that this data includes all tasks, not just tasks that used zoom).

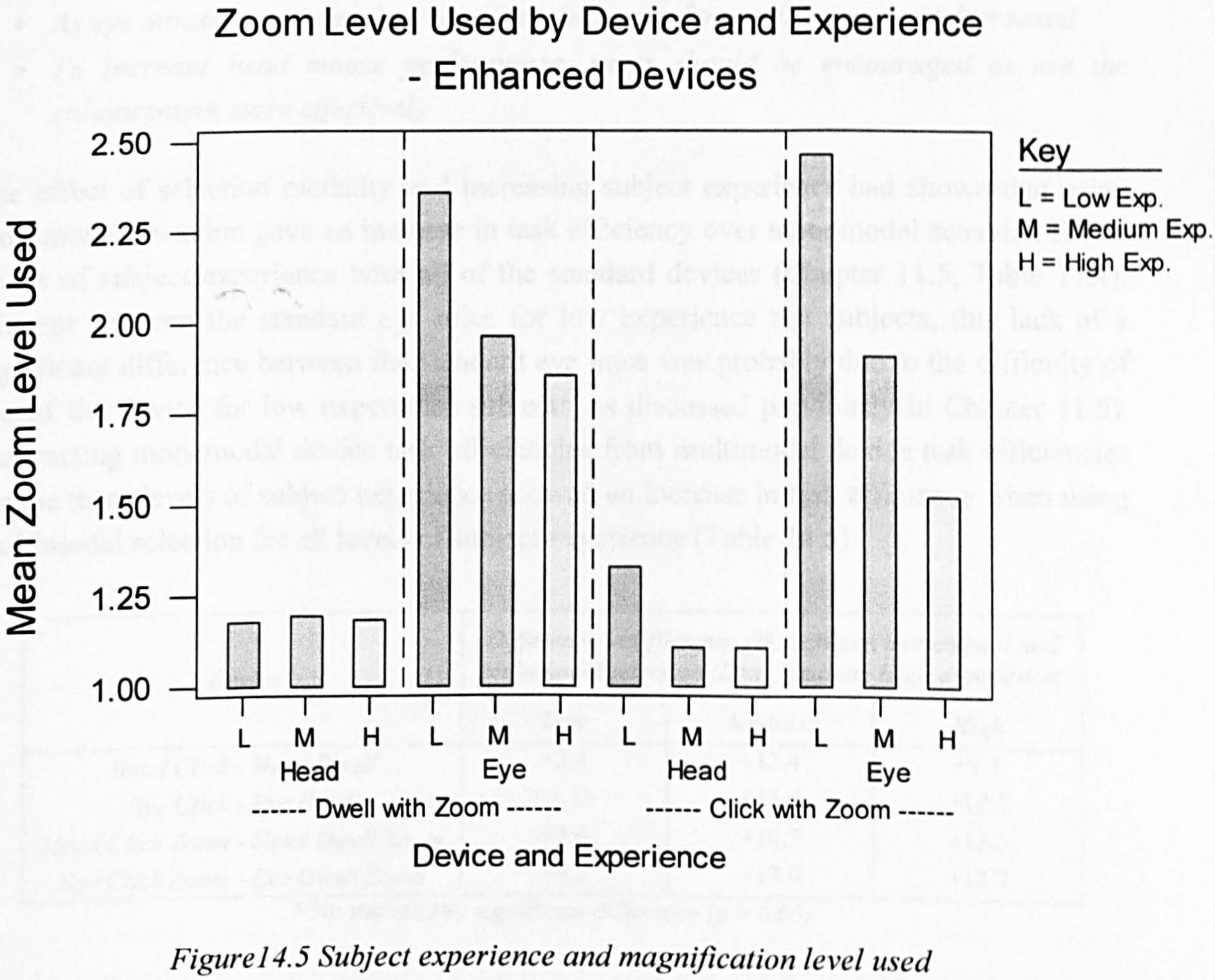


Figure14.5 Subject experience and magnification level used

This (Figure 14.5) showed that the level of magnification used with the head mouse was consistently low, possibly as subjects felt that the head mouse gave sufficient pointing accuracy to not require target object magnification, and hence the benefits of the enhancement were not realised for the head mouse. However, high levels of magnification were used with the eye mouse, with the level of magnification falling with increasing subject experience. This indicated that subjects felt the need to use the enhancement, possibly as subjects felt that the eye mouse exhibited poor pointing accuracy, but that with increased experience with the eye mouse, the level of magnification required was reduced but still remained higher than the head mouse. The results suggested that head mouse users should be encouraged to use the enhancement more to increase the potential performance of the head mouse.

In summary:

- *Subjects did not fully use the enhancement with the head mouse, but did use the enhancement fully with the eye mouse*
- *As eye mouse experience increased, so the level of magnification used decreased*
- *To increase head mouse performance, users should be encouraged to use the enhancement more effectively*

The effect of selection modality and increasing subject experience had shown that using multimodal selection gave an increase in task efficiency over monomodal selection for all levels of subject experience with all of the standard devices (Chapter 11.5, Table 11.4). (Except between the standard eye mice for low experience test subjects, this lack of a significant difference between the standard eye mice was probably due to the difficulty of use of the device for low experience subjects, as discussed previously in Chapter 11.5). Subtracting monomodal device task efficiencies from multimodal device task efficiencies for the three levels of subject experience showed an increase in task efficiency when using multimodal selection for all levels of subject experience (Table 14.5).

Device	*Difference in Efficiency (%) between monomodal and multimodal selection (Low, Medium, High experience)*		
	Low	*Medium*	*High*
Head Click - Head Dwell	+3.4	+12.4	+9.1
Eye Click - Eye Dwell	+1.5*	+16.6	+12.4
Head Click Zoom - Head Dwell Zoom	+5.6	+10.5	+13.3
Eye Click Zoom - Eye Dwell Zoom	+4.5	+17.0	+12.7

* No statistically significant difference ($p > 0.05$)

Table 14.5 All device difference in efficiency by selection modality and experience

In summary:

- Prediction 17 – *"that the effect of a supporting modality would be maintained irrespective of subject experience, with monomodal selection being less efficient that multimodal selection" – was supported, with monomodal performance being, overall, less efficient than multimodal performance for both the head and eye mice*

14.4 Choice of supporting modalities

The consistent penalty of monomodal selection, even when a device has the magnification enhancement added, gives rise to the question that if a user has a limited number of modalities available to control either selection or magnification, should the user control selection or magnification for the best possible performance? For instance, given one or two supporting modalities for either selection or zoom or both, which combination should be used? In addition, would this hold for all subject experience groups? It was predicted *(Prediction 18)* that the addition of the maximum number of supporting modalities controlling any device would result in the highest performance, irrespective of user experience. To determine this, the results for the devices were broken down by the number of controlling modalities required by each permutation of device. To do this, the number of modalities available, and hence control given over the devices, were illustrated by a matrix (Table 14.6).

<table>
<tr><th colspan="4">Available modalities and control options</th></tr>
<tr><td colspan="2" rowspan="2">Number of user modalities required</td><td colspan="2">Enhancement</td></tr>
<tr><td>Standard No Zoom</td><td>Enhanced Zoom</td></tr>
<tr><td rowspan="2">Selection</td><td>Monomodal Dwell</td><td>0</td><td>1</td></tr>
<tr><td>Multimodal Click</td><td>1</td><td>2</td></tr>
</table>

Table 14.6 Supporting modalities and control options

Here (Table 14.6) for example, the matrix shows that no additional modalities (such as control over a switch) are required for standard monomodal operation, but two modalities are required for enhanced multimodal operation (such as control over a switch for selection, and control over an additional switch for changing the zoom level). Next a matrix was completed with the task efficiency and rank in brackets of each device at each subject experience level (Tables 14.7 and 14.8).

Head Mouse – Low Experience			
Efficiency and (Rank)		Enhancement	
		Standard No Zoom	Enhanced Zoom
Selection	Monomodal Dwell	51.6 (3)	49.4 (4)
	Multimodal Click	55.0 (1)	55.0 (1)

Head Mouse – Medium Experience			
Efficiency and (Rank)		Enhancement	
		Standard No Zoom	Enhanced Zoom
Selection	Monomodal Dwell	54.5 (4)	59.7 (3)
	Multimodal Click	66.9 (2)	70.2 (1)

Head Mouse – High Experience			
Efficiency and (Rank)		Enhancement	
		Standard No Zoom	Enhanced Zoom
Selection	Monomodal Dwell	63.9 (3)	62.0 (4)
	Multimodal Click	73.0 (2)	75.3 (1)

Table 14.7 Head mouse performance and control options

It was expected that the multimodal enhanced devices would all have the highest performance and this was found to be true for all experience levels. Thus where possible, subjects should always use a supporting modality for selection if available, and also a further supporting modality for magnification control. It was also expected that given only one supporting modality, this should be used for multimodal selection as monomodal selection had a consistent performance penalty (Prediction 14).

This was the case for the head mice and the high experience eye mouse, but not for the medium and low experience eye mice. Here the use of the zoom enhancement resulted in a higher performance than controlling the selection method when only one supporting modality was available. This was probably due to the zoom enhancement compensating for the poor inherent pointing inaccuracy of the eye mouse by making targets larger, more than the ability to accurately time selection attempts when the eye mouse was used by low and medium experience subjects.

Eye Mouse – Low Experience			
Efficiency and (Rank)		Enhancement	
		Standard No Zoom	Enhanced Zoom
Selection	Monomodal Dwell	19.7 (4)	44.1 (2)
	Multimodal Click	21.2 (3)	48.6 (1)

Eye Mouse – Medium Experience			
Efficiency and (Rank)		Enhancement	
		Standard No Zoom	Enhanced Zoom
Selection	Monomodal Dwell	31.8 (4)	57.1 (2)
	Multimodal Click	48.4 (3)	74.1 (1)

Eye Mouse – High Experience			
Efficiency and (Rank)		Enhancement	
		Standard No Zoom	Enhanced Zoom
Selection	Monomodal Dwell	61.1 (4)	66.2 (3)
	Multimodal Click	73.5 (2)	78.9 (1)

Table 14.8 Eye mouse performance and control options

Finally it was expected that the use of no supporting modalities would result in the lowest performance. This was true for all devices and experience levels except for the low experience head mouse where controlling the zoom enhancement with monomodal dwell selection actually resulted in lower performance than using monomodal dwell selection on its own. There was only a small but significant 2.2 percentage points ($p < 0.050$)

difference between the results. This reversal in the expected result could be explained by the additional complexity of controlling the enhancement whilst also controlling the dwell click device reducing performance for novice users.

In summary:

- *Prediction 18 – "that the addition of the maximum number of supporting modalities controlling any device would result in the highest performance for any given level of experience" – was supported, with two modes of control, one for selection and one for magnification, being most efficient for any given level of experience*
- *For low and medium experience eye mouse users with a single available control modality; this was best used controlling magnification rather than selection. For all other devices and experience levels, a single control modality was best used controlling selection*

14.5 Enhancement, target size and subject experience

The maximum achievable performance of the head and eye mice was now investigated. It had been predicted that for high experience subjects with the maximum number of supporting modalities, the performance of the eye mouse would potentially be increased sufficiently that for large target sizes requiring less magnification, the eye mouse would outperform the head mouse and approach the baseline hand mouse performance *(Prediction 19).* To determine this predicted maximum performance, and to show if this work did sufficiently increase the performance of head and eye mice to the baseline of a hand mouse, the results for the multimodal enhanced head and eye mice were analysed by target object size for high experience subjects only (Appendices Table A14.8, Graphed Figure 14.6).

The statistical comparisons (Appendices Table A14.9) showed that the hand mouse outperformed the enhanced devices even when the enhanced devices were manipulating larger target sizes (requiring less use of the enhancement and hence giving less zoom time and zoom quality penalties) and when they were operated by high experience subjects. However, the differences between the head and eye mice devices and the hand mouse baseline for each target size were dramatically reduced. This is illustrated by subtracting the enhanced multimodal high experience subject only head and eye mice task efficiencies from the baseline hand mouse efficiency for each target size (Table 14.8).

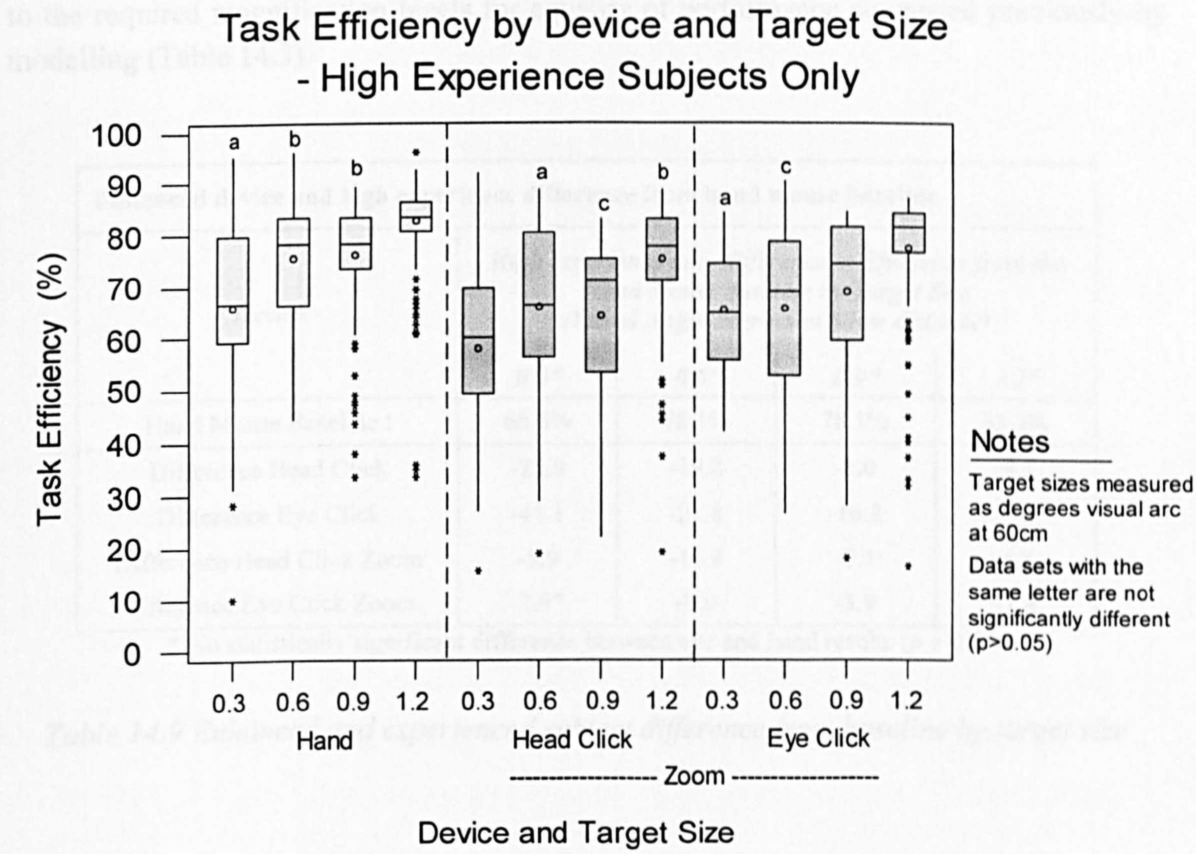


Figure14.6 Enhanced device task efficiency for high experience subjects by target size

A highly significant result was found from this analysis - **for high experience users, the enhanced multimodal eye mouse outperforms the enhanced multimodal head mouse for all target sizes.** In addition the eye mouse reaches to within a few percentage points of the hand mouse baseline performance for all target sizes, and equals the hand mouse performance for the smallest target size (S1 = 0.3°) (Mann-Whitney two-sample rank test $p = 0.575$). It is this result that is most interesting, where eye mouse subjects used the enhancement to magnify the smallest target object sufficiently that it could be manipulated more efficiently than using the hand mouse. Clearly the smallest object would attract the most willingness to use the enhancement to aid selection. This suggested that if both head and eye mouse users could be encouraged to use the enhancement more and to higher magnification levels (from Figure 14.5) then both devices might equal the hand mouse baseline for any target size. This might be accomplished by providing a magnification facility that was more automated and easier to use than a manual series of button presses; perhaps by sensing the size of target objects and automatically zooming in onto the object to the required magnification levels for equality of performance suggested previously by modelling (Table 14.3).

Enhanced device and high experience difference from hand mouse baseline				
Device	*High experience only difference in Efficiency from the Hand Mouse baseline by Target Size (Visual Angle degrees at 60cm distance)*			
	0.3°	*0.6°*	*0.9°*	*1.2°*
Hand Mouse Baseline :	66.6%	78.1%	78.1%	83.3%
Difference Head Click	-20.8	-19.2	-8.0	-8.1
Difference Eye Click	-41.1	-21.8	-16.8	-5.2
Difference Head Click Zoom	-5.9	-11.8	-7.1	-5.6
Difference Eye Click Zoom	-2.9*	-9.0	-3.9	-1.8

* No statistically significant difference between eye and hand results ($p > 0.05$)

Table 14.9 Enhanced and experienced subject difference from baseline by target size

In summary:

- *Prediction 19 – "that for high experience subjects with the maximum number of supporting modalities the performance of the eye mouse would potentially be increased sufficiently that for large target sizes requiring less magnification, the eye mouse would outperform the head mouse and approach the baseline hand mouse performance" – was supported, with the eye mouse exceeding the head mouse and approaching, and equalling for the smallest target size, the performance of the hand mouse*

14.6 Enhancement, device satisfaction and subject experience

As before (Chapter 11.9) the questionnaires were administered after a complete test had been completed, rather than after each task had been completed (which would be impractical) hence it was not possible to separate satisfaction by individual task properties, such as target size or interaction technique. However, the effect of test subject experience and the effect on enhancement with the devices on the satisfaction questionnaire categories could be investigated, by comparing the ratings and rating differences for all subjects to the ratings of the high experience category subjects on both the standard and enhanced devices (Table 11.7). Note that the largest differences are shown in bold type.

Device	*Satisfaction Questionnaire response (1-7)*								
	Workload (low=good)			*Comfort (high=good)*			*Ease of Use (high=good)*		
Experience	*All*	*High*	*Diff.*	*All*	*High*	*Diff.*	*All*	*High*	*Diff.*
Hand	1.8	1.8	-	6.5	6.5	-	6.3	6.3	-
Head Dwell	4.1	4.1	-	5.9	6.2	+0.3	3.5	3.5	-
Eye Dwell	5.8	4.2	-1.0	4.5	5.5	+1.0	2.6	3.5	+0.9
Head Click	3.8	3.7	-0.1	6.3	6.4	+0.1	4.3	4.6	+0.1
Eye Click	5.1	4.0	-1.1	5.2	5.9	+0.7	4.4	4.7	+0.3
Head Dwell Zoom	3.8	3.8	-	5.9	6.3	+0.4	3.9	3.6	-0.3
Eye Dwell Zoom	5.6	5.2	-0.4	4.7	4.9	+0.2	3.4	3.6	+0.2
Head Click Zoom	3.8	2.6	**-1.2**	6.1	6.3	+0.2	4.6	4.6	-
Eye Click Zoom	4.4	3.4	**-1.0**	5.2	5.7	+0.5	5.1	5.3	+0.2

(Larger differences within devices are highlighted in **bold**).

Table 14.10 Satisfaction factors for all subjects against high experience subjects

Examining the effects of the enhancement on the differences between all subjects and just high experience subjects for each device showed that the combination of enhancement and subject experience had little effect on the subjective satisfaction of all devices. The only notable exceptions to this result were for workload ratings for the enhanced multimodal devices, with large differences in workload ratings between all and just high experience subjects (Table 14.10 shown in **bold** type).

For the multimodal devices the enhancement allowed users to reduce or drop their required workload more rapidly with increasing experience when using the enhanced devices than with standard devices. This indicated that there was a longer learning time for users to reduce the workload needed to control both the monomodal dwell selection device and the magnification enhancement, than there was for users to reduce the workload needed to control the multimodal click selection device and the magnification enhancement. This showed that there was an additional workload penalty for monomodal selection, with even high experience users showing little reduction in workload when using the monomodal devices. Some care should be taken with these results due to the small sample sizes for experienced participants; however they did offer indications into the effect of participant experience on the performance and satisfaction of the devices.

In summary, the addition of the enhancement in combination with subject experience had little effect on the subjective satisfaction of the devices between all experience and high experience subjects, indicating the enhancement did not increase satisfaction over differing experience levels. The exception to this was for the multimodal devices, where the enhancement allowed workload to be appreciably reduced with increasing subject experience.

14.7 A summary of the effect of enhancement on eye and head mice

Reviewing the success of the predictions outlined at the start of this chapter showed that eight of the nine were proven. These results are summarised:

Results of Predictions:

11. *The benefit* of increased *task efficiency* for *smaller target sizes* - **TRUE**

12. Magnification levels used to give *effective target sizes sufficiently large* to achieve *parity of performance* with the baseline *hand mouse* - **FALSE**

13. That the magnification enhancement will be used such that *effective target sizes* are *constant*, regardless of the original target size - **TRUE**

14. The effect of a *supporting modality* would be maintained *irrespective of effective target size due to enhancement*, with monomodal selection being less efficient that multimodal selection - **TRUE**

15. The effect of *differing interaction techniques* would be maintained *irrespective of effective target size*, with task efficiency reducing with increasing complexity of manipulation - **TRUE**

16. The progression of *increasing performance* with *increasing subject experience* would be *maintained* and not affected by the enhancement - **TRUE**

17. The effect of a *supporting modality* would be maintained *irrespective of subject experience*, with monomodal selection being less efficient that multimodal selection - **TRUE**

18. The addition of the *maximum available number of supporting modalities* controlling any device would result in the *highest performance* for any given level of user experience - **TRUE**

19. For *high experience subjects* with the *maximum number of supporting modalities* the performance of the eye mouse would potentially be increased sufficiently that for large target sizes requiring less magnification, the eye mouse would outperform the head mouse and approach the baseline hand mouse performance - **TRUE**

Prediction 12 did not hold as subjects did not sufficiently magnify targets to such an extent that they would be large enough to be selected very easily, and hence achieve parity of performance with the hand mouse baseline. It is possible that alteration of the control method for the enhancement may encourage greater use or the automation of the control of magnification so that zoom to an adequate target size is automatic. Preliminary results (Bates et al. 2005) suggest that this may be effective.

It was shown that that the ability to temporarily increase target object sizes on the interface did lead to increased performance, and that this increase was disproportionately large for smaller target sizes. In addition that control over target size reduced the effect original target sizes had on the performance of the head and eye mice and that subjects tended to magnify targets, irrespective of their original size, to a near constant size to ease of manipulation. It was also found that this size was not sufficiently large to allow the head and eye mice to equal the performance of the hand mouse baseline, suggesting that subjects should be encouraged to use the enhancement to a greater extent (see previous comments above), perhaps by making the enhancement automated to zoom to idealised magnification levels and so be easier to control.

The effect of differing interaction techniques was maintained irrespective of effective target size, with enhancement having no effect on the reduction of performance with increasing complexity of manipulation, and that unrestricted manipulation was consistently the most efficient, and that even with control over object sizes, interfaces designed for head and particularly eye based interaction should use this form of object interaction technique wherever possible.

The progression of increasing performance with increasing user experience was unaffected with the addition of the enhancement. However the enhancement disproportionately increased performance for low and medium experience subjects, indicating that its use could increase the uptake or acceptability of, particularly, eye based devices with new users. Enhancement and experience did not affect the performance penalties associated with monomodal operation, with multimodal operation giving consistently higher performance than monomodal operation, even with high subject experience, indicating that when available multimodal operation should be used for highest performance irrespective of user experience.

The best use of modalities was investigated and it was found that for low and medium experience eye mouse users with a single available control modality this was best used controlling the enhancement rather than selection as these users had most difficulty locating the cursor on a target over and above control over how and when the target was then selected. For all other devices and experience levels, a single control modality was best used controlling selection rather than the enhancement, as the head mouse and high experience eye mouse users were sufficiently accurate in pointing to benefit more for selection control than object size control.

The maximum achievable performance of the head and eye mice was determined so that for high experience subjects, with the maximum number of supporting modalities, the performance of the eye mouse could be increased sufficiently that the eye mouse outperformed the head mouse and approached the baseline hand mouse performance.

These final results, particularly for the eye mouse, showed that this work did sufficiently increase the performance of head and eye mice to rival the baseline of a hand mouse, and that both devices could be simply enhanced to the extent that they could offer genuine alternatives for pointing interaction on unmodified graphical user interfaces.

Chapter 15

Conclusions

15.1 Summary

The aim of this work was to investigate the performance of eye and head based pointing devices during direct manipulation on a standard graphical user interface, with the aim of enhancing the performance of these devices so that they might reach a comparable performance to the known baseline of a hand mouse.

This work first introduced the properties of eye and head based pointing (Chapter 2), characterising head pointing as slow, effortful, but accurate and under conscious control, and eye pointing as rapid, fatigue free but inherently inaccurate and difficult to consciously control. The work then introduced existing eye and head pointing systems, illustrating their technological strengths and weaknesses, with head based systems tending to produce devices that did not exhibit a high degree of accuracy, or responsiveness, but did have a moderately high degree of ease of use, and eye based systems typically difficult to set up and use, sometimes invasive, but could offer greater tracking accuracy and higher responsiveness than head mouse systems. Supporting software devices were also investigated for text entry and object selection. On-screen virtual keyboards were typically found to require large keys and offer low speed on text entry, with object selection haphazard and based on the time a cursor would dwell on an object of interest.

The characterisation of the properties of eye and head pointing, together with the illustration of the properties of the systems used to enable eye and head pointing, showed how these systems are both potentially enabling for disabled users but also limited in their performance.

This work then started the investigation into the performance of eye and head based pointing devices (Chapter 3) by searching for and reviewing existing methods that could be used to assess and investigate the performance of these devices, such as simple abstract target acquisition tests and more 'real world' interaction scenarios. No suitable existing assessment method was found, with abstract tests such as those based on Fitts' law found to lack detail and insight and perhaps not be applicable to eye based pointing as a non-Fitts device, and 'real world' tests lacking repeatability and detail due to their random nature or lack of task breadth and variation. Hence a new suitable method was required.

A method was required (Chapter 4) that would be both 'real world' in operation and mimic as best as possible a natural and unrestricted use of an interface, but also be repeatable and allow detailed examination of performance at a task and object level. This work first conducted a survey of the structure of the interface to determine the most fundamental elements of interaction with a 'real world' interface, breaking down the interface into object size categories and object interaction requirement types. It then determined a test sequence, based on previous work, that as best as possible mimicked 'real world' interaction but that was also based on repeatable tasks with single fundamental object size types and interaction techniques, thus detailed analysis of performance could be undertaken at the most fundamental levels.

A method of performance measurement was now required (Chapter 5) to assess the performance of the eye and head mice on the 'real world' test scenario. This work then constructed a detailed performance measurement method based on existing metrics of objective efficiency comprised of tasks times and quality of task interaction, and subjective satisfaction based on questionnaire responses to workload, comfort and ease of use. This then allowed insight not only into the performance of the pointing devices in this work, but could also indicate how the performance of the head and eye mice devices might be enhanced.

Subjective questionnaire measurement scales were then investigated (Chapter 6) and a diversity of scales found in use from 5 intervals to 20 intervals with a variation of labelling schemes, but with little reason given or found for their use. The work thus experimentally generated a comprehensive range of 5 to 11 interval new bipolar and unipolar scales suitable for subjective satisfaction rating. The work then went on to experimentally determine which of these candidate scales would be most suitable for the assessment method (Chapter 7) by being both the most sensitive to assess small variations in device assessment, but also have sufficient range to measure a wide range of possible device performances. The work also showed experimental results that questioned the use of, and range and discrimination of, many of the subjective scales that have been used previously. Finally, a suitable 7 interval labelled questionnaire scale was determined that could be used in both bipolar and unipolar formats, and that was the most suitable scale for the subjective assessment of eye and head mice.

The final step in construction of the assessment method for eye and head mice was a validation of the method (Chapter 8) against a known benchmark of a hand mouse with known degraded levels of performance. A hand mouse was degraded by applying random cursor displacement and the level of performance of the device experimentally measured.

These known levels of device performance were then applied to the test method, with both the objective and subjective elements of the method experimentally evaluated and found to be valid.

Before the performance of eye and head mice could be investigated, this work determined that eye and head mouse systems were required (Chapter 9) that gave repeatable and accurate performance measurements, and also allowed modification to their operation so that any performance enhancements proposed by this work could be applied. The work showed that existing devices were not wholly suitable, and then showed the construction of new eye and head based pointing devices, together with suitable software devices to support their operation.

The performance of eye and head based pointing devices could now be determined (Chapter 10). The verified assessment method was now used to experimentally test both the hand mouse baseline, and eye and head mice in both monomodal and multimodal selection configurations. The performance of the head and eye mice were characterised as predicted, with the head mouse having slower but more accurate pointing and the eye mouse more rapid but less accurate pointing. Pointing accuracy had a strong influence on eye mouse performance but little influence on head mouse performance, and pointing speed had a strong influence on head mouse performance but less influence on eye mouse performance. The numbers of modalities available were also determined to have a strong effect on performance with multimodal selection outperforming monomodal selection. This work showed that interaction was feasible with all devices, although the performance of the eye-based devices was poor and would certainly prove to be unsatisfactory in comparison with the performance of a hand mouse. This work showed that neither the head nor eye based devices achieved the same performance as the hand mouse baseline, showing the penalty users of these devices incur in comparison to hand mouse users.

The detailed construction of the test method was used to examine how object sizes, interaction techniques, and subject experience (Chapter 11) affected the performance of the devices. This work showed that unrestricted interaction was most beneficial to head and eye based pointing due to the uncertainty of cursor location, that increasing target size on the interface increased the performance of the devices, and that extensive test subject experience with the devices enhanced performance. This work then showed how the combined effects of target size, experience and multimodal operation with the devices could maximise performance, although even with this neither the head nor eye based devices could yet attain the level of performance of the hand mouse. This work then predicted that, of these factors, the most practical on an existing 'real world' interface was to provide some means of increasing target object size on the interface as an enhancement to the devices to further increase performance to hand mouse levels.

A method of enhancing the eye and head devices by interface magnification, or zooming, was then described (Chapter 12). Previous methods were described and rejected due to their limitations of indirect manipulation or lack of suitability for eye and head based systems, and a new interface magnification enhancement developed for the eye and head mice.

This work then experimentally determined the effects of adding the object magnification enhancement (Chapter 13) to the head and eye mice by applying the verified test method to these new devices. Predictions were made and examined on the expected benefits of the enhancement and the expected costs of the enhancement. This work found that the zoom enhancement exhibited benefits that far outweighed costs and disproportionately enhanced eye mouse performance in comparison to head mouse performance, resulting in equality of performance between the enhanced head and eye mice for either monomodal or multimodal operation. This work showed that equality of performance between head and eye could be achieved with enhancement thus allowing users to potentially choose either device for similar performance. However, neither enhanced head nor eye devices approached the performance of the baseline hand mouse.

The final analysis of this work was to determine the highest performance of the enhanced head and eye mice against the performance of the baseline hand mouse (Chapter 14) and thus to determine if the aims of this work to raise eye and head mouse performance to a known baseline were achieved. The detailed construction of the test method was again used to examine how object sizes, interaction techniques, and subject experience affected the performance of the devices. This work examined the effect of the enhancement on target size and performance and found that the ability to temporarily increase target object sizes on the interface did lead to increased performance as predicted. It found that unrestricted manipulation was again consistently the most efficient, and that even with control over object sizes, interfaces designed for head and particularly eye based interaction should use this form of object interaction technique wherever possible. This work showed that the enhancement disproportionately increased performance for low and medium experience subjects, indicating that its use could increase the uptake or acceptability of, particularly, eye based devices with new users. Finally, this work determined that for high experience subjects, with the maximum number of supporting modalities, the performance of the eye mouse could be increased sufficiently that for large target sizes requiring less magnification, the eye mouse outperformed the head mouse and approached the baseline hand mouse performance. These final results, particularly for the eye mouse, showed that within this work the performance of head and eye mice was increased sufficiently to rival the baseline of a hand mouse, and that both devices could be simply enhanced to the extent that, when used on similar tasks, they could offer genuine alternatives for pointing interaction on unmodified graphical user interfaces.

15.2 Outcomes of this research

The outcomes of this work can be briefly summarised by the following:

Chapter 2 - A new survey and comparison of head and eye based pointing systems highlighting the properties, advantages and disadvantages of using these modalities and devices for direct interaction on a 'real world' interface.

Chapter 3 - A new survey and comparison of methods suitable for assessing the performance of assistive technology pointing devices during direct interaction on a 'real world' interface, resulting in:

3.1 The conclusion that no such suitable method was available.

3.2 A new definition of the requirements of a suitable method.

Chapter 4 - The construction of a new method for analysing the component objects, interaction techniques, and object sizes present on a graphical user interface in order to construct a new set of 'Real World' test tasks allowing multiple levels of analysis, resulting in:

4.1 A new analysis of the interaction requirements of the Windows graphical user interface.

4.2 The construction of a new test task set that mimics 'real world' interaction in the two domains of Word-processing and Web browsing, and that also contains all component objects, interaction techniques, and object sizes present on the Windows graphical user interface.

Chapter 5 - A new method suitable for measuring the objective and subjective performance of assistive technology pointing devices during direct interaction on a 'real world' interface, resulting in:

5.1 A new detailed objective task performance assessment method based on existing concepts of task efficiency, task time and task quality.

5.2 A new detailed subjective performance assessment method based on existing concepts of user satisfaction, workload, comfort, and ease of device use.

Chapter 6 - A survey and comparison of assessment questionnaire scales suitable for measuring the subjective user satisfaction with assistive technology pointing devices during direct interaction on a 'real world' interface, resulting in:

6.1 The conclusion that no such suitable assessment questionnaire scales were available.

6.2 A new experimentally determined set of candidate assessment questionnaire scale quantifiers.

6.3 A new experimentally determined set of assessment questionnaire scales of 5, 7, 9 and 11 intervals.

Chapter 7 - A new experimentally determined analysis of the performance of assessment questionnaire scales of 5, 7, 9, 11 and 20 intervals, resulting in:

7.1 The conclusion that 7-interval fully labelled scales held the highest discrimination and range in both bipolar and unipolar formats for pointing device assessment.

7.2 The conclusion that the use of other scales such as the 20 interval NASA-tlx scale may be invalid for pointing device assessment.

Chapter 8 - A new method for validating methods suitable for assessing the performance of assistive technology pointing devices during direct interaction on a 'real world' interface, resulting in:

8.1 A new experimentally determined method of degrading the known benchmark performance of a hand mouse that mimics assistive technology pointing devices.

8.2 A new experimentally determined validation of the full assessment method in this work using the benchmark performance of a hand mouse mimicking assistive technology pointing devices.

Chapters 10 and 11 - The experimental application of the new method suitable for assessing the performance of assistive technology pointing devices during direct interaction on a 'real world' interface to hand, head and eye based pointing devices, resulting in:

10.1 The task domain has no effect on the performance of hand, head and eye based pointing.

10.2 Hand based pointing has higher objective performance and higher subjective satisfaction than head and eye based pointing.

10.3 Head based pointing has higher objective performance and higher subjective satisfaction than eye based pointing.

10.4 Multimodal selection operation has higher objective performance and higher subjective satisfaction than monomodal selection operation.

11.1 For equivalent objective performance head based pointing requires larger target sizes than hand based pointing.

11.2 For equivalent objective performance eye based pointing requires larger target sizes than hand and head based pointing.

11.3 For equivalent objective performance monomodal selection operation requires larger targets than multimodal object selection.

11.4 Objective performance reduces with increasing complexity of object manipulation requirements for head and eye based pointing.

11.5 Interfaces designed for head and eye based manipulation should use unrestricted in preference to restricted object manipulation characteristics.

11.6 Objective performance increases with increasing user experience for head and eye based pointing.

11.7 For equivalent objective performance eye based pointing requires considerably more user experience than the head based pointing.

11.8 The objective performance of head based pointing can approach the objective performance of hand based pointing if object sizes are sufficiently large, multimodal object selection is used, and users are sufficiently experienced.

11.9 The objective performance of eye based pointing can approach the objective performance of hand and head based pointing if object sizes are sufficiently large, multimodal object selection is used, and users are sufficiently experienced.

11.10 Control over object size is a method of performance enhancement that could be applied to head, and particularly eye based pointing, that would greatly enhance the objective performance of these devices.

Chapter 12 - Based on these findings, the enhancement of head and eye based pointing devices with a new interface magnification tool for the magnification of interface target objects during direct interaction with the interface.

Chapters 13 and 14 - The experimental application of the new method for assessing the performance of assistive technology pointing devices during direct interaction on a 'real world' interface on the enhanced head and eye based devices, resulting in:

13.1 The task domain has an effect on the performance of enhanced eye based pointing due to the use of the magnification enhancement.

13.2 Objective performance and subjective satisfaction increases marginally for enhanced head based pointing due to the ability to control object size being only partially beneficial to a moderately accurate pointing device.

13.3 Objective performance and subjective satisfaction increases appreciably for enhanced eye based pointing due to the ability to control object size overcoming a highly inaccurate pointing device.

13.4 The benefits of controlling the new magnification tool outweighed the costs.

13.5 A disproportionate objective performance increase for eye based pointing resulting in the removal of the dominance of device choice for maximum assistive technology device performance, allowing users to choose either head or eye based pointing devices for interaction.

13.6 Enhanced devices have higher subjective satisfaction than non-enhanced standard devices.

14.1 Greater magnification is used in inverse proportion to object size, resulting in objects being magnified to a near constant target size irrespective of their original size.

14.2 Objects were not magnified sufficiently to achieve parity of performance with hand based pointing.

14.3 For low and medium experience eye based pointing users a single control modality was best used controlling magnification rather than selection, and for all other devices and experience levels, a single control modality was best used controlling selection.

14.4 Eye based pointing supported with high experience subjects, the maximum number of supporting modalities and magnified large object sizes can exceed the performance of head based pointing and approach the baseline hand based pointing performance.

Publications - *This work has also produced several publications:*

1. Bates, R. (1999) Multimodal Eye-based Interaction for Zoomed Target Selection on a Standard Graphical User Interface. Proceedings of Interact'99 vol II pages 7-8. British Computer Society.

2. Bates, R., Bierton, R. (2000) Experimental Determination of Quantifiers for Usability Questionnaire Design. Proceedings of Human Computer Interaction 2000 vol II pages 57-58. British Computer Society.

3. Bierton, R., Bates, R. (2000a) Experimental Determination of Optimal Scales for Usability Questionnaire Design. Proceedings of Human Computer Interaction 2000 vol II pages 55-56. British Computer Society.

4. Bierton, R., Bates, R. (2000b) A Method for Classifying Pointing Device Interaction Errors on Standard GUI Widgets. Proceedings of Human Computer Interaction 2000 vol II pages 61-62. British Computer Society.

5. Bates, R. (2000) Using Soft Devices as Enabling Translators - Standard GUI Interaction For All! Proceedings of Human Computer Interaction 2000 vol II pages 53-54. British Computer Society.

6. Bates, R. (2001) Have Patience With Your Eye Mouse! Proceedings of RAATE 2001 (Recent Advances in Assistive Technology and Engineering), 5th-6th November 2001, Birmingham, UK.

7. Bates, R. (2002a) A Computer Input Device Selection Method for Users with High-Level Spinal Cord Injuries - Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Trinity Hall, University of Cambridge 25th-27th March 2002.

8. Bates, R. (2002b) Have patience with your eye mouse! Eye-gaze interaction with computers can work- Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Trinity Hall, University of Cambridge 25th-27th March 2002.

9. Bates, R. Istance, H.O. (2002a) Why are eye mice unpopular? A detailed comparison of head and eye controlled assistive technology pointing devices-

Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Trinity Hall, University of Cambridge 25th-27th March 2002.

10. Bates, R. Istance, H.O. (2002b) Zooming interfaces! Enhancing the performance of eye controlled pointing devices - Proceedings of ASSETS 2002, The Fifth International ACM SIGCAPH Conference on Assistive Technologies, 8th–10th July 2002, Edinburgh, Scotland.

11. Bates, R. (2002c) Eye-gaze interaction with computers can work - Proceedings of the Scientific meeting of Institute of Physics in Engineering and Medicine (IPEM) 9th-12th September 2002, Durham.

12. Bates, R. Istance, H.O. (2003) Why are eye mice unpopular? A detailed comparison of head and eye controlled assistive technology pointing devices – Journal of the Association for the Advancement of Assistive Technology in Europe (AAATE): Universal Access in the Information Society Volume 2, Number 3, October 2003 Pages: 280 - 290 Special Issue on "Countering Design Exclusion", guest-edited by Simeon Keates and John Clarkson, Springer-Verlag Heidelberg ISSN: 1615-5289.

13. Bates, R. Istance, H.O. (2004) Towards eye based virtual environment interaction for users with high-level motor disabilities. International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT), pages 275-282, 20th-22nd September 2004, New College, Oxford, UK, ISBN 0704911442.

14. Bates, R., Istance, H.O., Donegan, M., Oosthuizen, L. (2005) Fly Where You Look: Enhancing Gaze Based Interaction in 3D Environments. In Proceedings of HCI International 2005, Volume 7 - Universal Access in HCI: Exploring New Interaction Environments, Caesars Palace, Las Vegas, USA, July 22-27 2005.

15. Donegan, M., Oosthuizen, L., Bates, R., Daunys, G., Hansen, J. P., Signorile, M. I., & Majaranta, P. (2005). Providing eye control for those who need it most - a study on user requirements. (Abstract) In Proceedings of ECEM13 - 13th European Conference on Eye Movements, Bern, Switzerland, August 2005.

15.3 Contribution of this research

The performance of and reasons why eye and head based pointing devices seem to exhibit poor performance during direct manipulation with a standard graphical user interface had not previously been fully determined, and possible methods of enhancing the performance of these devices to near parity with basic hand mouse interaction had not been investigated. In addition, no validated method for determining the performance of these devices and determining possible enhancements to these devices had been demonstrated. The contribution of this work was to address these questions and present a validated assessment method for eye and head based pointing, and hence give a full investigation of eye and head based pointing direct manipulation on a standard graphical user interface during 'real world' interaction. From this a target object magnification performance enhancement was determined such that the performance of these devices could approach that of basic hand mouse interaction.

This work contributed an assessment method and evaluated enhancement to eye and head based pointing that allowed users with a high-level motor disability to approach the direct interaction performance of all other hand based pointing users.

15.4 A discussion on eye-based interaction design

With the outcomes of this work recommendations and conclusions on magnification and cursor control can be drawn to aid eye gaze based interaction with graphical user interfaces. These recommendations may be addressed in two areas, the first aiding interaction with pre-existing standard graphical user interfaces by the use of additional gaze interaction assisting tools, and the second aiding interaction with specially designed interfaces produced specifically for eye gaze manipulation by forming design guidelines for these interfaces.

Examining what tools could be added to pre-existing graphical interfaces this work has shown that the addition of a target magnification tool can greatly enhance performance. In this work the tool was controlled manually via a simple switch, with step-wise increases in magnification (x1, x2, x4, x8 and so on) requiring a button press for each increase or decrease in magnification level. This approach contains task time and possible interaction error penalties, as well as possibly increasing the effort required to control the magnification tool, as often several presses may be required for a user to achieve the required level of magnification. Hence a design recommendation would be to add some knowledge of the objects on the interface to the magnification enhancement. Perhaps the magnification enhancement could scan the interface and determine object positions and sizes (one method would be to employ Microsoft Active Accessibility[1] software to achieve this). With knowledge of the pointing accuracy of eye gaze and the size of the interface object under or closest to the cursor, when the user presses the zoom switch the magnification enhancement would automatically generate sufficient magnification to make the target object large enough to select. This work has shown that given the magnification tool, all target sizes are magnified to a near-constant 1.7° visual angle at 60cm from the screen (Table 14.2), so the automated magnification level could be set to near this figure, irrespective of the original target object size. This would remove the need for multiple switch presses to control magnification levels. Operation of the magnification tool could also be enhanced for users who cannot operate a switch and who use dwell selection. Here as the cursor is held steady on or near a target and the dwell selection time starts to expire, the magnification tool could zoom into the target to magnify the target to a near-constant 1.7° (Table 14.2) just before an automatic dwell selection is generated, thus increasing dwell selection accuracy and removing the need for the user to control the magnification enhancement at all. Such 'auto-zoom' could be further enhanced and automated by monitoring both the eye gaze pattern on the screen near a target object together with the size of that object. If the interest of the user in the target can be determined (via gaze pattern analysis) then the target may be magnified automatically purely via the level of interest detected by the gaze patterns around that object. This work addressed magnification by magnifying the complete interface screen centred on the cursor location, but magnification could be employed only within the vicinity of the cursor (rather like a 'fish-eye' view or magnification lens attached to the cursor). Again this magnification could be automated and dynamic as described above, and only be invoked once user interest in an object is determined.

Another approach to enhancing pre-existing interfaces would be to use a 'virtual' magnification scheme. Here no actual magnification of the screen is caused, instead the area of effect of the cursor is enlarged (an 'area' cursor). Thus once the user invokes a magnification either voluntarily or by the tool automatically generating a magnification (as described previously above) the cursor is magnified to cover a larger area, thus encompassing any small object within a given radius of the existing cursor position. This would produce an effect very similar to magnifying the screen, but without the need to do so, so leaving the entire screen available for manipulation.

A separate but related issue concerns the amount of cursor positional corrections made during eye interaction. Due to the inaccuracy of pointing with the eye, users produce jitter of the cursor position as the eye flicks onto the intended target, but due to tracking

[1] http://msdn.microsoft.com/library/default.asp?url=/library/en-us/msaa/msaastart_9w2t.asp (12/2005)

inaccuracies the cursor remains a little off target. The eye is then drawn to the cursor and its inaccurate position, thus moving the tracked gaze position, and thus displacing the cursor further. These positional control corrections were the main source of low efficiency for the eye mouse (Figures 10.3, 10.4) and were present for both smaller and larger target object sizes (Figures 11.2, 11.3). Even with the magnification enhancement applied (Figures 13.3, 13.4) these control corrections were still evident, although reduced. One method of reducing these positional control corrections would be to use a secondary cursor or 'puck' object. The puck would be drawn around and highlight any object that was the last object that was closest to the eye driven cursor, and would indicate which object would be manipulated or selected by any button press or dwell click generated, irrespective of where the actual eye driven cursor has subsequently moved to. Once the puck is located on the target that the user wishes to manipulate then the eye is free to move away from fixing on that object, but the object is still subject to any selection actions as the puck will remain on that object. This detaches the eye cursor away from the object and frees the eye to receive information from the interface at any position, rather than tethering the eye to the object during its manipulation. The decision on which object the puck is placed around can be determined by monitoring object locations on the interface and gaze patterns of the user around objects, as described previously, with the object that receives most gaze attention being highlighted by the puck, or the puck 'snapping-on' to the closest object. This also enhances dwell selection as it allows accumulative dwell times by allowing gaze to leave and re-enter a target object without the cursor losing focus on that target. The puck may also be used with target magnification and area cursor techniques as described previously, with the puck being placed instead of a selection being generated after magnification. Selection of the highlighted target can then take place at any time irrespective of the gaze position of the user.

Another method of addressing the distraction of the cursor and the control corrections that are generated would be to hide or remove the visible cursor completely. Thus the eye would not be drawn to an inaccurate cursor but could maintain gaze on the intended target object. Due to tracking inaccuracies the non-visible cursor may not be on the intended target, and there would be no way the user could see that they were not locating the cursor on the target. However, this can be solved by employing the automated zoom magnification methods described earlier. If the target object were magnified quite considerably (perhaps to a greater extent than found by the work here using a visible cursor) then the non-visible cursor would always be encompassed by the magnified object, thus enabling accurate selection. This approach of non-visible cursor and large (always obtainable) magnified targets has the further benefit of 'fooling' the user into believing that all target objects can easily be selected, and that there are no pointing inaccuracies when using eye based pointing. This could dramatically improve user acceptance of gaze based interaction.

Turning to recommendations for specially designed interfaces that are produced specifically for eye gaze manipulation opens a wider range of design possibilities. From this work it is clear that perhaps the most effective change would be to design an interface comprised solely of large target objects. Clearly this would resolve any eye gaze based pointing inaccuracies. However this would impact on available interface screen area and the number of buttons and menus the interface could display, and thus would probably give a trade-off in usability between device performance and interface utility. Examining the results of this work shows that if objects were larger than approximately 1.7° visual angle at 60cm from the screen (Table 14.2) then manipulation by eye gaze should become usable and approach hand and head based manipulation. Hence an interface could be designed with no object subtending less than 1.7° to give ease of manipulation via eye gaze, but also with no objects subtending excessively greater angles than 1.7° as this would allow the maximum number of objects to be fitted onto the screen.

The opportunity to design a custom interface for gaze manipulation would allow the removal of specific manipulation types. Referring back to Section 11.4 of this work (Table 11.1) showed that performance was degraded considerably when attempting to manipulate objects that required double clicks, dragging, and restricted manipulation[1]. These can all be addressed by a new interface. Double clicks can be replaced by single clicks, dragging could be replaced by a 'cut' and 'paste' move where the object is not visible and not attached to the cursor and hence the eye position during the drag (where it distracts the user during dragging), and finally all restriction on retaining the cursor within a target or menu during manipulation can be removed. (It is still not clear why restricted manipulation is present on any graphical user interface, as it offers no apparent advantages and often causes manipulation difficulties).

Another method of enhancing performance on custom designed interfaces for eye gaze manipulation would be the use of probability in object manipulation. Here the interface

[1] As described in Chapter 4, there are three basic forms of pointing device object manipulation on a Windows interface. These are a single click on an object, a double click on an object, and a drag of an object. These actions can be either *restricted*, where cursor movement is confined within the area of the object to retain control of that object, or *unrestricted* where cursor movement may move from the object and then return to the object and still retain control of that object. For example, a button object on the Windows interface has *unrestricted* manipulation as, provided the mouse button is held down, the cursor may be moved away from the button and returned without losing permanent control of the button. In contrast, a hypertext link has *restricted* manipulation, as any movement of the cursor away from the object with the mouse button held down will lose control of the object.

could accumulate a history of object usage in differing task contexts and hence build a model of 'typical' interaction during a given task which would allow the prediction of which target object or subset of target objects are most likely to be manipulated next. For example, it is probable that after performing a 'print preview' a user is likely to either select 'print' or 'cancel' rather than any other action. Hence the most probable targets for the next manipulation could be temporarily magnified to aid selection – thus the zoom magnification tool would be an intrinsic part of the interface rather than an add-on. Also any dwell times required to select 'more probable' targets may be made shorter than dwell times for 'less probable' targets. In addition, this metric of probability could be applied to any of the magnification, area cursor or puck techniques described previously. Of these the puck technique is of interest – if the puck could jump automatically to the next most probable object then no pointing and only selection clicks would be required, with any erroneous predictions ignored by manipulation of alternate objects in the usual way.

This approach of probability could allow the interface to be designed from the outset with improbable objects placed adjacent to each other with little spacing, and probable targets spaced far apart. When the inaccurate eye gaze cursor is in close proximity with a probable and improbable object placed next to each other the probable object of the group will automatically be selected. By placing two similar probability objects sufficiently far apart (greater than the pointing inaccuracy of eye gaze) the object of the pair that the user wishes to manipulate can always be determined as the cursor will always be closer to one or other of the objects. Such an approach where objects that are likely to be selected in a sequence are widely spaced on the interface can only be effective for eye gaze input devices as the eye exhibits a ballistic type of movement (the further the distance, the more rapid the movement). Hence (as discussed in Section 3.2) these devices are not true Fitt's Law devices and the distance needed to be travelled between objects does not greatly affect movement time or performance.

A final approach to interface design would be to use very large target objects (greater than 1.7° visual angle at 60cm from the screen) to give a high ease of selection, and also to address the problem of only being able to fit a few large objects on the screen, by producing a dynamically changing interface. Such an interface would initially present a few large targets, and based on which target object was selected the interface would then be redrawn with a new set of objects representing the next logical steps for interaction, with this process of layers under layers of objects continuing indefinitely. Such an approach has been discussed for keyboard and simple task entry (Section 2.6, Figures 2.8 and 2.9) but this could be extended to the full function of the interface. This approach opens a new area of interfaces for eye gaze control that could offer greater performance for eye interaction than any other interaction type, by the use of eye gaze to 'fly' into a 3-dimensional interface. Here the user appears to fly into the interface with the flight path guided by where they are gazing, as small (and hence the possibility of many) objects appear in the distance so the user gazes toward the objects they wish to manipulate and is 'flown' toward them so that the objects become larger in the field of view until the point where the required object is 'flown through' for selection. Such an interface has been attempted (Bates et. al., 2005) where preliminary results found that eye gaze offered greater performance than hand.

15.5 Conclusions

This work posed the question could eye and head based interaction performance approach that of basic hand mouse interaction. Eye and head based pointing devices are often used by high-level motor disabled people to enable computer interaction in the place of a standard desktop hand mouse. However, the performance of these devices when used for direct manipulation on a standard graphical user interface has generally been regarded as poor in comparison to that of a standard desktop hand mouse, putting users of head and eye mice at a disadvantage when interacting with computers.

The performance of eye and head based pointing devices during direct manipulation on a standard graphical user interface had not previously been fully investigated, and the reasons why these devices seemed to demonstrate poor performance had not been determined in detail. Few proven methods had been demonstrated and investigated that enhanced the performance of these devices based on their performance during direct manipulation. In addition, no validated assessment method had been constructed to allow such an investigation.

This work addressed these questions by constructing and verifying a test method suitable for the detailed performance assessment of eye and head based assistive technology pointing devices. It then used this method to determine the factors influencing the performance of eye and head based pointing devices during direct manipulation. After identifying these factors, this work hypothesised, and then demonstrated that applying suitable methods for addressing these factors did result in enhanced performance for eye and head based pointing devices. It showed that the performance of these enhanced devices does approach the performance of a standard desktop hand mouse when the benefits of highly experienced users, a supporting modality for object manipulation, and a supporting interface enhancement for control over object size are available.

This work posed and addressed the question 'could eye and head based interaction performance approach that of basic hand mouse interaction' by showing that with zoom enhancement eye and head based interaction performance can approach that of basic hand mouse interaction, and hence be a viable and usable interaction method for people with high-level motor disabilities.

15.6 Future work

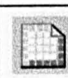
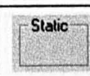
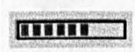
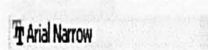
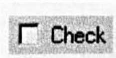
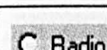
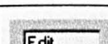
This work has continued after this thesis with the extension of eye and head based pointing for users with high level motor disabilities to interaction with 3-dimensional environments (Bates and Istance 2004). Based on this and other eye-based work the author has also extended this work by being invited to join the managing Steering Board and also lead a key Work Package on the European Union 6th Framework Information Society Technologies "Communication by Gaze Interaction" (COGAIN) Network of Excellence[1]. This five year €3 million research project has gathered Europe's leading expertise in eye tracking integration with computers into a research project on assistive technologies for citizens with motor impairments. The project aims to improve existing gaze-based interaction techniques, and facilitate the implementation of cutting edge eye gaze based systems for everyday communication. Major parts of this work will be integrated into this European project, including the surveys of eye based systems from Chapter 2, the assessment method presented in Chapters 3 to 8 and the magnification enhancement presented and examined in Chapters 9 to 14.

[1] COGAIN: www.cogain.org

Appendices

Appendix Chapter 4

Table A4.1 Typical Windows interface objects

Typical Windows interface objects		
Object name	*Representation*	*Function*
Static Text		Show information
Picture / animation		Show information
Group Box		Group objects
Progress Indicator		Show information
Text characters	Hello	Display text, allow editing of text
Spin button		Change spin value
Drop down list button		Show or hide list
List box item		Picks item from list
Window control button		Control window size, close window
Check Box		Set true/false value
Radio Button		Set one value from a set
Scrollbar button		Scroll slider
Scrollbar slider		Scroll slider
Scrollbar channel		Scroll slider
Standard toolbar button		Start command action
Edit Box		Input of text
Window size Control		Resize window

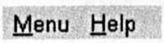
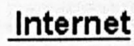
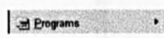
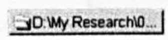
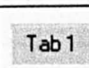
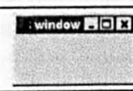
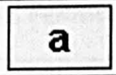
Menu		Select one command action from a set
Textual hypertext link		Move to new page
Command Button		Start command action
Window title bar		Move/size application window
Start menu entry		Start application or display menu entry
Task bar button		Give application focus
Icon		Start application
Graphic hypertext link		Move to new page
Large toolbar button		Start command action
Scroll Bar		Scroll window contents
Spin control		Increment/decrement a value
List box		Group objects, allow selection of objects
Drop down list box		Group objects, allow selection of objects
Tab control		Select one page
Window		Container for objects
Soft keyboard key		Generate keyboard input

Table A4.2 Taxonomy of Windows active fundamental objects

Taxonomy of Windows active fundamental objects			*Interaction area*		*Pointing device interaction*				
					Restricted			*Un – restricted*	
Object name	*Representation*	*Function*	*Description*	*Interaction area size category*	*Single click*	*Double click*	*Drag*	*Single click*	*Drag*
Text	**Hello**	Display text, allow editing of text	Text character	S1		✓		✓	✓
Spin button		Change spin value	Button area	S1				✓	
Drop down list button		Show or hide list	Button area	S2				✓	
Scrollbar button		Scroll slider	Button area	S2				✓	
Window control button		Control window size, close window	Button area	S2				✓	
Scrollbar slider		Scroll slider	Slider area	S2			✓		
Scrollbar channel		Scroll slider	Channel area	S2				✓	
Radio Button	Radio	Set one value from a set	Circle and text area	S2				✓	
Check Box	Check	Set true/false value	Box and text area	S2				✓	
Textual hypertext link	Internet	Move to new page	Text area	S2	✓				
List box item	Arial Narrow	Picks item from list	Item area	S2		✓		✓	
Edit Box	Edit	Input of text	Text area	S2				✓	
Standard toolbar button		Start command action	Button area	S3				✓	
Window size Control		Resize window	Lined area	S3					✓

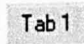
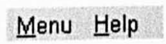
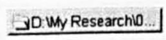
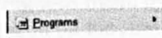
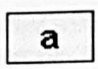
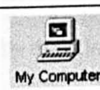
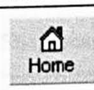
Tab		Select one page	Tab area	S3				✓	
Menu		Select one command action from a set	Text area	S3				✓	
Command Button		Start command action	Button area	S3				✓	
Task bar button		Give application focus	Button area	S3				✓	
Start menu entry		Start application or display menu entry	Menu entry area	S3				✓	
Window title bar		Move/size application window	Title bar area	S3		✓			✓
Graphic hypertext link		Move to new page	Graphic area	S4	✓				
Soft keyboard key		Generate text	Key area	S4				✓	
Icon		Start application	Icon area	S4		✓			✓
Large toolbar button		Start command action	Button area	S4				✓	

Table A4.3 Test tasks

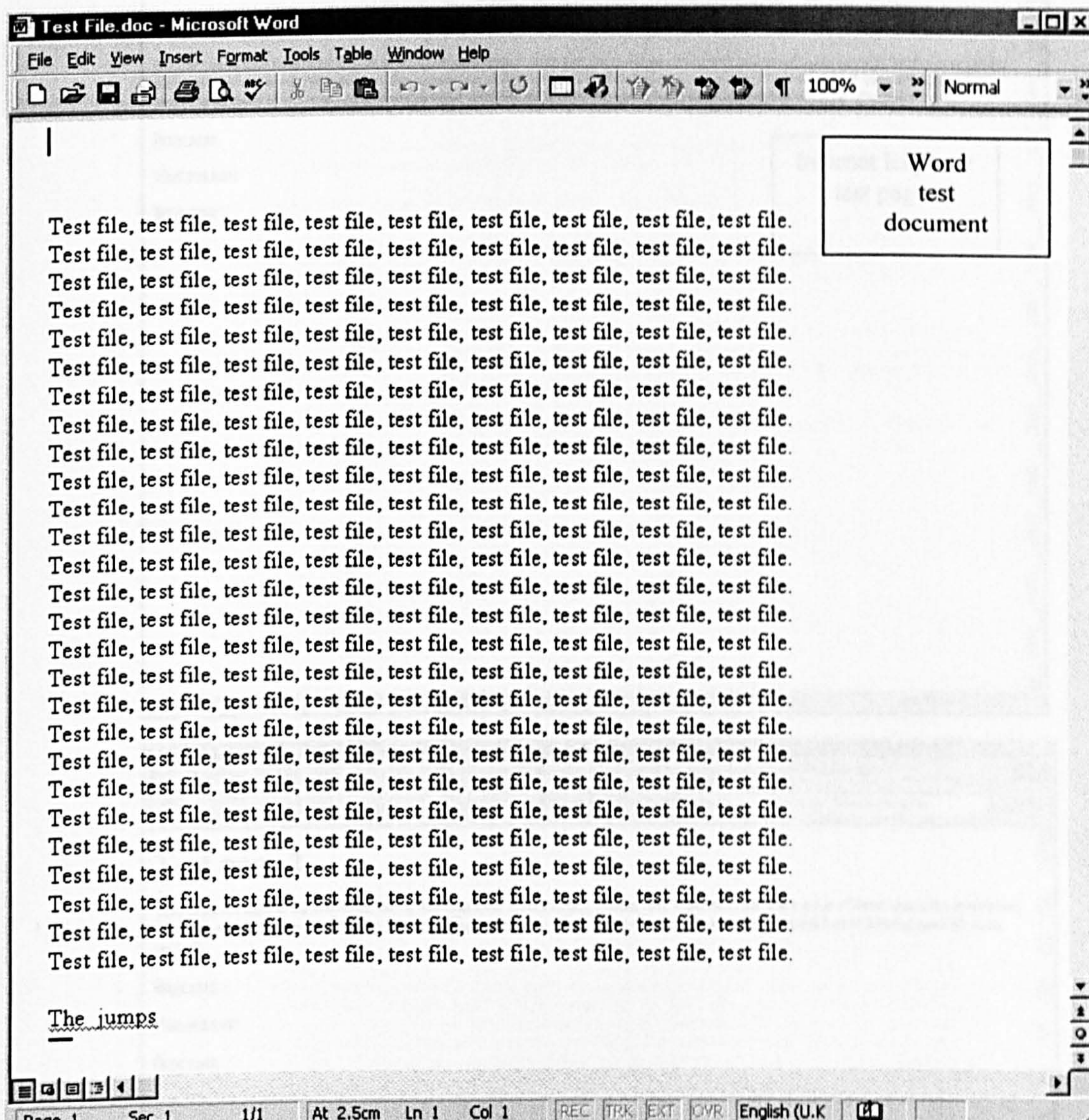
Task element	Task description	Object used	Interaction object size category	Pointing device interaction				
				Restricted			Un-restricted	
				Single click	Double click	Drag	Single click	Drag
1	Click the [Start] button on the task bar	Task bar button	S3				✓	
2	Open the Programs menu by clicking the [Programs] icon on the start menu	Start menu entry	S3				✓	
3	Start Word by clicking the [Microsoft Word] icon from the start menu (maximised)	Start menu entry	S3				✓	
4	Click the [Soft Keyboard] button on the task bar	Task bar button	S3				✓	
5	Resize Word by double clicking the window title bar	Window title bar	S3		✓			
6	Move the Word window to the top left of the screen by dragging the window title bar	Window title bar	S3					✓
7	Resize the Word window to fill the top 2/3 of the screen by dragging the bottom right size handle	Window size control	S3					✓
8	Click the [File] menu	Menu	S3				✓	
9	Click the [Open] menu item	Menu	S3				✓	
10	Double click the [Test File.doc] filename in the list box	List box item	S2		✓			
11-55	Type [The quick brown fox jumps over the lazy dog.] at the top of the document using [Shift] for the capital at the start	Keyboard key	S4				✓	
56	Highlight the word [fox] by double clicking the word	Text	S1		✓			
57	Click the [Format] menu	Menu	S3				✓	
58	Click the [Font] menu item	Menu	S3				✓	
59-61	Click on the [Up] and [Down] scrollbar buttons 3 times to display the [Courier] font	Scrollbar buttons	S2				✓	
62	Click on the [Courier] font name to change the font	List box item	S2				✓	
63	Click the [Strikethrough] check box	Check box	S2				✓	
64	Click the [Character spacing] tab	Tab	S3				✓	

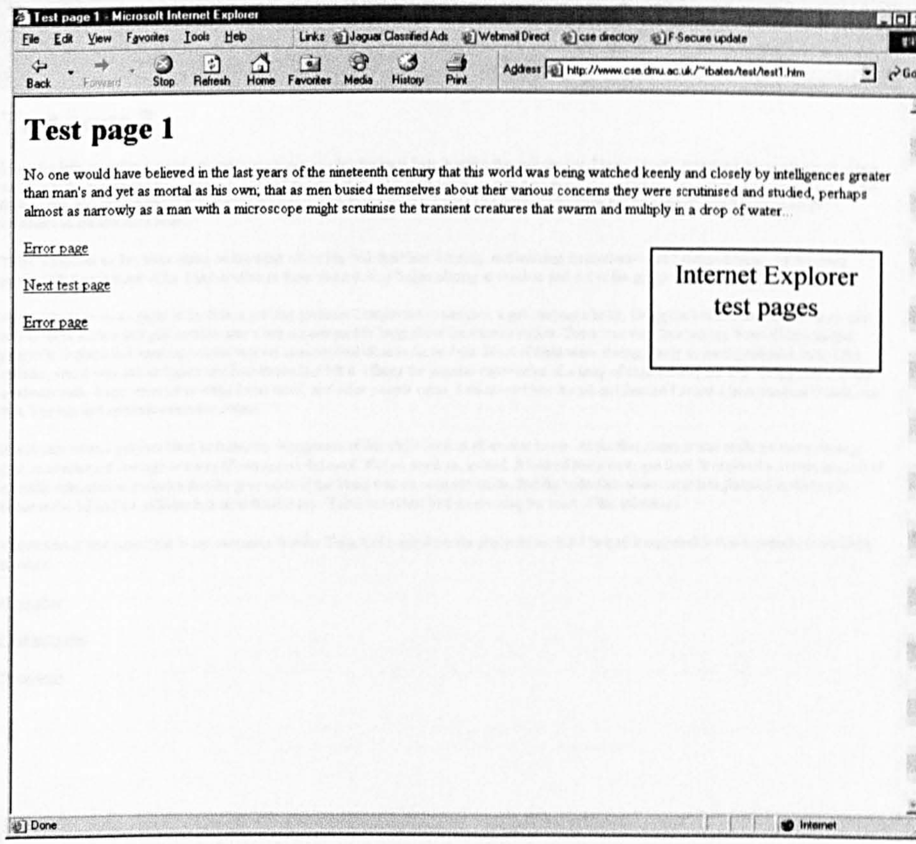
65-66	Click the [Up] spin button of the [Position] spin control two times to set the text position to 2pt raised	Spin button	S1				✓	
67	Click the [OK] button to close the dialog box	Command button	S3				✓	
68	Click the [Copy] toolbar button to copy the word [fox]	Standard toolbar button	S3				✓	
69	Scroll to the bottom of the page by dragging the window scrollbar slider	Scrollbar slider	S2			✓		
70	Click in between the words [The] and [jumps] to place the typing point between the words	Text	S1				✓	
71	Click the [Paste] toolbar button	Standard toolbar button	S3				✓	
72	Click the [File] menu	Menu	S3				✓	
73	Click the [Save As] menu item	Menu	S3				✓	
74-77	Type [test] for the file name	Keyboard button	S4				✓	
78	Click the [Save] button	Command button	S3				✓	
79	Minimise the Soft Keyboard window by clicking the [Minimise] button on the window title bar	Window control button	S2				✓	
80	Close Word by clicking the [Close] button on the window title bar	Window control button	S2				✓	
81	Double click the [Test files] folder icon	Icon	S4		✓			
82	Drag the Word [test] file icon into the [Test files] folder window	Icon	S4					✓
83	Click the [Start] button on the task bar	Task bar button	S3				✓	
84	Open the Programs menu by clicking the [Programs] icon on the start menu	Start menu entry	S3				✓	
85	Start Internet Explorer by clicking the [Internet Explorer] icon from the start menu (windowed)	Start menu entry	S3				×	
86	Maximise Internet Explorer by double clicking the window title bar	Window title bar	S3		✓			
87	Click in the Internet Explorer address box	Edit box	S2				✓	
88	Click the [Soft Keyboard] button on the task bar to show the Keyboard	Task bar button	S3				✓	
89-108	Delete the existing address using [Backspace] and type [c:\test\test1.htm] ([Shift] to access [:]) followed by [Return]	Keyboard key	S4				✓	
109	Minimise the Soft Keyboard window by clicking the [Minimise] button on the window title bar	Window control button	S2				✓	

110	(page1) Click on the [Next test page] hypertext link	Textual hypertext link	S2	✓				
111	(page2) Click on the [Next test page] hypertext link	Textual hypertext link	S2	✓				
112-114	(page3) Click on the [Down] scrollbar button three times to display the [Next test page] hypertext link	Scrollbar buttons	S2				✓	
115	(page3) Click on the [Next test page] hypertext link	Textual hypertext link	S2	✓				
116	(page4) Click in the [Down] scrollbar channel to display the [Next test page] hypertext link	Scrollbar buttons	S2				✓	
117	(page4) Click on the [Next test page] hypertext link	Textual hypertext link	S2	✓				
118	(page5) Drag the scrollbar slider down to display the [Next page] graphical hypertext link	Scrollbar slider	S2			✓		
119	(page5) Click on the [Next page] graphical hypertext link	Graphic hypertext link	S4	✓				
120	(page6) Click on the [Next page] graphical hypertext link	Graphic hypertext link	S4	✓				
121-126	(page7) Click the [Back] toolbar button 6 times until you see the [Test page 1] page	Internet Explorer Toolbar button	S4				✓	
127	(page1) Click the [Favourites] menu	Menu	S3				✓	
128	(page1) Click the [Test page 8] favourites menu item	Menu	S3				✓	
129	(page8) Highlight the text [fox] by dragging the cursor over the text	Text	S1					✓
130	Click the [Edit] menu	Menu	S3				✓	
131	Click the [Copy] menu item	Menu	S3				✓	
132	Scroll down the page by dragging the scrollbar slider to the [My Message] edit box containing the text [The ... jumps]	Scrollbar slider	S2			✓		
133	Click in between [The] and [jumps] in the [My Message] edit box to place the typing point between the words	Text	S1				✓	
134	Click the [Edit] menu	Menu	S3				✓	
135	Click the [Paste] menu item	Menu	S3				✓	
136	Click on the [Message priority] button	List box button	S2				✓	
137	Click on the [Urgent] list box item	List box item	S2				✓	

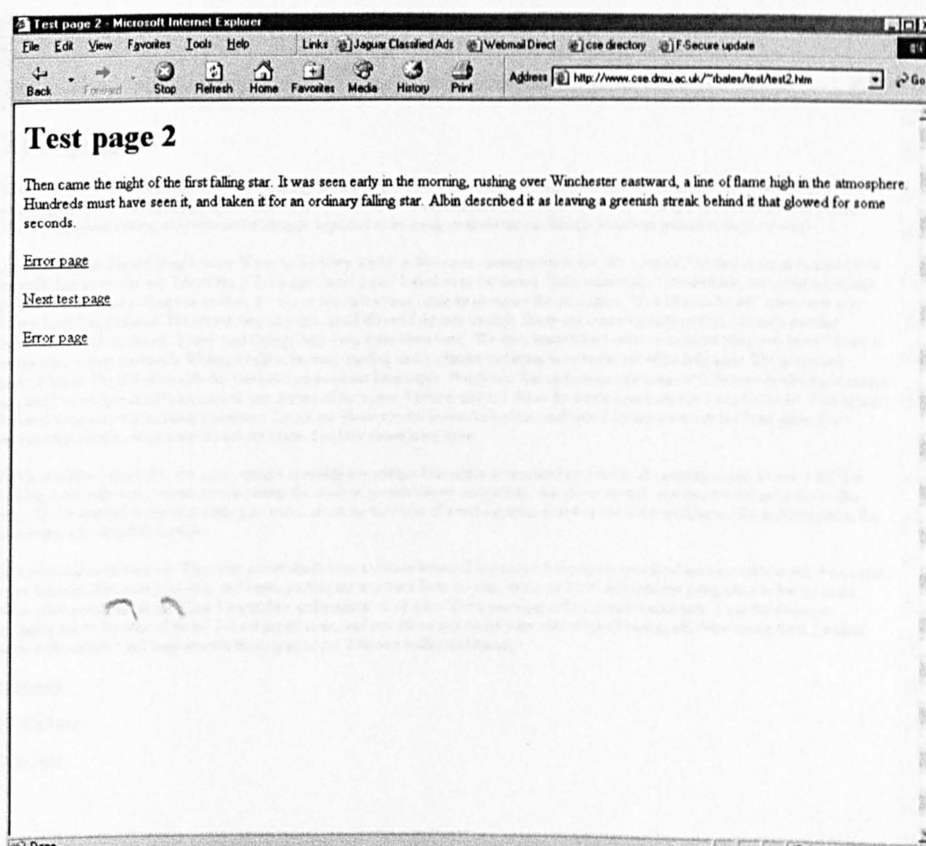
138	Click the [My Address] radio button	Radio button	S2				✓	
139	Click the [Send Message] button	Command button	S3				✓	
140	Click the [File] menu	Menu	S3				✓	
141	Click the [Save As] menu item	Menu	S3				✓	
142	Click the [Soft Keyboard] button on the task bar to show the Keyboard	Task bar button	S3				✓	
143-146	Type [test] in the filename edit box	Keyboard key	S4				✓	
147	Click the [Save] button	Command button	S3				✓	
148	Close Internet Explorer by clicking the [Close] button on the window title bar	Window control button	S2				✓	
149	Minimise the Soft Keyboard window by clicking the [Minimise] button on the window title bar	Window control button	S2				✓	
150	Drag the Internet Explorer [test] file icon into the [Test files] folder window	Icon	S4					✓

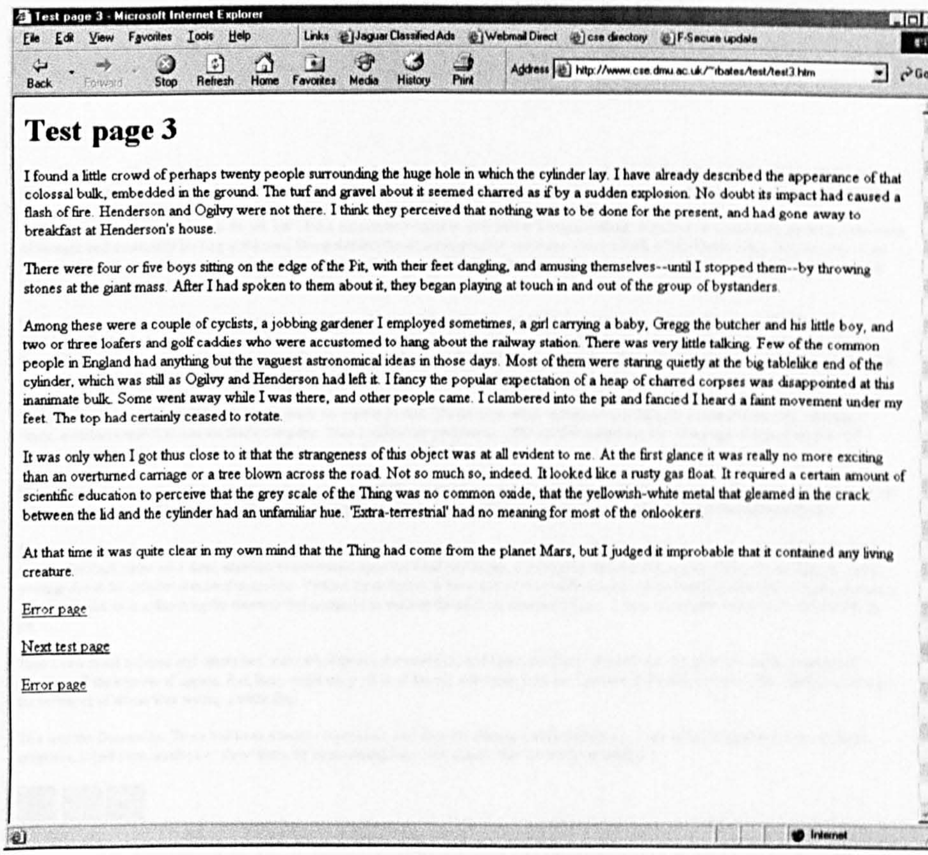
Figure A4.1 Pre-test set up





Internet Explorer test pages





Test page 3

I found a little crowd of perhaps twenty people surrounding the huge hole in which the cylinder lay. I have already described the appearance of that colossal bulk, embedded in the ground. The turf and gravel about it seemed charred as if by a sudden explosion. No doubt its impact had caused a flash of fire. Henderson and Ogilvy were not there. I think they perceived that nothing was to be done for the present, and had gone away to breakfast at Henderson's house.

There were four or five boys sitting on the edge of the Pit, with their feet dangling, and amusing themselves--until I stopped them--by throwing stones at the giant mass. After I had spoken to them about it, they began playing at touch in and out of the group of bystanders.

Among these were a couple of cyclists, a jobbing gardener I employed sometimes, a girl carrying a baby, Gregg the butcher and his little boy, and two or three loafers and golf caddies who were accustomed to hang about the railway station. There was very little talking. Few of the common people in England had anything but the vaguest astronomical ideas in those days. Most of them were staring quietly at the big tablelike end of the cylinder, which was still as Ogilvy and Henderson had left it. I fancy the popular expectation of a heap of charred corpses was disappointed at this inanimate bulk. Some went away while I was there, and other people came. I clambered into the pit and fancied I heard a faint movement under my feet. The top had certainly ceased to rotate.

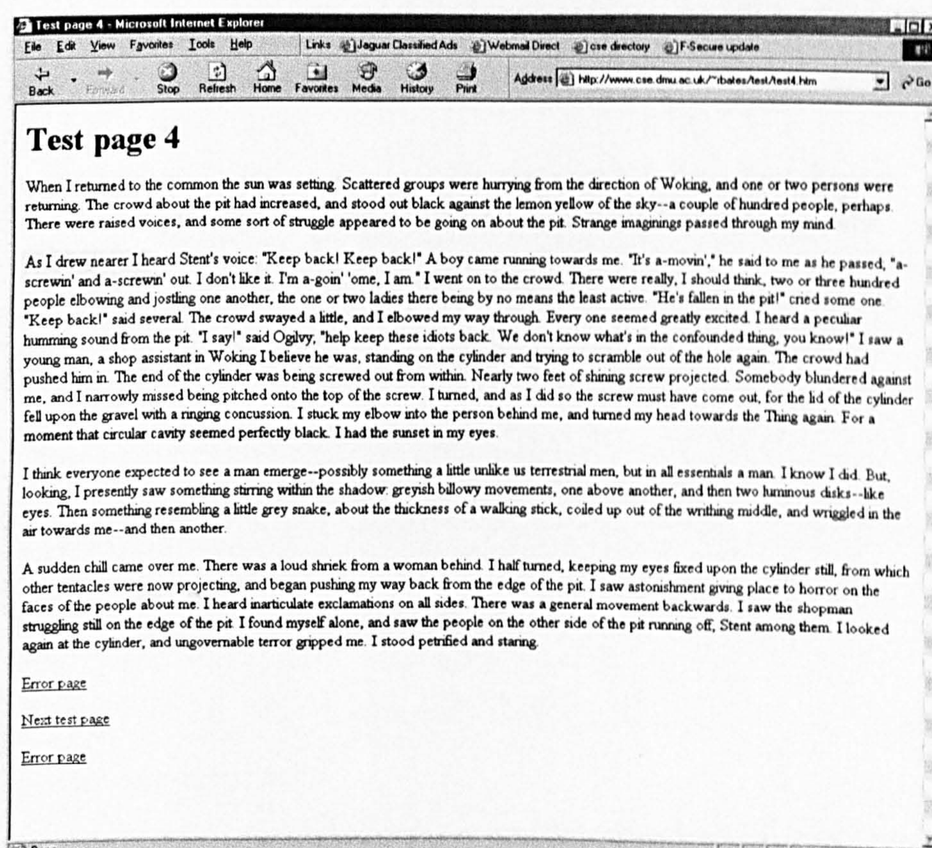
It was only when I got thus close to it that the strangeness of this object was at all evident to me. At the first glance it was really no more exciting than an overturned carriage or a tree blown across the road. Not so much so, indeed. It looked like a rusty gas float. It required a certain amount of scientific education to perceive that the grey scale of the Thing was no common oxide, that the yellowish-white metal that gleamed in the crack between the lid and the cylinder had an unfamiliar hue. 'Extra-terrestrial' had no meaning for most of the onlookers.

At that time it was quite clear in my own mind that the Thing had come from the planet Mars, but I judged it improbable that it contained any living creature.

Error page

Next test page

Error page



Test page 4

When I returned to the common the sun was setting. Scattered groups were hurrying from the direction of Woking, and one or two persons were returning. The crowd about the pit had increased, and stood out black against the lemon yellow of the sky--a couple of hundred people, perhaps. There were raised voices, and some sort of struggle appeared to be going on about the pit. Strange imaginings passed through my mind.

As I drew nearer I heard Stent's voice: "Keep back! Keep back!" A boy came running towards me. "It's a-movin'," he said to me as he passed, "a-screwin' and a-screwin' out. I don't like it. I'm a-goin' 'ome, I am." I went on to the crowd. There were really, I should think, two or three hundred people elbowing and jostling one another, the one or two ladies there being by no means the least active. "He's fallen in the pit!" cried some one. "Keep back!" said several. The crowd swayed a little, and I elbowed my way through. Every one seemed greatly excited. I heard a peculiar humming sound from the pit. "I say!" said Ogilvy; "help keep these idiots back. We don't know what's in the confounded thing, you know!" I saw a young man, a shop assistant in Woking I believe he was, standing on the cylinder and trying to scramble out of the hole again. The crowd had pushed him in. The end of the cylinder was being screwed out from within. Nearly two feet of shining screw projected. Somebody blundered against me, and I narrowly missed being pitched onto the top of the screw. I turned, and as I did so the screw must have come out, for the lid of the cylinder fell upon the gravel with a ringing concussion. I stuck my elbow into the person behind me, and turned my head towards the Thing again. For a moment that circular cavity seemed perfectly black. I had the sunset in my eyes.

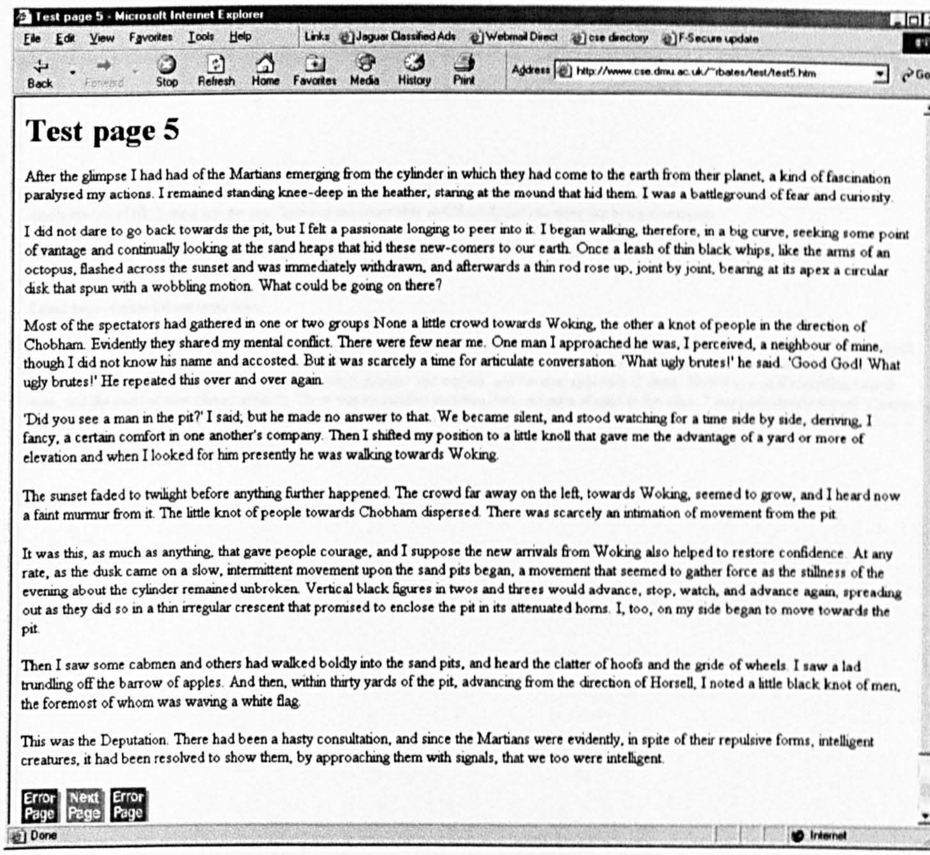
I think everyone expected to see a man emerge--possibly something a little unlike us terrestrial men, but in all essentials a man. I know I did. But, looking, I presently saw something stirring within the shadow: greyish billowy movements, one above another, and then two luminous disks--like eyes. Then something resembling a little grey snake, about the thickness of a walking stick, coiled up out of the writhing middle, and wriggled in the air towards me--and then another.

A sudden chill came over me. There was a loud shriek from a woman behind. I half turned, keeping my eyes fixed upon the cylinder still, from which other tentacles were now projecting, and began pushing my way back from the edge of the pit. I saw astonishment giving place to horror on the faces of the people about me. I heard inarticulate exclamations on all sides. There was a general movement backwards. I saw the shopman struggling still on the edge of the pit. I found myself alone, and saw the people on the other side of the pit running off, Stent among them. I looked again at the cylinder, and ungovernable terror gripped me. I stood petrified and staring.

Error page

Next test page

Error page



Test page 5 - Microsoft Internet Explorer

Test page 5

After the glimpse I had had of the Martians emerging from the cylinder in which they had come to the earth from their planet, a kind of fascination paralysed my actions. I remained standing knee-deep in the heather, staring at the mound that hid them. I was a battleground of fear and curiosity.

I did not dare to go back towards the pit, but I felt a passionate longing to peer into it. I began walking, therefore, in a big curve, seeking some point of vantage and continually looking at the sand heaps that hid these new-comers to our earth. Once a leash of thin black whips, like the arms of an octopus, flashed across the sunset and was immediately withdrawn, and afterwards a thin rod rose up, joint by joint, bearing at its apex a circular disk that spun with a wobbling motion. What could be going on there?

Most of the spectators had gathered in one or two groups None a little crowd towards Woking, the other a knot of people in the direction of Chobham. Evidently they shared my mental conflict. There were few near me. One man I approached he was, I perceived, a neighbour of mine, though I did not know his name and accosted. But it was scarcely a time for articulate conversation. 'What ugly brutes!' he said. 'Good God! What ugly brutes!' He repeated this over and over again.

'Did you see a man in the pit?' I said; but he made no answer to that. We became silent, and stood watching for a time side by side, deriving, I fancy, a certain comfort in one another's company. Then I shifted my position to a little knoll that gave me the advantage of a yard or more of elevation and when I looked for him presently he was walking towards Woking.

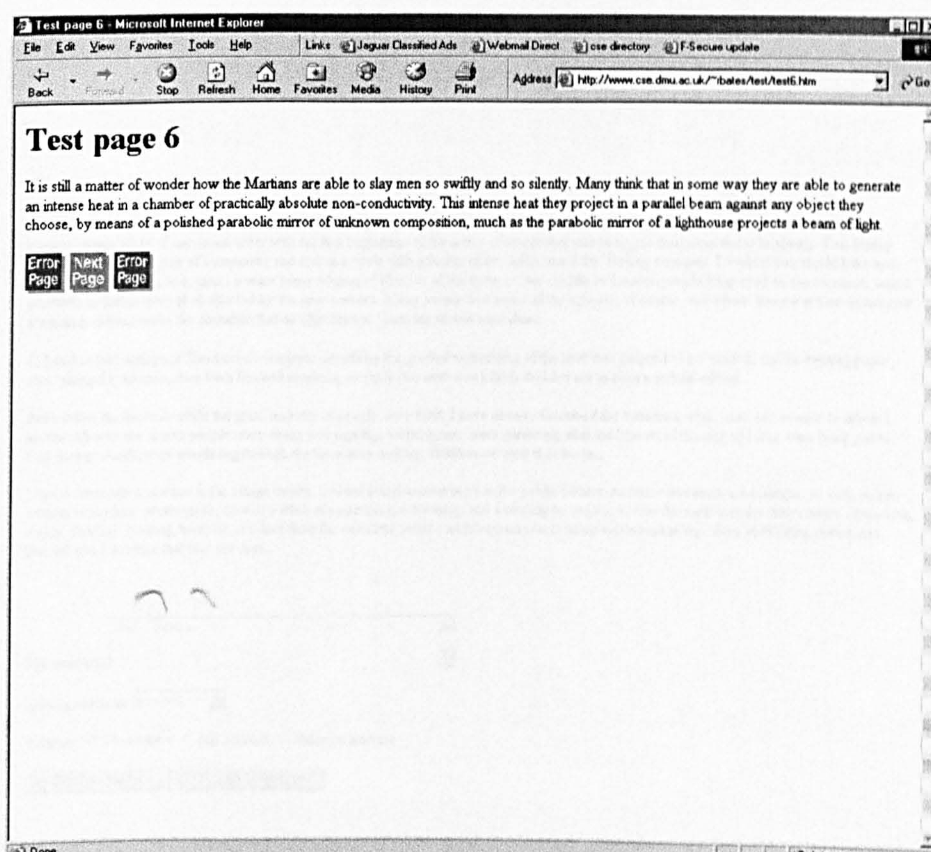
The sunset faded to twilight before anything further happened. The crowd far away on the left, towards Woking, seemed to grow, and I heard now a faint murmur from it. The little knot of people towards Chobham dispersed. There was scarcely an intimation of movement from the pit.

It was this, as much as anything, that gave people courage, and I suppose the new arrivals from Woking also helped to restore confidence. At any rate, as the dusk came on a slow, intermittent movement upon the sand pits began, a movement that seemed to gather force as the stillness of the evening about the cylinder remained unbroken. Vertical black figures in twos and threes would advance, stop, watch, and advance again, spreading out as they did so in a thin irregular crescent that promised to enclose the pit in its attenuated horns. I, too, on my side began to move towards the pit.

Then I saw some cabmen and others had walked boldly into the sand pits, and heard the clatter of hoofs and the gride of wheels. I saw a lad trundling off the barrow of apples. And then, within thirty yards of the pit, advancing from the direction of Horsell, I noted a little black knot of men, the foremost of whom was waving a white flag.

This was the Deputation. There had been a hasty consultation, and since the Martians were evidently, in spite of their repulsive forms, intelligent creatures, it had been resolved to show them, by approaching them with signals, that we too were intelligent.

Error Page | Next Page | Error Page

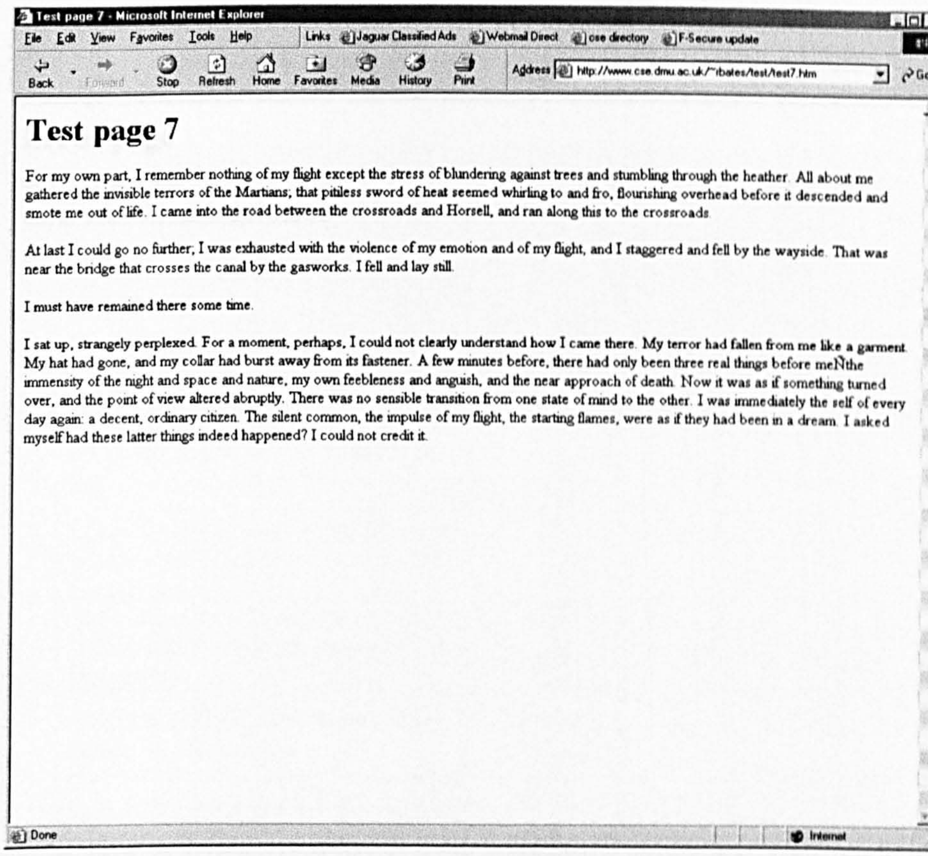


Test page 6 - Microsoft Internet Explorer

Test page 6

It is still a matter of wonder how the Martians are able to slay men so swiftly and so silently. Many think that in some way they are able to generate an intense heat in a chamber of practically absolute non-conductivity. This intense heat they project in a parallel beam against any object they choose, by means of a polished parabolic mirror of unknown composition, much as the parabolic mirror of a lighthouse projects a beam of light.

Error Page | Next Page | Error Page



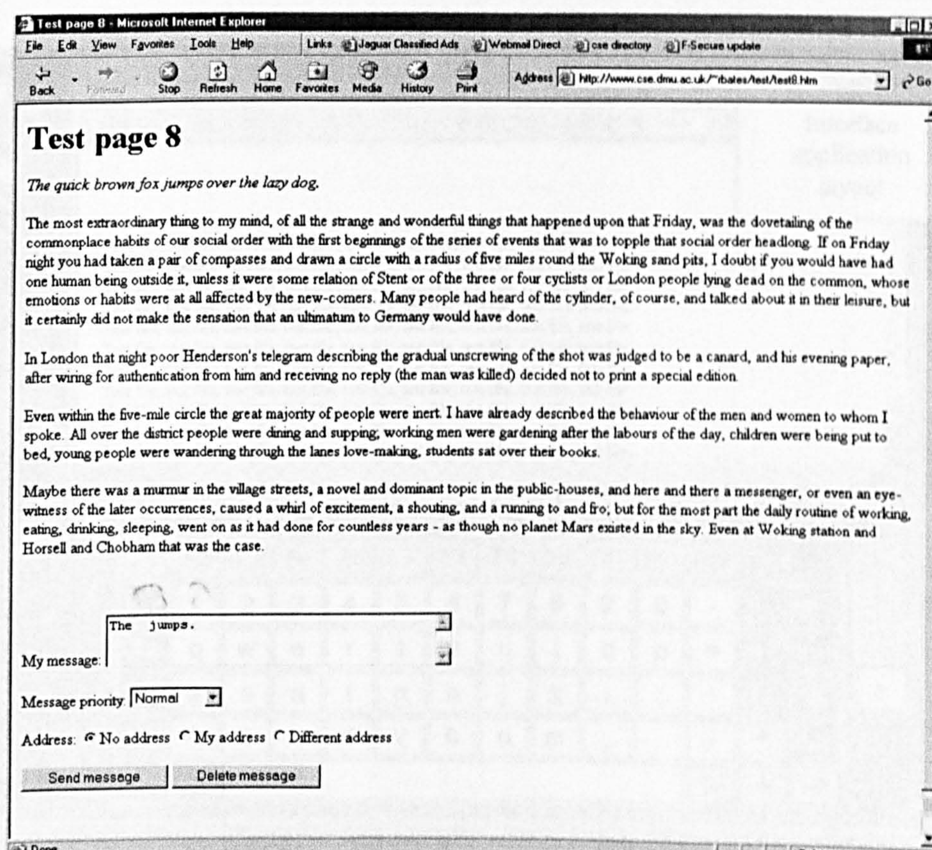
Test page 7

For my own part, I remember nothing of my flight except the stress of blundering against trees and stumbling through the heather. All about me gathered the invisible terrors of the Martians; that pitiless sword of heat seemed whirling to and fro, flourishing overhead before it descended and smote me out of life. I came into the road between the crossroads and Horsell, and ran along this to the crossroads.

At last I could go no further; I was exhausted with the violence of my emotion and of my flight, and I staggered and fell by the wayside. That was near the bridge that crosses the canal by the gasworks. I fell and lay still.

I must have remained there some time.

I sat up, strangely perplexed. For a moment, perhaps, I could not clearly understand how I came there. My terror had fallen from me like a garment. My hat had gone, and my collar had burst away from its fastener. A few minutes before, there had only been three real things before meÑthe immensity of the night and space and nature, my own feebleness and anguish, and the near approach of death. Now it was as if something turned over, and the point of view altered abruptly. There was no sensible transition from one state of mind to the other. I was immediately the self of every day again: a decent, ordinary citizen. The silent common, the impulse of my flight, the starting flames, were as if they had been in a dream. I asked myself had these latter things indeed happened? I could not credit it.



Test page 8

The quick brown fox jumps over the lazy dog.

The most extraordinary thing to my mind, of all the strange and wonderful things that happened upon that Friday, was the dovetailing of the commonplace habits of our social order with the first beginnings of the series of events that was to topple that social order headlong. If on Friday night you had taken a pair of compasses and drawn a circle with a radius of five miles round the Woking sand pits, I doubt if you would have had one human being outside it, unless it were some relation of Stent or of the three or four cyclists or London people lying dead on the common, whose emotions or habits were at all affected by the new-comers. Many people had heard of the cylinder, of course, and talked about it in their leisure, but it certainly did not make the sensation that an ultimatum to Germany would have done.

In London that night poor Henderson's telegram describing the gradual unscrewing of the shot was judged to be a canard, and his evening paper, after wiring for authentication from him and receiving no reply (the man was killed) decided not to print a special edition.

Even within the five-mile circle the great majority of people were inert. I have already described the behaviour of the men and women to whom I spoke. All over the district people were dining and supping; working men were gardening after the labours of the day, children were being put to bed, young people were wandering through the lanes love-making, students sat over their books.

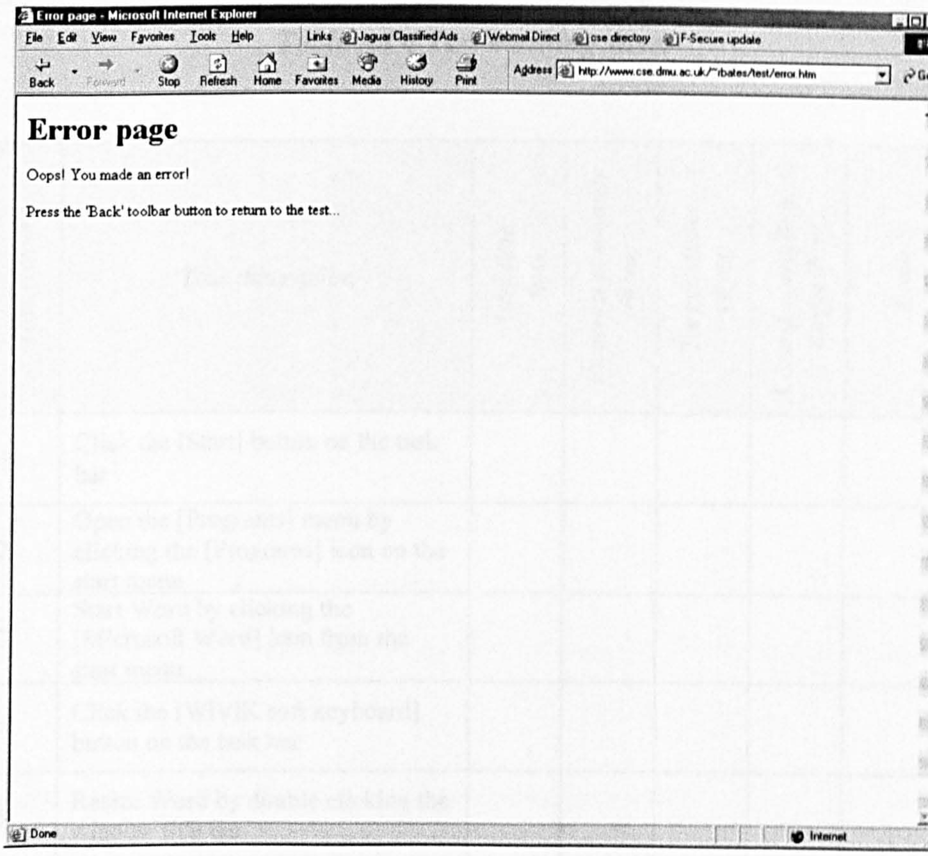
Maybe there was a murmur in the village streets, a novel and dominant topic in the public-houses, and here and there a messenger, or even an eye-witness of the later occurrences, caused a whirl of excitement, a shouting, and a running to and fro; but for the most part the daily routine of working, eating, drinking, sleeping, went on as it had done for countless years - as though no planet Mars existed in the sky. Even at Woking station and Horsell and Chobham that was the case.

My message: The jumps.

Message priority: Normal

Address: No address My address Different address

Send message | Delete message


Error page - Microsoft Internet Explorer
Error page
Oops! You made an error!
Press the 'Back' toolbar button to return to the test...
Done
Internet

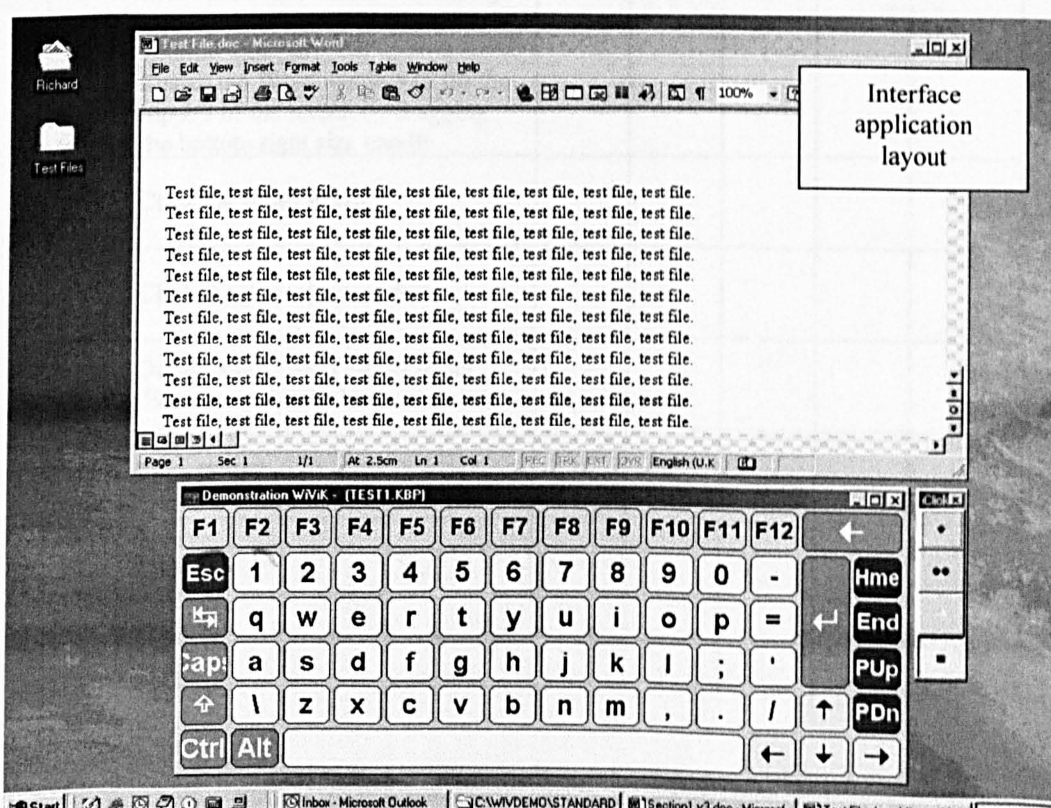

Interface application layout
Test File.doc - Microsoft Word
Richard
Test Files
Test file, test file, test file, test file, test file, test file, test file, test file, test file.
Demonstration WiVik - (TEST1.KBP)
Start
Inbox - Microsoft Outlook

Table A4.4 Test marking sheet

Task element	*Task description*	*Task time (ms)*	*Incorrect commands (# ms)*	*Target Misses (# ms)*	*Control corrections at target (# ms)*	*Notes*
1	Click the [Start] button on the task bar					
2	Open the [Programs] menu by clicking the [Programs] icon on the start menu					
3	Start Word by clicking the [Microsoft Word] icon from the start menu					
4	Click the [WiViK soft keyboard] button on the task bar					
5	Resize Word by double clicking the window title bar					
6	Move the Word window to the top left of the screen by dragging the window title bar					
7	Resize the Word window to fill the top 2/3 of the screen by dragging the bottom right size handle					
8	Click the [File] menu					
9	Click the [Open] menu item					
10	Double click the [Test File.doc] filename in the list box					
11	Type [Shift]					
12	Type 'T'					

Appendix Chapter 6

Figure A6.1 Consent form

Assessment of Questionnaire Design

**DE MONTFORT UNIVERSITY SCHOOL OF COMPUTING SCIENCE
EXPERIMENTAL SUBJECT CONSENT FORM:**

This experiment is designed determine a questionnaire design

The experimental procedure will involve:

- Rating questionnaire factors
- Completing a series of ratings
- The test will last approximately 30 minutes.

This experiment will ask you to assess what you feel is the score for a word, and then indicate on a line where you think that word lies.

You may:
At any reasonable time you may discontinue the tests for a break. Subjects may withdraw from the experiment at any time and will suffer no personal consequences as a result.

Your rights:
All data will remain anonymous and names will not be disclosed without the express written consent of the test subject.

The experiments are to be conducted by the experimenter ______________________________ in the HCI laboratory in the School of Computing Science at De Montfort University.

CONSENT

I ____ agree to participate in the above titled experiment.
I am satisfied that I have received a full explanation of experimental procedures and have had any queries answered to the extent of the experimenters current knowledge. I understand that I may terminate the experiment at any time and will suffer no personal consequences as a result.

Subject: ______________________________ Date: ___________ Witness: _________________________

Table A6.1 Bipolar questionnaire quantifier experimental results

	Quantifier	*Percentiles*			*Inter-quartile distance*
		50th	*25th*	*75th*	
1.	*Completely happy*	*100.0*	-	-	-
2.	Extremely happy	93.3	90.0	96.7	6.7
3.	Very happy	80.0	73.3	86.7	13.3
4.	Greatly happy	80.0	70.0	86.7	16.7
5.	Really happy	76.7	66.7	86.7	20.0
6.	Very much happy	76.7	63.3	86.7	23.3
7.	Considerably happy	66.7	60.0	76.7	16.7
8.	Happy	48.3	40.0	66.7	26.7
9.	Pretty much happy	46.7	30.0	66.7	36.7
10.	Not at all sad	41.7	0.0	70.0	70.0
11.	Fairly happy	38.3	30.0	50.0	20.0
12.	Quite happy	36.7	26.7	50.0	23.3
13.	Moderately happy	36.7	20.0	43.3	23.3
14.	Somewhat happy	26.7	13.3	36.7	23.3
15.	Just happy	25.0	6.7	43.3	36.7
16.	A little happy	13.3	10.0	20.0	10.0
17.	Slightly happy	13.3	10.0	16.7	6.7
18.	A bit happy	13.3	6.7	20.0	13.3
19.	Not very sad	10.0	-10.0	26.7	36.7
20.	Very slightly happy	6.7	3.3	10.0	6.7
21.	Scarcely happy	3.3	-3.3	10.0	13.3
22.	*Neither Happy nor Sad*	*0.0*	-	-	-
23.	Scarcely sad	-10.0	-16.7	0.0	16.7
24.	Very slightly sad	-10.0	-20.0	-6.7	13.3
25.	A bit sad	-20.0	-26.7	-10.0	16.7
26.	Slightly sad	-20.0	-26.7	-16.7	10.0
27.	A little sad	-20.0	-30.0	-13.3	16.7
28.	Not very happy	-20.0	-43.3	3.3	46.7
29.	Somewhat sad	-30.0	-40.0	-20.0	20.0
30.	Just sad	-40.0	-56.7	-13.3	43.3
31.	Not at all happy	-40.0	-76.7	-3.3	73.3
32.	Moderately sad	-50.0	-56.7	-36.7	20.0
33.	Quite sad	-51.7	-63.3	-36.7	26.7
34.	Fairly sad	-51.7	-63.3	-36.7	26.7
35.	Sad	-56.7	-66.7	-50.0	16.7
36.	Pretty much sad	-61.7	-73.3	-46.7	26.7
37.	Considerably sad	-73.3	-83.3	-63.3	20.0
38.	Very much sad	-86.7	-90.0	-76.7	13.3
39.	Greatly sad	-86.7	-90.0	-80.0	10.0
40.	Really sad	-86.7	-93.3	-76.7	16.7
41.	Very sad	-90.0	-93.3	-83.3	10.0
42.	Extremely sad	-96.7	-100.0	-93.3	6.7
43.	*Completely sad*	*-100.0*	-	-	-

Table A6.2 Unipolar questionnaire quantifier experimental results

	Quantifier	*Percentiles*			*Inter-quartile distance*
		50th	*25th*	*75th*	
1.	*Completely happy*	*100.0*	-	-	-
2.	Extremely happy	96.7	93.3	98.3	5.0
4.	Very much happy	87.5	81.7	91.7	10.0
3.	Very happy	87.5	78.3	93.3	15.0
5.	Really happy	85.8	78.3	93.3	15.0
6.	Greatly happy	85.0	75.0	93.3	18.3
7.	Considerably happy	73.3	61.7	85.0	23.3
8.	Pretty much happy	58.3	45.0	73.3	28.3
9.	Quite happy	56.7	41.7	65.0	23.3
10.	Happy	54.2	45.0	73.3	28.3
11.	Fairly happy	53.3	45.0	61.7	16.7
12.	Moderately happy	48.3	41.7	58.3	16.7
13.	Just happy	47.5	35.0	56.7	21.7
14.	Somewhat happy	45.8	35.0	56.7	21.7
15.	A bit happy	41.7	18.3	51.7	33.3
16.	A little happy	35.0	18.3	48.3	30.0
17.	Slightly happy	28.3	13.3	51.7	38.3
18.	Very slightly happy	18.3	5.0	43.3	38.3
19.	Scarcely happy	14.2	5.0	30.0	25.0
20.	Not very happy	6.7	1.7	18.3	16.7
21.	*Not at all happy*	*0.0*	-	-	-

Figure A6.2 Calculating an optimal scale

For a 5-point full-range bipolar scale:

Use the bipolar data (Appendix 1, Table 6).

Negative end anchor = 'Extremely Sad' = -96.7
Midpoint anchor = 'Neither Happy nor Sad' = 0
Positive end anchor = 'Extremely Happy' = 93.3

To obtain a symmetrical scale, take the average of the two endpoints as the full-scale value: (96.7 + 93.3) / 2 = 95

For a 5-point bipolar scale we get:
(2 * 95) / 4 = 47.5 per interval, giving intervals at: -95, -47.5, 0, + 47.5, +95

We already have the anchor-point quantifiers, so choosing the two symmetrical intermediate quantifiers and aggregating their distances from the interval point we have a close choices of:

'Sad/Happy' (-56.7 and +48.3) total distance from ideal point:
(56.7 – 47.5) + (48.3 - 47.5) = 9.2 + 0.8 = 10

'Moderately' (-50.0 and +36.7) total distance from ideal point:
(50.0 – 47.5) + (47.5 – 36.7) = 2.5 + 10.8 = 13.3

This suggests using 'Sad/Happy'. Looking at the potential overlaps with the other chosen quantifiers we see no overlap and a highest to lowest percentile separations of:

'Extremely Sad' to 'Sad' = 93.3 – 66.7 = 26.6
'Sad' to 'Neither Happy Nor Sad' = 50.0
'Neither Happy Nor Sad' to 'Happy' = 40.0
'Happy' to 'Extremely Happy' = 90.0 – 66.7 = 23.3
Total = 26.6 + 50.0 + 40.0 + 23.3 = 139.9

'Extremely Sad' to 'Moderately Sad' = 93.3 – 56.7 = 36.6
'Moderately Sad' to 'Neither Happy Nor Sad' = 36.7
'Neither Happy Nor Sad' to 'Moderately Happy' = 20.0
'Moderately Happy' to 'Extremely Happy' = 89.2 – 43.3 = 45.9
Total = 36.6 + 36.7 + 20.0 + 45.9 = 139.2

There is little difference between the candidate quantifiers. Finally, looking at the 25^{th} to 75^{th} interquartile distances:

'Sad/Happy' = 16.7 + 26.7 = 43.4
'Moderately' = 20.0 + 23.3 = 43.3

Again, there is little difference between the quantifiers. Since 'Sad/Happy' is the closest to the ideal point on the scale, it is chosen as the quantifier for the scale.

Appendix Chapter 7

Figure A7.1 Jitter generation loop

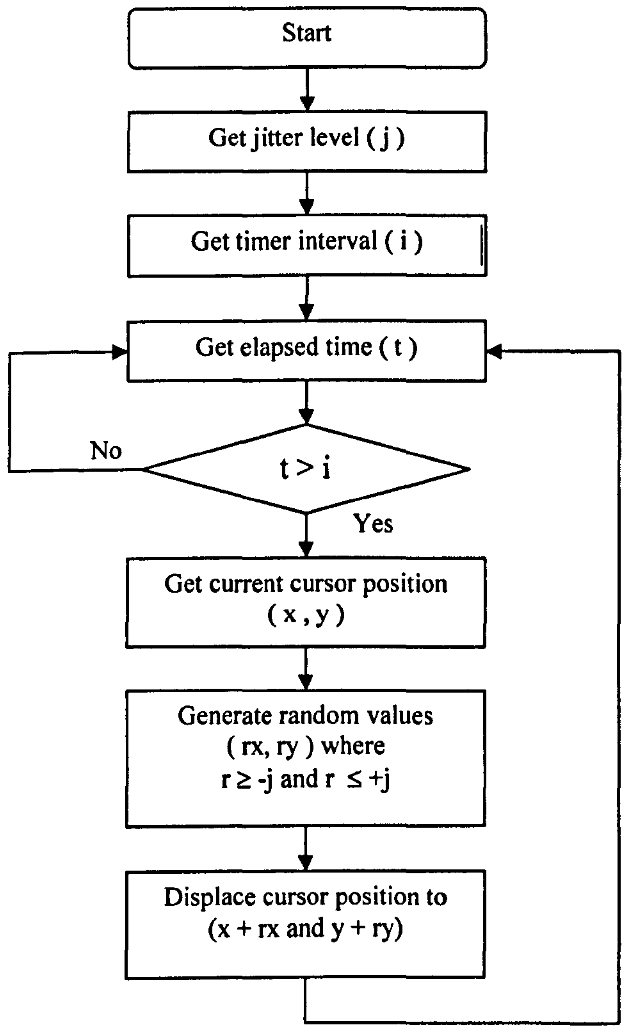


Figure A7.2 Consent form

Assessment of Jitter Discrimination

DE MONTFORT UNIVERSITY SCHOOL OF COMPUTING SCIENCE EXPERIMENTAL SUBJECT CONSENT FORM:

This experiment is designed determine the discrimination between differing target acquisition difficulties.

The experimental procedure will involve:

- Using a mouse to select a target.
- Assess which targets are more or less or the same difficult to select.
- The test will last approximately 30 minutes.

This experiment will ask you to use a desktop mouse to select a target, the mouse cursor may be 'jittery', making the task more difficult. After each task you will be asked to rate which targets are more, or less, or the same difficulty.

You may:
At any reasonable time you may discontinue the tests for a break. Subjects may withdraw from the experiment at any time and will suffer no personal consequences as a result.

Your rights:
All data will remain anonymous and names will not be disclosed without the express written consent of the test subject.

The experiments are to be conducted by the experimenter ______________________________ in the HCI laboratory in the School of Computing Science at De Montfort University.

CONSENT

I ______________________________________ agree to participate in the above titled experiment. I am satisfied that I have received a full explanation of experimental procedures and have had any queries answered to the extent of the experimenters current knowledge. I understand that I may terminate the experiment at any time and will suffer no personal consequences as a result.

Subject: ____________________________ Date: __________ Witness: ________________________

Table A7.1 Paired comparison of jitter test conditions

Test condition	*Baseline Jitter Level (Red or Blue Target)*	*Comparative Jitter Level (Red or Blue Target)*
1	R (2.0mm)	B (2.0mm)
2	R (2.0mm)	B (2.5mm)
3	R (2.0mm)	B (3.0mm)
4	R (2.0mm)	B (3.5mm)
5	B (2.0mm)	R (2.0mm)
6	B (2.0mm)	R (2.5mm)
7	B (2.0mm)	R (3.0mm)
8	B (2.0mm)	R (3.5mm)

Table A7.2 Paired comparison of jitter, balanced Latin-square design

Subject	*Test conditions*							
1	1	8	2	7	3	6	4	5
2	2	1	3	8	4	7	5	6
3	3	2	4	1	5	8	6	7
4	4	3	5	2	6	1	7	8
5	5	4	6	3	7	2	8	1
6	6	5	7	4	8	3	1	2
7	7	6	8	5	1	4	2	3
8	8	7	1	6	2	5	3	4

Table A7.3 Paired comparison of jitter, summary of test results

Baseline Red Jitter level	*Comparative Blue Jitter Level*	*Subjective evaluations*			*Percentage correct answers*
		Red more difficult than Blue	*Red and Blue the same difficulty*	*Blue more difficult than Red*	
2.0	2.0	28.1%	56.3%	15.6%	56.3%
2.0	2.5	9.4%	34.4%	56.3%	56.3%
2.0	3.0	6.3%	21.9%	71.9%	71.9%
2.0	3.5	0.0%	9.4%	90.6%	90.6%

Note: Results re-ordered to remove randomisation of test conditions

Figure A7.3 Consent form

Assessment of Task Difficulty Due to Jitter

DE MONTFORT UNIVERSITY SCHOOL OF COMPUTING SCIENCE EXPERIMENTAL SUBJECT CONSENT FORM:

This experiment is designed determine the rating of task difficulty between differing target acquisition tasks.

The experimental procedure will involve:

- Using a mouse to select a target.
- Assess the level of difficulty of each task.
- The test will last approximately 30 minutes.

This experiment will ask you to use a desktop mouse to select a target, the mouse cursor may be 'jittery', making the task more difficult. After each task you will be asked to rate the level of difficulty of the task on a questionnaire.

You may:
At any reasonable time you may discontinue the tests for a break. Subjects may withdraw from the experiment at any time and will suffer no personal consequences as a result.

Your rights:
All data will remain anonymous and names will not be disclosed without the express written consent of the test subject.

The experiments are to be conducted by the experimenter ______________________________ in the HCI laboratory in the School of Computing Science at De Montfort University.

CONSENT

I ____ agree to participate in the above titled experiment. I am satisfied that I have received a full explanation of experimental procedures and have had any queries answered to the extent of the experimenters current knowledge. I understand that I may terminate the experiment at any time and will suffer no personal consequences as a result.

Subject: ____________________________ Date: __________ Witness: ________________________

Table A7.4 Bipolar quantifier jitter test conditions

Test condition	*Questionnaire*
1	5pt Bipolar part labeled
2	7pt Bipolar part labeled
3	9pt Bipolar part labeled
4	11pt Bipolar part labeled
5	5pt Bipolar labeled
6	7pt Bipolar labeled
7	9pt Bipolar labeled
8	11pt Bipolar labeled

Table A7.5 Bipolar quantifier jitter test, balanced Latin-square design

Subject	*Conditions*							
1	1 1 2 3 4	8 2 4 1 3	2 1 4 3 2	7 2 1 3 4	3 4 1 3 2	6 3 2 1 4	4 1 4 2 3	5 3 1 2 4
2	2 1 3 2 4	1 2 3 4 1	3 2 4 3 1	8 4 2 3 1	4 2 1 4 3	7 3 1 4 2	5 4 2 1 3	6 3 2 4 1
3	3 3 4 2 1	2 4 1 2 3	4 4 3 1 2	1 1 3 4 2	5 1 2 4 3	8 4 3 2 1	6 2 1 4 3	7 3 1 2 4
4	4 1 3 4 2	3 2 3 4 1	5 4 3 1 2	2 3 4 2 1	6 2 1 3 4	1 2 4 3 1	7 2 3 1 4	8 4 2 1 3
5	5 1 3 2 4	4 1 4 3 2	6 2 4 1 3	3 3 2 4 1	7 4 1 3 2	2 3 4 1 2	8 3 1 4 2	1 4 1 2 3
6	6 4 3 2 1	5 1 4 2 3	7 1 2 4 3	4 3 1 2 4	8 4 1 3 2	3 1 4 2 3	1 2 4 1 3	2 3 4 1 2
7	7 2 1 3 4	6 1 2 3 4	8 2 3 4 1	5 3 2 1 4	1 3 4 2 1	4 1 3 4 2	2 3 1 4 2	3 3 2 4 1
8	8 1 3 2 4	7 1 4 3 2	1 2 3 1 4	6 4 2 1 3	2 4 1 2 3	5 4 3 1 2	3 2 1 4 3	4 2 4 3 1

Key: First number in cell is the test condition; the following 4 numbers are the order of the jitter levels to be applied: 1 = 0.0mm, 2 = 2.0mm, 3 = 3.5mm, 4 = 6.0mm

Table A7.6 Bipolar quantifier jitter test, summary of test results

Questionnaire	*Jitter levels*	*Percentiles* 50th	25th	75th	*Inter-quartile distance*
5pt Bipolar part labeled	1	8.0	4.0	8.0	4.0
	2	16.0	11.0	16.0	5.0
	3	12.0	11.0	16.0	5.0
	4	20.0	16.0	20.0	4.0
7pt Bipolar part labeled	1	2.9	2.9	5.7	2.9
	2	10.0	8.6	14.3	5.7
	3	14.3	13.6	14.3	0.7
	4	20.0	17.1	20.0	2.9
9pt Bipolar part labeled	1	2.2	2.2	4.4	2.2
	2	11.1	8.9	13.3	4.4
	3	13.3	10.6	13.9	3.3
	4	17.8	17.8	20.0	2.2
11pt Bipolar part labeled	1	3.6	3.6	5.5	1.8
	2	10.9	9.1	11.4	2.3
	3	12.7	10.5	14.5	4.1
	4	18.2	18.2	18.2	0.0
5pt Bipolar labeled	1	6.0	4.0	8.0	4.0
	2	12.0	8.0	12.0	4.0
	3	12.0	8.0	16.0	8.0
	4	20.0	19.0	20.0	1.0
7pt Bipolar labeled	1	2.9	2.9	5.7	2.9
	2	11.4	8.6	12.1	3.6
	3	14.3	14.3	14.3	0.0
	4	20.0	17.1	20.0	2.9
9pt Bipolar labeled	1	4.4	4.4	6.7	2.2
	2	11.1	8.9	13.3	4.4
	3	13.3	10.6	15.6	5.0
	4	17.8	17.2	17.8	0.6
11pt Bipolar labeled	1	3.6	3.6	4.1	0.5
	2	9.1	7.3	12.7	5.5
	3	10.0	7.3	13.2	5.9
	4	18.2	16.4	18.2	1.8

Note: Results normalised to 20-interval scale

Table A7.7 Unipolar quantifier jitter test conditions

Test condition	*Questionnaire*
1	5pt Unipolar part labeled
2	7pt Unipolar part labeled
3	9pt Unipolar part labeled
4	11pt Unipolar part labeled
5	5pt Unipolar labeled
6	7pt Unipolar labeled
7	9pt Unipolar labeled
8	11pt Unipolar labeled
9	20pt Unipolar part labeled

Table A7.8 Unipolar quantifier jitter test, balanced Latin-square design

Subject	*Conditions*								
1	1 1 2 3 4	9 2 4 1 3	2 1 4 3 2	8 2 1 3 4	3 4 1 3 2	7 3 2 1 4	4 1 4 2 3	6 3 1 2 4	5 2 3 1 4
2	2 1 3 2 4	1 2 3 4 1	3 2 4 3 1	9 4 2 3 1	4 2 1 4 3	8 3 1 4 2	5 4 2 1 3	7 3 2 4 1	6 3 4 1 2
3	3 3 4 2 1	2 4 1 2 3	4 4 3 1 2	1 1 3 4 2	5 1 2 4 3	9 4 3 2 1	6 2 1 4 3	8 3 1 2 4	7 3 2 1 4
4	4 1 3 4 2	3 2 3 4 1	5 4 3 1 2	2 3 4 2 1	6 2 1 3 4	1 2 4 3 1	7 2 3 1 4	9 4 2 1 3	8 1 2 3 4
5	5 1 3 2 4	4 1 4 3 2	6 2 4 1 3	3 3 2 4 1	7 4 1 3 2	2 3 4 1 2	8 3 1 4 2	1 4 1 2 3	9 4 2 3 1
6	6 4 3 2 1	5 1 4 2 3	7 1 2 4 3	4 3 1 2 4	8 4 1 3 2	3 1 4 2 3	9 2 4 1 3	2 3 4 1 2	1 1 2 4 3
7	7 2 1 3 4	6 1 2 3 4	8 2 3 4 1	5 3 2 1 4	9 3 4 2 1	4 1 3 4 2	1 3 1 4 2	3 3 2 4 1	2 4 2 3 1
8	8 1 3 2 4	7 1 4 3 2	9 2 3 1 4	6 4 2 1 3	1 4 1 2 3	5 4 3 1 2	2 2 1 4 3	4 2 4 3 1	3 4 3 2 1
9	9 1 3 4 2	8 3 1 4 2	1 4 1 2 3	7 2 1 4 3	2 4 2 1 3	6 3 1 2 4	3 2 3 4 1	5 4 3 1 2	4 1 2 3 4

Key: First number in cell is the test condition; the following 4 numbers are the order of the jitter levels to be applied: 1 = 0.0mm, 2 = 2.0mm, 3 = 3.5mm, 4 = 6.0mm

Table A7.9 Unipolar quantifier jitter test, summary of test results

Questionnaire	*Jitter levels*	Percentiles			Inter-quartile distance
		50th	*25th*	*75th*	
5pt Unipolar part labeled	1	4.0	4.0	5.0	1.0
	2	8.0	8.0	12.0	4.0
	3	12.0	8.0	16.0	8.0
	4	18.0	16.0	20.0	4.0
7pt Unipolar part labeled	1	2.9	2.9	8.6	5.7
	2	8.6	5.7	11.4	5.7
	3	12.9	8.6	14.3	5.7
	4	17.1	16.4	20.0	3.6
9pt Unipolar part labeled	1	2.2	2.2	4.4	2.2
	2	8.9	6.7	11.1	4.4
	3	12.2	10.6	13.3	2.8
	4	15.6	13.3	18.3	5.0
11pt Unipolar part labeled	1	1.8	1.8	3.6	1.8
	2	9.1	7.3	10.9	3.6
	3	10.9	8.6	12.7	4.1
	4	18.2	15.9	18.6	2.7
5pt Unipolar labeled	1	4.0	4.0	4.0	0.0
	2	8.0	8.0	12.0	4.0
	3	12.0	8.0	16.0	8.0
	4	20.0	16.0	20.0	4.0
7pt Unipolar labeled	1	2.9	2.9	2.9	0.0
	2	8.6	7.9	11.4	3.6
	3	11.4	10.7	14.3	3.6
	4	17.1	17.1	20.0	2.9
9pt Unipolar labeled	1	4.4	2.2	4.4	2.2
	2	8.9	6.7	11.1	4.4
	3	11.1	8.9	13.3	4.4
	4	15.6	15.0	18.3	3.3
11pt Unipolar labeled	1	3.6	1.8	3.6	1.8
	2	9.1	7.3	12.7	5.5
	3	10.9	7.3	10.9	3.6
	4	17.3	14.5	18.2	3.6
20pt Unipolar not labeled	1	3.0	2.0	5.0	3.0
	2	10.0	8.0	11.0	3.0
	3	11.0	8.0	12.0	4.0
	4	17.0	15.0	18.0	3.0

Note: Results normalised to 20-interval scale

Table A7.10 Combined quantifier jitter test, summary of test results

Questionnaire	*Jitter levels*	*Percentiles*			*Inter-quartile distance*
		50th	*25th*	*75th*	
5pt part labelled	1	4.0	4.0	8.0	4.0
	2	12.0	8.0	16.0	8.0
	3	12.0	8.0	16.0	8.0
	4	20.0	16.0	20.0	4.0
7pt part labeled	1	2.9	2.9	5.7	2.9
	2	8.6	8.6	12.1	3.6
	3	14.3	11.4	14.3	2.9
	4	18.6	17.1	20.0	2.9
9pt part labeled	1	2.2	2.2	4.4	2.2
	2	10.0	6.7	13.3	6.7
	3	13.3	10.6	13.3	2.8
	4	17.8	15.6	20.0	4.4
11pt not labeled	1	3.6	1.8	4.1	2.3
	2	9.1	7.3	10.9	3.6
	3	10.9	9.1	13.2	4.1
	4	18.2	16.4	18.2	1.8
5pt labeled	1	4.0	4.0	8.0	4.0
	2	12.0	8.0	12.0	4.0
	3	12.0	8.0	16.0	8.0
	4	20.0	16.0	20.0	4.0
7pt labeled	1	2.9	2.9	3.6	0.7
	2	11.4	8.6	11.4	2.9
	3	14.3	11.4	14.3	2.9
	4	20.0	17.1	20.0	2.9
9pt labeled	1	4.4	2.2	5.0	2.8
	2	8.9	6.7	11.1	4.4
	3	12.2	8.9	13.3	4.4
	4	17.8	15.6	17.8	2.2
11pt labeled	1	3.6	1.8	3.6	1.8
	2	9.1	7.3	12.7	5.5
	3	10.9	7.3	10.9	3.6
	4	18.2	15.9	18.2	2.3

Note: Results normalised to 20-interval scale

Figure A7.4 Workload assessment questionnaire

Workload Assessment Questionnaire

Please circle the 'X' closest to your opinion

← low workload ratings high →

1. How much *physical* effort or activity was required to operate the system?

X X	X	X	X	X	X	X
Extremely low physical effort	Considerably low physical effort	Somewhat low physical effort	Neither high nor low physical effort	Somewhat high physical effort	Considerably high physical effort	Extremely high physical effort

2. How much *mental* effort or concentration was required to operate the system?

X	X	X	X	X	X	X
Extremely low mental effort	Considerably low mental effort	Somewhat low mental effort	Neither high nor low mental effort	Somewhat high mental effort	Considerably high mental effort	Extremely high mental effort

3. How much *temporal* or time pressure did you feel under?

X	X	X	X	X	X	X
Extremely low temporal pressure	Considerably low temporal pressure	Somewhat low temporal pressure	Neither high nor low temporal pressure	Somewhat high temporal pressure	Considerably high temporal pressure	Extremely high temporal pressure

4. What level of *frustration* did you experience when using the system?

X	X	X	X	X	X	X
Extremely low level frustration	Considerably low level frustration	Somewhat low level frustration	Neither high nor low level frustration	Somewhat high level frustration	Considerably high level frustration	Extremely high level frustration

5. How well do you think you *performed* on the test, was your performance high or low?

X	X	X	X	X	X	X
Extremely high performance	Considerably high performance	Somewhat high performance	Neither high nor low performance	Somewhat low performance	Considerably low performance	Extremely low performance

6. Overall, how hard did you have to *work* during the test?

X	X	X	X	X	X	X
Not at all hard work	Scarcely hard work	A little hard work	Moderately hard work	Considerably hard work	Really hard work	Extremely hard work

Figure A7.5 Comfort assessment questionnaire

Comfort Assessment Questionnaire

Please circle the 'X' closest to your opinion

← comfortable comfort ratings uncomfortable →

1. Do you have *headache* pain of any kind?

X	X	X	X	X	X	X
Not at all painful	Scarcely painful	A little painful	Moderately painful	Considerably painful	Really painful	Extremely painful

2. Do your *eyes* feel tired, strained or painful?

X	X	X	X	X	X	X
Not at all tired, strained or painful	Scarcely Tired, strained or painful	A little tired, strained or painful	Moderately tired, strained or painful	Considerably tired, strained or painful	Really tired, strained or painful	Extremely tired, strained or painful

3. Do your *facial muscles* feel tired, strained or painful?

X	X	X	X	X	X	X
Not at all tired, strained or painful	Scarcely Tired, strained or painful	A little tired, strained or painful	Moderately tired, strained or painful	Considerably tired, strained or painful	Really tired, strained or painful	Extremely tired, strained or painful

4. Does your *mouth* or *throat* feel tired, strained or painful?

X	X	X	X	X	X	X
Not at all tired, strained or painful	Scarcely Tired, strained or painful	A little tired, strained or painful	Moderately tired, strained or painful	Considerably tired, strained or painful	Really tired, strained or painful	Extremely tired, strained or painful

5. Does your *neck* feel tired, stiff or painful?

X	X	X	X	X	X	X
Not at all tired, stiff or painful	Scarcely tired, stiff or painful	A little tired, stiff or painful	Moderately tired, stiff or painful	Considerably tired, stiff or painful	Really tired, stiff or painful	Extremely tired, stiff or painful

6. Overall, during the test did using the system make you feel comfortable or uncomfortable?

X	X	X	X	X	X	X
Extremely comfortable	Considerably comfortable	Somewhat comfortable	Neither comfortable nor uncomfortable	Somewhat uncomfortable	Considerably uncomfortable	Extremely uncomfortable

Figure A7.6 Ease of use assessment questionnaire

Ease of Use Questionnaire

Please circle the 'X' closest to your opinion

← easy ease of use ratings hard →

1. Did you find that *pointing* with the system was accurate or inaccurate?

X	X	X	X	X	X	X
Extremely accurate pointing	Considerably accurate pointing	Somewhat accurate pointing	Neither accurate nor inaccurate pointing	Somewhat inaccurate pointing	Considerably inaccurate pointing	Extremely inaccurate pointing

2. Did you find that the *speed* of *pointing* with the system was fast or slow?

X	X	X	X	X	X	X
Extremely fast pointing	Considerably fast pointing	Somewhat fast pointing	Neither fast nor slow pointing	Somewhat slow pointing	Considerably slow pointing	Extremely slow pointing

3. Did you find that *clicking* with the system was accurate or inaccurate?

X	X	X	X	X	X	X
Extremely accurate clicking	Considerably accurate clicking	Somewhat accurate clicking	Neither accurate nor inaccurate clicking	Somewhat inaccurate clicking	Considerably inaccurate clicking	Extremely inaccurate clicking

4. Did you find that the *speed* of *clicking* with the system was fast or slow?

X	X	X	X	X	X	X
Extremely fast clicking	Considerably fast clicking	Somewhat fast clicking	Neither fast nor slow clicking	Somewhat slow clicking	Considerably slow clicking	Extremely slow clicking

5. Did you find it easy or difficult to *control* other parts of the system?

X	X	X	X	X	X	X
Extremely easy to control	Considerably easy to control	Somewhat easy to control	Neither easy nor difficult to control	Somewhat difficult to control	Considerably difficult to control	Extremely difficult to control

6. Overall, did you find the system easy or difficult to use?

X	X	X	X	X	X	X
Extremely easy to use	Considerably easy to use	Somewhat easy to use	Neither easy nor difficult to use	Somewhat difficult To use	Considerably difficult to use	Extremely difficult to use

Appendix Chapter 8

Table A8.1 IP for four test jitter levels, summary of test results

Jitter levels	IP Percentiles 50th	25th	75th	IP Inter-quartile distance
1	3.33	3.09	3.63	0.54
2	2.43	2.03	2.77	0.74
3	1.86	1.49	2.24	0.75
4	0.00	0.00	0.00	0.00

Figure A8.1 Fitts target test jitter level IP regression

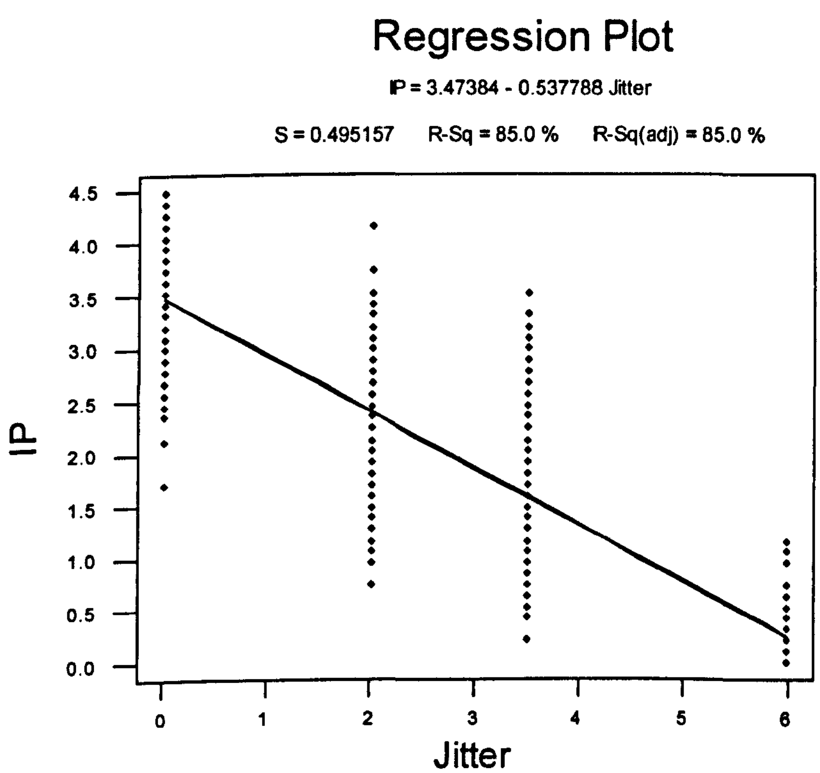


Figure A8.2 Consent form

Validation of the Test Method

DE MONTFORT UNIVERSITY SCHOOL OF COMPUTING SCIENCE EXPERIMENTAL SUBJECT CONSENT FORM:

This experiment is designed determine the validity of a test method. You will be asked to rate the difficulty of performing some tasks on an interface with a desktop hand mouse.

The experimental procedure will involve:

- Using a mouse to perform some 'real world' tasks.
- Assess the level of difficulty of each set of tasks.
- The test will last approximately 45 minutes.

This experiment will ask you to use a desktop mouse to perform some simple 'real world' tasks on an interface. The mouse cursor may be 'jittery', making the task more difficult. After each task you will be asked to rate the level of difficulty of the tasks on a questionnaire.

You may:
At any reasonable time you may discontinue the tests for a break. Subjects may withdraw from the experiment at any time and will suffer no personal consequences as a result.

Your rights:
All data will remain anonymous and names will not be disclosed without the express written consent of the test subject.

The experiments are to be conducted by the experimenter ______________________________ in the HCI laboratory in the School of Computing Science at De Montfort University.

CONSENT

I ____ agree to participate in the above titled experiment.
I am satisfied that I have received a full explanation of experimental procedures and have had any queries answered to the extent of the experimenters current knowledge. I understand that I may terminate the experiment at any time and will suffer no personal consequences as a result.

Subject: ______________________________ Date: __________ Witness: _________________________

Table A8.2 Test method jitter test conditions

Test condition	*Jitter Level (mm)*
1	0.0
2	2.0
3	3.5
4	6.0

Table A8.3 Test method, incomplete Latin-square design

Subject	*Test conditions*			
1	1	4	2	3
2	2	1	3	4
3	3	2	4	1
4	4	3	1	2
5	1	4	2	3
6	2	1	3	4

Table A8.4 IP for four test jitter levels, summary of test results

Jitter levels	*IP Percentiles*			*IP Inter-quartile distance*
	50^{th}	*25^{th}*	*75^{th}*	
1	83.3	78.1	86.2	8.1
2	80.6	64.5	86.1	21.6
3	71.4	44.4	83.3	38.9
4	44.1	14.2	66.4	52.2

Figure A8.3 Test method jitter level IP regression

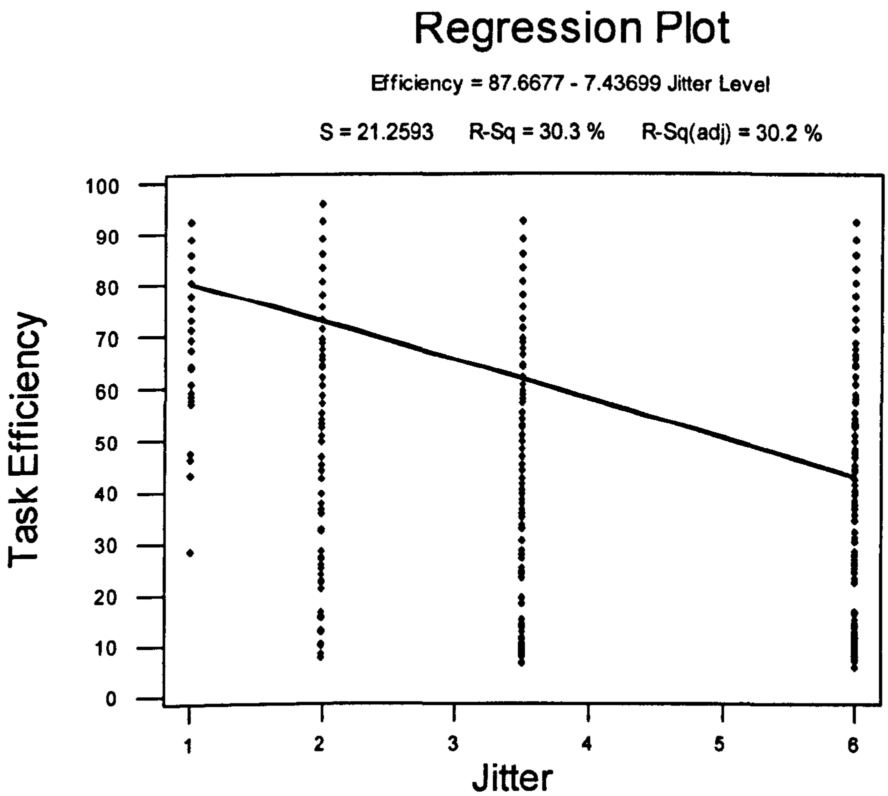


Table A8.5 Questionnaire results

Questionnaire factors		*Questionnaire Results (1-7) for Jitter level* 1	2	3	4
Workload:	Physical	1.3	3.8	4.0	5.3
	Mental	2.0	3.3	5.1	6.7
	Temporal	1.8	2.5	3.1	4.0
	Frustration	1.6	4.0	5.6	6.8
	Performance	1.8	4.1	4.8	5.8
Comfort:	Head	6.7	6.5	6.7	6.8
	Eye	6.3	6.5	6.8	6.3
	Facial	6.6	6.8	6.8	6.8
	Mouth	7.0	6.8	7.0	7.0
	Neck	6.0	6.7	6.8	6.7
Ease of Use:	Pointing Accuracy	6.5	4.0	2.4	1.1
	Pointing Speed	6.1	4.5	4.5	3.1
	Clicking Accuracy	6.3	4.1	2.7	2.1
	Clicking Speed	6.3	4.7	4.7	3.5
	System Control	6.0	5.8	5.7	4.1
Overall:	Workload	1.7	3.6	4.6	5.7
	Comfort	6.5.	6.7	6.8	6.2
	Ease of Use	6.2	4.6	3.9	2.8

Appendix Chapter 9

Figure A9.1. Head mouse software driver

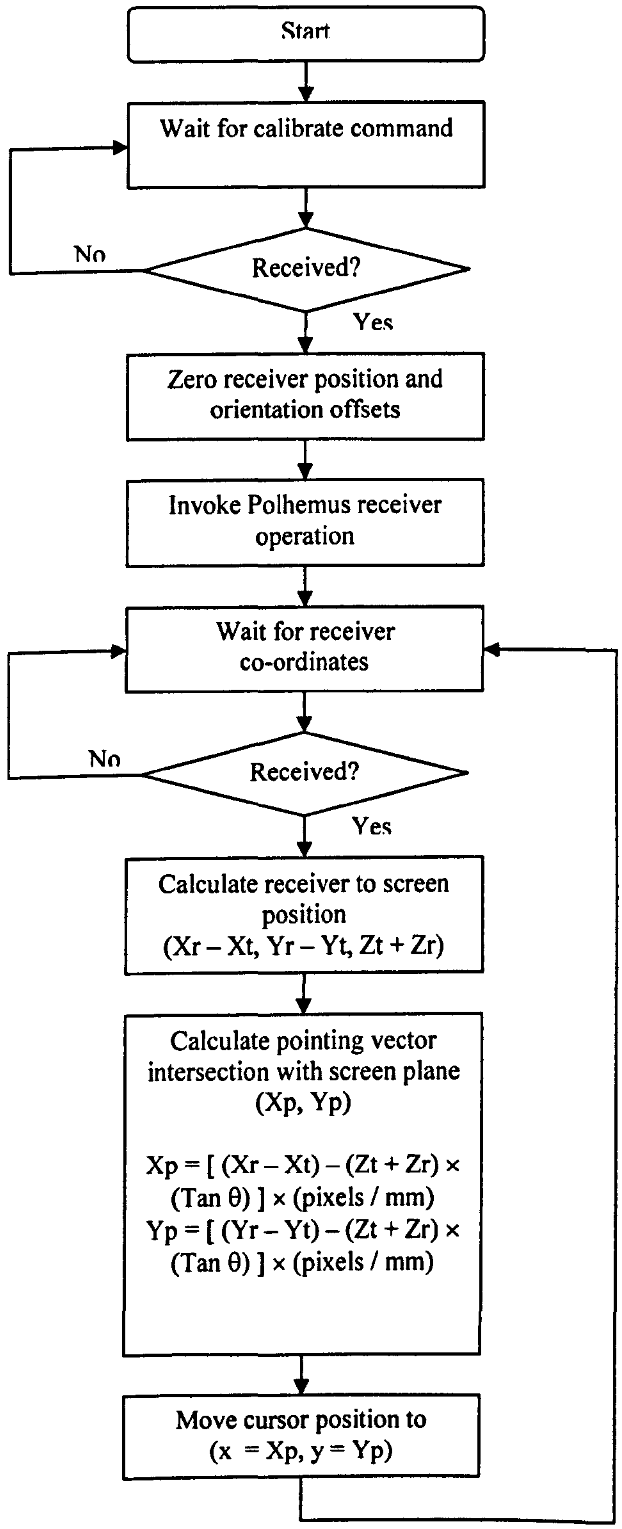
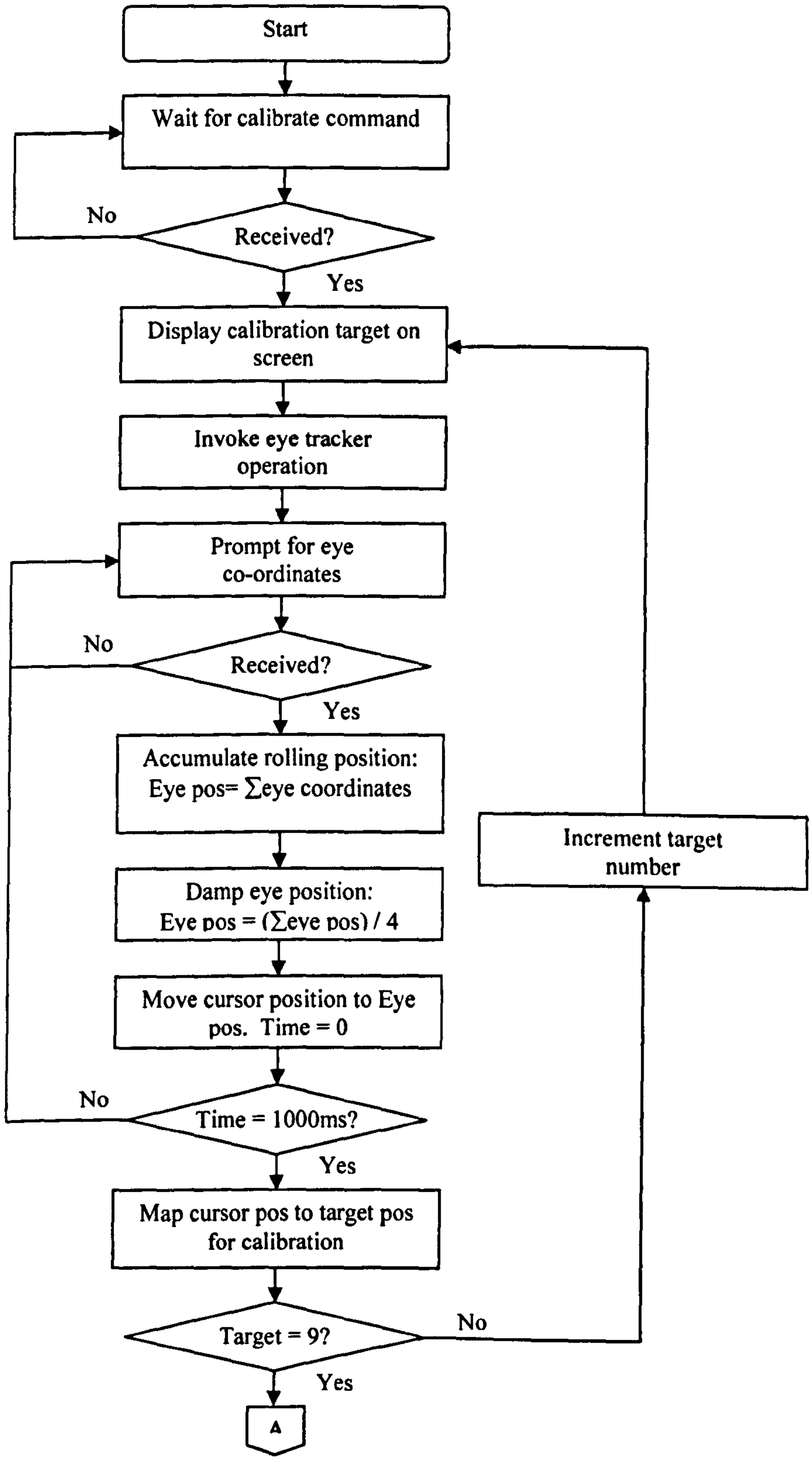
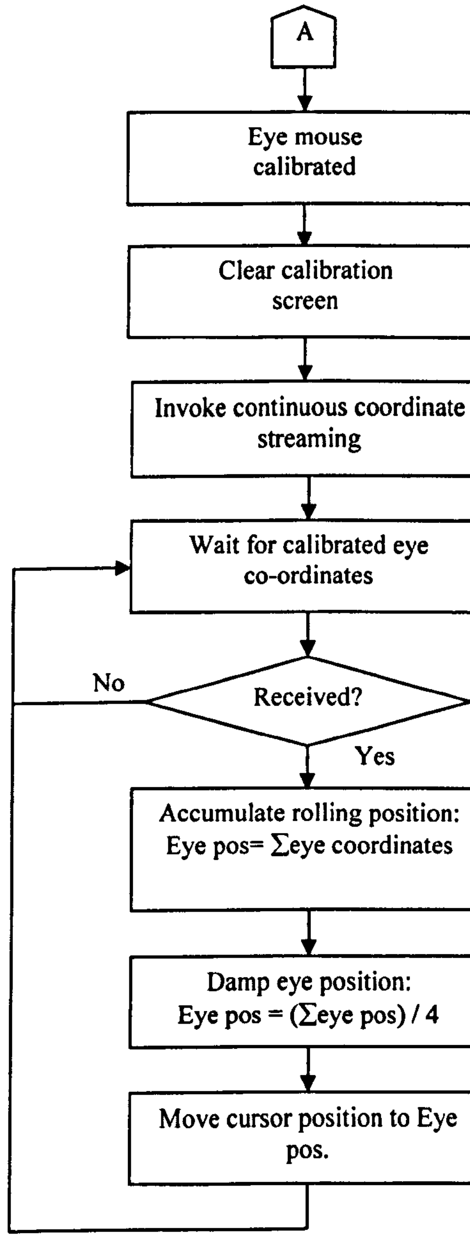


Figure A9.2. Eye mouse software driver




A
Eye mouse calibrated
Clear calibration screen
Invoke continuous coordinate streaming
Wait for calibrated eye co-ordinates
No
Received?
Yes
Accumulate rolling position: Eye pos= Σeye coordinates
Damp eye position: Eye pos = (Σeye pos) / 4
Move cursor position to Eye pos.

Appendix Chapter 10

Figure A10.1 Consent form

Usability of Hand, Head and Eye Mice on a Standard GUI

DE MONTFORT UNIVERSITY SCHOOL OF COMPUTING SCIENCE EXPERIMENTAL SUBJECT CONSENT FORM:

This experiment is designed to study the usability of hand, head and eye mice when operating on a standard graphical user interface.

The experimental procedure will involve:

- Calibrating and using a head mouse based on an electromagnetic head tracker
- Calibrating and using an eye-mouse based on an infrared eye-tracker.
- Completing a practice period lasting 5-10 minutes.
- Completing five sets of tests.
- Completing a series of short questionnaires.
- Sitting still in front of the eye-tracker for several periods of approximately 10 minutes.
- The tests will last approximately 30 minutes.
- Short rest periods of 20 minutes will take place between test sessions.

The experiment will expose the subject to electromagnetic fields and infrared light shone at the eyes. This is within safe exposure limits to the knowledge of the experimenter and the manufacturer of the head and eye tracking equipment and to their knowledge should not damage the body or eyes.

You may:
At any reasonable time you may discontinue the tests for a break. Subjects may withdraw from the experiment at any time and will suffer no personal consequences as a result.

Your rights:
All data will remain anonymous and names will not be disclosed without the express written consent of the test subject.

The experiments are to be conducted by the experimenter ______________________________ in the HCI laboratory in the School of Computing Science at De Montfort University.

CONSENT

I ____ agree to participate in the above titled experiment.
I am satisfied that I have received a full explanation of experimental procedures, their possible complications and side effects and have had any queries answered to the extent of the experimenters current knowledge.
I understand that I may terminate the experiment at any time and will suffer no personal consequences as a result.

Subject: ______________________________ Date: __________ Witness: __________________________

Table A10.1 Device assessment conditions

Test condition	*Device*
1	Hand mouse baseline
2	Monomodal Head mouse
3	Monomodal Eye mouse
4	Multimodal Head mouse
5	Multimodal Eye mouse

Table A10.2 Device assessment order, incomplete Latin-square design

Subject	*Test conditions*				
1	1	5	2	4	3
2	2	1	3	5	4
3	3	2	4	1	5
4	4	3	5	2	1
5	5	4	1	3	2
6	1	5	2	4	3

Table A10.3 Device efficiency

Device	*Metric*	*Percentiles*			*Inter-quartile distance*
		50th	*25th*	*75th*	
Hand mouse baseline	Overall Efficiency	83.3	75.7	86.2	10.5
Monomodal Head mouse	"	56.4	43.3	67.2	23.9
Monomodal Eye mouse	"	42.9	11.4	63.2	51.8
Multimodal Head mouse	"	65.2	47.3	75.3	28.0
Multimodal Eye mouse	"	51.1	17.8	73.2	55.4
Hand mouse baseline	Word Domain Efficiency	83.3	78.1	86.2	8.1
Monomodal Head mouse	"	56.5	44.0	66.8	22.8
Monomodal Eye mouse	"	43.2	10.8	63.0	52.2
Multimodal Head mouse	"	65.2	47.3	75.2	27.9
Multimodal Eye mouse	"	51.5	19.3	71.8	52.5
Hand mouse baseline	Browser Domain Efficiency	80.6	75.7	86.2	10.5
Monomodal Head mouse	"	55.4	43.1	67.3	24.2
Monomodal Eye mouse	"	42.9	13.2	63.2	50.0
Multimodal Head mouse	"	65.1	47.2	75.5	28.3
Multimodal Eye mouse	"	51.1	15.0	74.8	59.8

Table A10.4. Statistics of task domain efficiency

Device	*Domain*	*Efficiency (%)*	*Mann-Whitney between-domain comparisons*
Head Dwell	Word Browser	56.5 [a] 43.2 [a]	$p = 0.675$
Eye Dwell	Word Browser	65.2 [b] 51.5 [b]	$p = 0.515$
Head Click	Word Browser	55.4 [c] 42.9 [c]	$p = 0.767$
Eye Click	Word Browser	65.1 [d] 51.1 [d]	$p = 0.810$

Data with the same letter are not significantly different ($p > 0.050$)

Table A10.5 Statistics of overall device efficiency

Device	*Efficiency (%)*
Hand	83.3
Head Dwell	56.4
Eye Dwell	42.9
Head Click	65.2
Eye Click	51.1

Wilcoxon between-device comparisons				
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*
Hand	-	-	-	-
Head Dwell	$p < 0.050$	-	-	-
Eye Dwell	$p < 0.050$	$p < 0.050$	-	-
Head Click	$p < 0.050$	$p < 0.050$	$p < 0.050$	-
Eye Click	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$

Data with the same letter are not significantly different ($p > 0.050$)

Table A10.6 Statistics of device pointing accuracy

Subject	*Pooled pointing accuracy (VA)*		*Device efficiency (%)*			
			Monomodal		*Multimodal*	
	Head	*Eye* [a, b]	*Head*	*Eye* [a]	*Head*	*Eye* [b]
S1	0.161	0.559	63.1	60.2	67.9	69.3
S2	0.402	0.720	55.4	49.1	60.8	60.3
S3	0.519	0.936	51.9	42.5	64.6	49.8
S4	0.251	1.005	51.5	25.7	55.6	36.0
S5	0.253	1.533	50.2	32.8	54.2	37.6
S6	0.307	1.375	47.0	25.4	52.4	24.3

Data with the same letter are significantly correlated ($p < 0.050$)

Spearman between-device efficiency to pointing accuracy correlations		
Device / metric	*Head point accuracy*	*Eye point accuracy*
Mono head	S=-0.143 $p = 0.787$	-
Mono eye	-	S=-0.829 $p < 0.050$
Multi head	S=0.086 $p = 0.872$	-
Multi eye	-	S=1.000 $p < 0.050$

Table A10.7 Device task time

Device	*Metric*	*Time per Task (ms)*
Hand mouse baseline	Total Task Time	1246
Monomodal Head mouse	"	3489
Monomodal Eye mouse	"	3668
Multimodal Head mouse	"	2537
Multimodal Eye mouse	"	3289
Hand mouse baseline	Productive Task Time	1234
Monomodal Head mouse	"	3106
Monomodal Eye mouse	"	1646
Multimodal Head mouse	"	2069
Multimodal Eye mouse	"	1699
Hand mouse baseline	Incorrect Commands Task Time	4
Monomodal Head mouse	"	13
Monomodal Eye mouse	"	62
Multimodal Head mouse	"	9
Multimodal Eye mouse	"	14
Hand mouse baseline	Target Misses Task Time	3
Monomodal Head mouse	"	64
Monomodal Eye mouse	"	155
Multimodal Head mouse	"	32
Multimodal Eye mouse	"	113
Hand mouse baseline	Control Corrections Task Time	4
Monomodal Head mouse	"	305
Monomodal Eye mouse	"	1751
Multimodal Head mouse	"	427
Multimodal Eye mouse	"	1356

Table A10.8 Statistics of device task time

Device	*Task time (ms)*
Hand	1246
Head Dwell	3489 [a]
Eye Dwell	3668 [a]
Head Click	2537
Eye Click	3289

Wilcoxon between-device comparisons				
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*
Hand	-	-	-	-
Head Dwell	$p<0.050$	-	-	-
Eye Dwell	$p<0.050$	$p=0.233$	-	-
Head Click	$p<0.050$	$p<0.050$	$p<0.050$	-
Eye Click	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$

Data with the same letter are not significantly different ($p > 0.050$)

Table A10.9 Device task quality

Device	*Metric*	*Quality*
Hand mouse baseline	Overall quality (1-5)	4.90
Monomodal Head mouse	"	4.26
Monomodal Eye mouse	"	3.25
Multimodal Head mouse	"	4.23
Multimodal Eye mouse	"	3.42
Hand mouse baseline	Incorrect Commands Count/Task	0.004
Monomodal Head mouse	"	0.026
Monomodal Eye mouse	"	0.102
Multimodal Head mouse	"	0.023
Multimodal Eye mouse	"	0.033
Hand mouse baseline	Target Misses Count/Task	0.007
Monomodal Head mouse	"	0.107
Monomodal Eye mouse	"	0.284
Multimodal Head mouse	"	0.081
Multimodal Eye mouse	"	0.218
Hand mouse baseline	Control Corrections Count/Task	0.070
Monomodal Head mouse	"	0.482
Monomodal Eye mouse	"	1.072
Multimodal Head mouse	"	0.576
Multimodal Eye mouse	"	1.145

Table A10.10 Statistics of device task quality

Device	*Task quality (1-5)*	*Device rank*
Hand	4.90	1
Head Dwell	4.26 [a]	=2
Eye Dwell	3.25	5
Head Click	4.23 [a]	=2
Eye Click	3.42	4

Wilcoxon between-device comparisons				
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*
Hand	-	-	-	-
Head Dwell	$p<0.050$	-	-	-
Eye Dwell	$p<0.050$	$p<0.050$	-	-
Head Click	$p<0.050$	$p=0.194$	$p<0.050$	-
Eye Click	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$

Data with the same letter are not significantly different ($p > 0.050$)

Table A10.11 Statistics of satisfaction

Device	*Satis-faction (1-7)*	*Device rank*
Hand	6.20	1
Head Dwell	4.36 [a]	3
Eye Dwell	2.93 [b]	5
Head Click	4.73	2
Eye Click	3.90 [a, b]	4

Wilcoxon between-device comparisons				
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*
Hand	-	-	-	-
Head Dwell	p <0.050	-	-	-
Eye Dwell	p <0.050	p <0.050	-	-
Head Click	p <0.050	p <0.050	p <0.050	-
Eye Click	p <0.050	*p=0.521*	*p =0.148*	p <0.050

Data with the same letter are not significantly different (p > 0.050)

Table A10.12 Statistics of satisfaction categories - Workload

Device	*Workload (1-7)*
Hand	1.80
Head Dwell	4.10
Eye Dwell	5.80 [a]
Head Click	3.80
Eye Click	5.10 [a]

Data with the same letter are not significantly different (p > 0.050)

Wilcoxon between-device comparisons				
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*
Head Dwell	p <0.050	-	-	-
Eye Dwell	p<0.050	p <0.050	-	-
Head Click	p<0.050	p <0.050	p <0.050	-
Eye Click	p<0.050	p <0.050	*p=0.295*	p <0.050

Table A10.13 Statistics of satisfaction categories - Comfort

Device	*Comfort (1-7)*
Hand	6.50 ^{a}
Head Dwell	5.90 ^{b}
Eye Dwell	4.50 ^{c}
Head Click	6.30 ^{b}
Eye Click	5.20 ^{c}

Data with the same letter are not significantly different ($p > 0.050$)

Wilcoxon between-device comparisons				
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*
Head Dwell	p <0.050	-	-	-
Eye Dwell	p<0.050	p <0.050	-	-
Head Click	*p =0.834*	*p =0.281*	p <0.050	-
Eye Click	p<0.050	p <0.050	*p =0.281*	p <0.050

Table A10.14 Statistics of satisfaction categories – Ease of Use

Device	*Ease of Use (1-7)*
Hand	6.30
Head Dwell	3.50 ^{a}
Eye Dwell	2.60
Head Click	4.30 ^{abc}
Eye Click	4.40 ^{bc}

Data with the same letter are not significantly different ($p > 0.050$)

Wilcoxon between-device comparisons				
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*
Head Dwell	p <0.050	-	-	-
Eye Dwell	p<0.050	p <0.050	-	-
Head Click	p <0.050	*p=0.059*	p <0.050	-
Eye Click	p<0.050	*p =0.093*	p <0.050	*p =0.295*

Appendix Chapter 11

Table A11.1 Device task efficiency by target size

Device	*Metric*	*Efficiency (%)*
Hand mouse baseline	Efficiency at Target Size 0.3° at 60cm	66.6
Monomodal Head mouse	"	30.0
Monomodal Eye mouse	"	9.6
Multimodal Head mouse	"	29.8
Multimodal Eye mouse	"	17.1
Hand mouse baseline	Efficiency at Target Size 0.6° at 60cm	78.0
Monomodal Head mouse	"	48.9
Monomodal Eye mouse	"	17.8
Multimodal Head mouse	"	52.4
Multimodal Eye mouse	"	37.7
Hand mouse baseline	Efficiency at Target Size 0.9° at 60cm	78.0
Monomodal Head mouse	"	50.9
Monomodal Eye mouse	"	32.3
Multimodal Head mouse	"	56.9
Multimodal Eye mouse	"	40.3
Hand mouse baseline	Efficiency at Target Size 1.2° at 60cm	83.3
Monomodal Head mouse	"	62.3
Monomodal Eye mouse	"	51.2
Multimodal Head mouse	"	72.7
Multimodal Eye mouse	"	57.4

Table A11.2 Statistics of device efficiency by target size

Device	*Efficiency (%) by Target Size (degrees at 60cm distance)*			
	0.3°	*0.6°*	*0.9°*	*1.2°*
Hand	66.6	78.1 [a]	78.1 [a]	83.3
Head Dwell	30.0	48.9 [b]	50.9 [b]	62.3
Eye Dwell	9.6	17.8 [c]	32.3 [c]	51.2
Head Click	29.8	52.4	56.9	72.7
Eye Click	17.1	37.7 [d]	40.3 [d]	57.4

Data within the same device with the same letter are not significantly different ($p > 0.050$)

Mann-Whitney within-device comparisons				
Device	*Target Size*	*0.3°*	*0.6°*	*0.9°*
Hand	0.6°	$p<0.050$	-	-
	0.9°	$p<0.050$	$p=0.741$	-
	1.2°	$p<0.050$	$p<0.050$	$p<0.050$
Head Dwell	0.6°	$p<0.050$	-	-
	0.9°	$p<0.050$	$p=0.556$	-
	1.2°	$p<0.050$	$p<0.050$	$p<0.050$
Eye Dwell	0.6°	$p<0.050$	-	-
	0.9°	$p<0.050$	$p=0.062$	-
	1.2°	$p<0.050$	$p<0.050$	$p<0.050$
Head Click	0.6°	$p<0.050$	-	-
	0.9°	$p<0.050$	$p<0.050$	-
	1.2°	$p<0.050$	$p<0.050$	$p<0.050$
Eye Click	0.6°	$p<0.050$	-	-
	0.9°	$p<0.050$	$p=0.107$	-
	1.2°	$p<0.050$	$p<0.050$	$p<0.050$

Table A11.3 Device quality by target size

Device	*Metric*	*Count per task for Target Size (at 60cm)*			
		0.3°	*0.6°*	*0.9°*	*1.2°*
Hand mouse baseline	Total Quality				
"	Incorrect Commands	0.00	0.00	0.00	0.00
"	Target Misses	0.16	0.00	0.00	0.00
"	Control Corrections	0.30	0.21	0.07	0.00
Monomodal Head mouse	Total Quality				
"	Incorrect Commands	0.11	0.050	0.04	0.00
"	Target Misses	0.66	0.21	0.12	0.02
"	Control Corrections	1.05	0.70	0.67	0.29
Monomodal Eye mouse	Total Quality				
"	Incorrect Commands	0.08	0.19	0.10	0.07
"	Target Misses	1.08	0.51	0.41	0.10
"	Control Corrections	1.58	1.21	1.21	0.93
Multimodal Head mouse	Total Quality				
"	Incorrect Commands	0.08	0.02	0.02	0.01
"	Target Misses	0.25	0.18	0.10	0.02
"	Control Corrections	1.66	0.94	0.61	0.36
Multimodal Eye mouse	Total Quality				
"	Incorrect Commands	0.08	0.02	0.03	0.02
"	Target Misses	0.30	0.37	0.24	0.15
"	Control Corrections	2.02	1.35	1.35	0.93

Table A11.4 Device task time by target size

Device	*Metric*	*Time (ms) for Target Size (at 60cm)*			
		0.3°	*0.6°*	*0.9°*	*1.2°*
Hand mouse baseline	Total Task Time	1958	1429	1502	1035
"	Productive Task Time	1793	1429	1495	1029
"	Incorrect Commands Task Time	55	0	2	3
"	Target Misses Task Time	55	0	3	0
"	Control Corrections Task Time	55	0	2	3
"	Feedback point time	-	-	-	-
"	Calibration time	-	-	-	-
Monomodal Head mouse	Total Task Time	5442	3983	3896	3127
"	Productive Task Time	3788	3319	3306	2909
"	Incorrect Commands Task Time	55	25	22	3
"	Target Misses Task Time	544	101	79	12
"	Control Corrections Task Time	1000	483	434	148
"	Calibration time	53	53	53	53
"	Feedback point time	-	-	-	-
Monomodal Eye mouse	Total Task Time	6046	4714	4341	3380
"	Productive Task Time	2011	1487	1925	1555
"	Incorrect Commands Task Time	55	132	50	46
"	Target Misses Task Time	538	276	210	67
"	Control Corrections Task Time	3297	2625	2023	1260
"	Calibration time	131	131	131	131
"	Feedback point time	11	60	0	317
Multimodal Head mouse	Total Task Time	5036	3311	2827	2097
"	Productive Task Time	2561	2373	2299	1847
"	Incorrect Commands Task Time	38	7	11	6
"	Target Misses Task Time	88	74	40	11
"	Control Corrections Task Time	2294	803	422	178
"	Calibration time	52	52	52	52
"	Feedback point time	-	-	-	-
Multimodal Eye mouse	Total Task Time	6007	4203	4100	2761
"	Productive Task Time	2294	1791	2059	1481
"	Incorrect Commands Task Time	55	10	17	11
"	Target Misses Task Time	100	278	84	75
"	Control Corrections Task Time	3427	1903	1809	855
"	Calibration time	128	128	128	128
"	Feedback point time	0	91	0	207

Table A11.5 Device efficiency by interaction technique

Device	*Interaction Technique Efficiency (%)*				
	Restricted			*Unrestricted*	
	Single	*Double*	*Drag*	*Single*	*Drag*
Hand	80.6 [a]	76.3	64.9 [b]	83.3 [a]	62.5 [b]
Head Dwell	48.2 [c]	48.6 [c]	36.2 [d]	60.3	33.7 [d]
Eye Dwell	47.4 [e]	42.2 [e]	12.1 [f]	45.4	8.5 [f]
Head Click	57.2 [g]	49.5 [g]	30.9 [h]	68.9	39.9 [h]
Eye Click	30.8 [i]	30.4 [i]	27.1 [j]	52.7	17.1 [j]

Data within the same device with the same letter are not significantly different ($p > 0.050$)

Table A11.6 Statistics of efficiency by interaction technique

Mann-Whitney within-device comparisons							
Device	*Interaction Technique*		*Restricted*			*Unrestricted*	
			Single	*Double*	*Drag*	*Single*	*Drag*
Hand	Restrict.	*Single*	-	-	-	-	-
		Double	$p < 0.050$	-	-	-	-
		Drag	$p < 0.050$	$p < 0.050$	-	-	-
	Unrest.	*Single*	$p = 0.080$	$p < 0.050$	$p < 0.050$	-	-
		Drag	$p < 0.050$	$p < 0.050$	$p = 0.411$	$p < 0.050$	-
Head dwell	Restrict.	*Single*	-	-	-	-	-
		Double	$p = 0.515$	-	-	-	-
		Drag	$p < 0.050$	$p < 0.050$	-	-	-
	Unrest.	*Single*	$p < 0.050$	$p = 0.050$	$p < 0.050$	-	-
		Drag	$p < 0.050$	$p < 0.050$	$p = 0.773$	$p < 0.050$	-
Eye dwell	Restrict.	*Single*	-	-	-	-	-
		Double	$p = 0.350$	-	-	-	-
		Drag	$p < 0.050$	$p < 0.050$	-	-	-
	Unrest.	*Single*	$p = 0.370$	$p = 0.514$	$p < 0.050$	-	-
		Drag	$p < 0.050$	$p < 0.050$	$p = 0.757$	$p < 0.050$	-
Head click	Restrict.	*Single*	-	-	-	-	-
		Double	$p = 0.528$	-	-	-	-
		Drag	$p < 0.050$	$p < 0.050$	-	-	-
	Unrest.	*Single*	$p < 0.050$	$p < 0.050$	$p < 0.050$	-	-
		Drag	$p < 0.050$	$p < 0.050$	$p = 0.183$	$p < 0.050$	-
Eye click	Restrict.	*Single*	-	-	-	-	-
		Double	$p = 0.912$	-	-	-	-
		Drag	$p = 0.358$	$p = 0.245$	-	-	-
	Unrest.	*Single*	$p = 0.005$	$p < 0.050$	$p = 0.050$	-	-
		Drag	$p = 0.060$	$p = 0.057$	$p = 0.449$	$p < 0.050$	-

Table A11.7 Device efficiency by subject experience

Device	*Efficiency (%) by subject experience (Low, Medium, High)* L	M	H
Hand	-	-	83.3
Head Dwell	51.6	54.5	63.9 [a]
Eye Dwell	19.7	31.8	61.1 [a]
Head Click	55.0	66.9	73.0 [b]
Eye Click	21.2	48.4	73.5 [b]

Data within the same device with the same letter are not significantly different ($p > 0.050$)

Table A11.8 Statistics of efficiency by subject experience

Wilcoxon within-device comparisons			
Device	*Experience*	*H*	*M*
Head Dwell	M	p <0.050	-
	L	p <0.050	p <0.050
Eye Dwell	M	p <0.050	-
	L	p <0.050	p <0.050
Head Click	M	p <0.050	-
	L	p <0.050	p <0.050
Eye Click	M	p <0.050	-
	L	p <0.050	p <0.050

Mann-Whitney between-device, within experience comparisons				
Device	*Head Dwell*	*Eye Dwell*	*Head Click*	*Hand*
	High Experience			
Head Dwell	-	-	-	p <0.050
Eye Dwell	p =0.142	-	-	p <0.050
Head Click	p <0.050	p <0.050	-	p <0.050
Eye Click	p <0.050	p <0.050	p =0.215	p <0.050
	Medium Experience			
Eye Dwell	p <0.050	-	-	-
Head Click	p <0.050	p <0.050	-	-
Eye Click	p <0.050	p <0.050	p <0.050	-
	Low Experience			
Eye Dwell	p <0.050	-	-	-
Head Click	p =0.050	p <0.050	-	-
Eye Click	p <0.050	p =0.445	p <0.050	-

Table A11.9 Device quality by subject experience

Device	*Metric*	*Count per task for Target Size (at 60cm)*		
		Low	*Medium*	*High*
Hand mouse baseline	Incorrect Commands	-	-	0.00
"	Target Misses	-	-	0.00
"	Control Corrections	-	-	0.07
Monomodal Head mouse	Incorrect Commands	0.04	0.02	0.01
"	Target Misses	0.17	0.10	0.04
"	Control Corrections	0.64	0.53	0.26
Monomodal Eye mouse	Incorrect Commands	0.16	0.10	0.04
"	Target Misses	0.41	0.29	0.15
"	Control Corrections	1.29	1.34	0.58
Multimodal Head mouse	Incorrect Commands	0.02	0.02	0.02
"	Target Misses	0.10	0.10	0.04
"	Control Corrections	0.80	0.52	0.40
Multimodal Eye mouse	Incorrect Commands	0.050	0.04	0.01
"	Target Misses	0.39	0.19	0.07
"	Control Corrections	1.67	1.31	0.45

Table A11.10 Device task time by subject experience

Device	*Metric*	*Time (ms) for Target Size (at 60cm)*		
		Low	*Medium*	*High*
Hand mouse baseline	Total Task Time	-	-	1246
"	Productive Task Time	-	-	1234
"	Incorrect Commands Task Time	-	-	4
"	Target Misses Task Time	-	-	2
"	Control Corrections Task Time	-	-	4
"	Feedback point time	-	-	-
"	Calibration time	-	-	-
Monomodal Head mouse	Total Task Time	3676	3744	3208
"	Productive Task Time	3012	3301	3004
"	Incorrect Commands Task Time	24	12	4
"	Target Misses Task Time	115	53	24
"	Control Corrections Task Time	444	317	154
"	Feedback point time	-	-	-
"	Calibration time	80	60	20
Monomodal Eye mouse	Total Task Time	4444	4453	2911
"	Productive Task Time	1411	1615	1910
"	Incorrect Commands Task Time	98	59	30
"	Target Misses Task Time	214	157	94
"	Control Corrections Task Time	2324	2258	671
"	Feedback point time	215	260	90
"	Calibration time	179	101	114
Multimodal Head mouse	Total Task Time	3008	2493	2271
"	Productive Task Time	2227	2015	1965
"	Incorrect Commands Task Time	8	10	10
"	Target Misses Task Time	41	40	14
"	Control Corrections Task Time	661	372	246
"	Feedback point time	-	-	-
"	Calibration time	70	53	33
Multimodal Eye mouse	Total Task Time	4637	3423	2273
"	Productive Task Time	1788	1625	1683
"	Incorrect Commands Task Time	23	15	5
"	Target Misses Task Time	203	115	22
"	Control Corrections Task Time	2296	1370	402
"	Feedback point time	172	189	34
"	Calibration time	153	106	126

Table A11.11 Device efficiency by high subject experience and target size

Device	*High experience only Efficiency (%) by Target Size (Visual Angle degrees at 60cm distance)*			
	0.3°	*0.6°*	*0.9°*	*1.2°*
Hand	$66.6^{1,2,3}$	$78.1^{a,n}$	$78.1^{a,m}$	83.3
Head Dwell	33.2^{g}	$59.9^{b,h}$	$62.3^{b,c,j}$	$63.9^{c,l,3}$
Eye Dwell	34.2^{g}	$50.9^{d,i}$	53.6^{d}	65.9^{4}
Head Click	45.8^{g}	$58.9^{e,h}$	$70.1^{e,k,1}$	$75.2^{l,5}$
Eye Click	25.2^{g}	$56.3^{f,h,i}$	$61.3^{f,j,k,2}$	$78.1^{m,n}$

Data within the same device with the same letter or number are not significantly different ($p > 0.050$)

Table A11.12 Statistics of efficiency by high subject experience and target size

Mann-Whitney within-device comparisons				
Device	*Target Size*	*0.3°*	*0.6°*	*0.9°*
Hand	0.6°	$p<0.050$	-	-
	0.9°	$p<0.050$	$p=0.741$	-
	1.2°	$p<0.050$	$p<0.050$	$p<0.050$
Head Dwell	0.6°	$p<0.050$	-	-
	0.9°	$p<0.050$	$p=0.894$	-
	1.2°	$p=0.005$	$p<0.050$	$p=0.071$
Eye Dwell	0.6°	$p<0.050$	-	-
	0.9°	$p<0.050$	$p=0.436$	-
	1.2°	$p<0.050$	$p<0.050$	$p<0.050$
Head Click	0.6°	$p<0.050$	-	-
	0.9°	$p=0.050$	$p=0.298$	-
	1.2°	$p<0.050$	$p<0.050$	$p<0.050$
Eye Click	0.6°	$p<0.050$	-	-
	0.9°	$p<0.050$	$p=0.384$	-
	1.2°	$p<0.050$	$p<0.050$	$p<0.050$

Wilcoxon between-device, within target size comparisons				
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*
	0.3° VA at 60cm			
Head Dwell	$p=0.004$	-	-	-
Eye Dwell	$p<0.050$	$p=0.750$	-	-
Head Click	$p<0.050$	$p=0.885$	$p=0.506$	-
Eye Click	$p<0.050$	$p=1.000$	$p=1.000$	$p=0.673$

Device	0.6° VA at 60cm			
Head dwell	$p < 0.050$	-	-	-
Eye Dwell	$p < 0.050$	$p = 0.025$	-	-
Head Click	$p < 0.050$	*$p = 0.737$*	$p = 0.011$	-
Eye Click	$p < 0.050$	*$p = 0.971$*	*$p = 0.081$*	*$p = 0.889$*
	0.9° VA at 60cm			
Head Dwell	$p < 0.050$	-	-	-
Eye Dwell	$p < 0.050$	$p = 0.035$	-	-
Head Click	$p < 0.050$	$p = 0.024$	$p < 0.050$	-
Eye Click	$p < 0.050$	*$p = 0.198$*	$p = 0.013$	*$p = 0.741$*
	1.2° VA at 60cm			
Head Dwell	$p < 0.050$	-	-	-
Eye Dwell	$p < 0.050$	$p < 0.050$	-	-
Head Click	$p < 0.050$	*$p = 0.325$*	$p < 0.050$	-
Eye Click	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$

Mann-Whitney Hand mouse baseline to device and target size comparisons				
Device	*Hand mouse 0.3° VA at 60cm*			
	0.3°	0.6°	0.9°	1.2°
Head Dwell	$p < 0.050$	$p < 0.050$	$p < 0.050$	*$p = 0.057$*
Eye Dwell	$p < 0.050$	$p < 0.050$	$p < 0.050$	*$p = 0.108$*
Head Click	$p < 0.050$	$p < 0.050$	*$p = 0.151$*	*$p = 0.327$*
Eye Click	$p < 0.050$	$p < 0.050$	*$p = 0.160$*	$p < 0.050$
	Hand mouse 0.6° VA at 60cm			
Head dwell	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$
Eye Dwell	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$
Head Click	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$
Eye Click	$p < 0.050$	$p < 0.050$	$p < 0.050$	*$p = 0.476$*
	Hand mouse 0.9° VA at 60cm			
Head Dwell	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$
Eye Dwell	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$
Head Click	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$
Eye Click	$p < 0.050$	$p < 0.050$	$p < 0.050$	*$p = 0.484$*
	Hand mouse 1.2° VA at 60cm			
Head Dwell	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$
Eye Dwell	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$
Head Click	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$
Eye Click	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$

Appendix Chapter 13

Figure A13.1 Consent form

Usability of Enhanced Head and Eye Mice on a Standard GUI

DE MONTFORT UNIVERSITY SCHOOL OF COMPUTING SCIENCE EXPERIMENTAL SUBJECT CONSENT FORM:

This experiment is designed to study the usability of hand, head and eye mice when operating on a standard graphical user interface.

The experimental procedure will involve:

- Calibrating and using a head mouse based on an electromagnetic head tracker
- Calibrating and using an eye-mouse based on an infrared eye-tracker.
- Completing a practice period lasting 5-10 minutes.
- Completing four sets of tests.
- Controlling a 'zoom' feature with your hand.
- Completing a series of short questionnaires.
- Sitting still in front of the eye-tracker for several periods of approximately 10 minutes.
- The tests will last approximately 30 minutes.
- Short rest periods of 20 minutes will take place between test sessions.

The experiment will expose the subject to electromagnetic fields and infrared light shone at the eyes. This is within safe exposure limits to the knowledge of the experimenter and the manufacturer of the head and eye tracking equipment and to their knowledge should not damage the body or eyes.

You may:
At any reasonable time you may discontinue the tests for a break. Subjects may withdraw from the experiment at any time and will suffer no personal consequences as a result.

Your rights:
All data will remain anonymous and names will not be disclosed without the express written consent of the test subject.

The experiments are to be conducted by the experimenter ______________________________ in the HCI laboratory in the School of Computing Science at De Montfort University.

CONSENT

I ____ agree to participate in the above titled experiment.
I am satisfied that I have received a full explanation of experimental procedures, their possible complications and side effects and have had any queries answered to the extent of the experimenters current knowledge.
I understand that I may terminate the experiment at any time and will suffer no personal consequences as a result.

Subject: ______________________________ Date: __________ Witness: __________________________

Table A13.1 Test marking sheet

Task element	*Task description*	*Task time (ms)*	*Incorrect commands (# ms)*	*Target Misses (# ms)*	*Control corrections at target (# ms)*	*Zoom level (1-16)*	*Zoom time (mS)*	*Zoom immediate? (True / False)*	*Zoom held? (True / false)*	*Zoom level corrections (count)*	*Notes*
1	Click the [Start] button on the task bar										
2	Open the [Programs] menu by clicking the [Programs] icon on the										
3	Start Word by clicking the [Microsoft Word] icon from the start menu										
4	Click the [WiViK soft keyboard] button on the task bar										
5	Resize Word by double clicking the window title bar										
6	Move the Word window to the top left of the screen by dragging the										
7	Resize the Word window to fill the top 2/3 of the screen by										
8	Click the [File] menu										
9	Click the [Open] menu item										
10	Double click the [Test File.doc] filename in the list box										
11	Type [Shift]										
12	Type 'T'										

Table A13.2 Enhanced device assessment conditions

Test condition	*Device*
1	Head Dwell Zoom
2	Eye Dwell Zoom
3	Head Click Zoom
4	Eye Click Zoom

Table A13.3 Enhanced device assessment order, incomplete Latin-square design

Subject	*Test conditions*			
1	1	4	2	3
2	2	1	3	4
3	3	2	4	1
4	4	3	1	2
5	1	4	2	3
6	2	1	3	4

Table A13.4 Enhanced device efficiency

Device	*Metric*	*Percentiles*			*Inter-quartile distance*
		50th	*25th*	*75th*	
Head Dwell Zoom	Overall Efficiency	58.9	44.1	67.0	22.9
Eye Dwell Zoom	"	57.8	34.5	66.7	32.2
Head Click Zoom	"	67.4	52.1	77.1	25.0
Eye Click Zoom	"	69.9	49.1	79.9	30.8
Head Dwell Zoom	Word Domain Efficiency	59.0	44.8	66.9	22.1
Eye Dwell Zoom	"	56.0	35.8	66.2	30.4
Head Click Zoom	"	68.7	53.1	75.5	22.4
Eye Click Zoom	"	70.8	52.4	81.2	28.8
Head Dwell Zoom	Browser Domain Efficiency	55.7	44.0	67.0	23.0
Eye Dwell Zoom	"	60.9	34.0	68.0	34.0
Head Click Zoom	"	67.4	51.9	77.1	25.2
Eye Click Zoom	"	69.2	44.7	79.0	34.3

Table A13.5 Statistics of enhanced device task domain efficiency

Device	*Domain*	*Efficiency (%)*	*Mann-Whitney between-domain comparisons*
Head Dwell Zoom	Word Browser	59.0 [a] 55.7 [a]	$p = 0.643$
Eye Dwell Zoom	Word Browser	56.0 60.9	$p < 0.050$
Head Click Zoom	Word Browser	68.7 [b] 67.4 [b]	$p = 0.463$
Eye Click Zoom	Word Browser	70.8 69.2	$p < 0.050$

Data with the same letter are not significantly different ($p > 0.050$)

Table A13.6 Statistics of overall enhanced device efficiency

Device	*Efficiency (%)*
Hand	83.3
Head Dwell	56.4 [a]
Eye Dwell	42.9
Head Click	65.2
Eye Click	51.1 [b]
Head Dwell Zoom	58.9 [a]
Eye Dwell Zoom	57.8 [b]
Head Click Zoom	67.4 [c]
Eye Click Zoom	69.9 [c]

Data with the same letter are not significantly different ($p > 0.050$)

Wilcoxon between-device comparisons								
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*	*Eye Click*	*Head Dwell Zoom*	*Eye Dwell Zoom*	*Head Click Zoom*
Head Dwell	$p<0.050$	-	-	-	-	-	-	-
Eye Dwell	$p<0.050$	$p<0.050$	-	-	-	-	-	-
Head Click	$p<0.050$	$p<0.050$	$p<0.050$	-	-	-	-	-
Eye Click	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	-	-	-	-
Head Dwell Zoom	$p<0.050$	$p=0.568$	$p<0.050$	$p<0.050$	$p<0.050$	-	-	-
Eye Dwell Zoom	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p=0.316$	$p=0.180$	-	-
Head Click Zoom	$p<0.050$	$p<0.050$	$p<0.050$	$p=0.050$	$p<0.050$	$p<0.050$	$p<0.050$	-
Eye Click Zoom	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p=0.145$

Table A13.7 Enhanced device task time

Device	*Metric*	*Time per Task (ms)*
Head Dwell Zoom	Total Task Time	3480
Eye Dwell Zoom	"	3646
Head Click Zoom	"	2900
Eye Click Zoom	"	2225
Head Dwell Zoom	Productive Task Time	3178
Eye Dwell Zoom	"	2574
Head Click Zoom	"	2670
Eye Click Zoom	"	1546
Head Dwell Zoom	Incorrect Commands Task Time	6
Eye Dwell Zoom	"	34
Head Click Zoom	"	3
Eye Click Zoom	"	5
Head Dwell Zoom	Target Misses Task Time	41
Eye Dwell Zoom	"	130
Head Click Zoom	"	30
Eye Click Zoom	"	30
Head Dwell Zoom	Control Corrections Task Time	220
Eye Dwell Zoom	"	581
Head Click Zoom	"	152
Eye Click Zoom	"	282
Head Dwell Zoom	Zoom Task Time	35
Eye Dwell Zoom	"	214
Head Click Zoom	"	44
Eye Click Zoom	"	182

Table A13.8 Statistics of enhanced device task time

Device	*Task time (ms)*
Hand	1246
Head Dwell	3489 [a b]
Eye Dwell	3668 [a c]
Head Click	2537 [d]
Eye Click	3289 [e]
Head Dwell Zoom	3480 [b c]
Eye Dwell Zoom	3646
Head Click Zoom	2900 [d e]
Eye Click Zoom	2225

Data with the same letter are not significantly different ($p > 0.050$)

Wilcoxon between-device comparisons								
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*	*Eye Click*	*Head Dwell Zoom*	*Eye Dwell Zoom*	*Head Click Zoom*
Head Dwell	$p<0.050$	-	-	-	-	-	-	-
Eye Dwell	$p<0.050$	$p=0.233$	-	-	-	-	-	-
Head Click	$p<0.050$	$p<0.050$	$p<0.050$	-	-	-	-	-
Eye Click	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	-	-	-	-
Head Dwell Zoom	$p<0.050$	$p=0.459$	$p=0.114$	$p<0.050$	$p<0.050$	-	-	-
Eye Dwell Zoom	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	-	-
Head Click Zoom	$p<0.050$	$p<0.050$	$p<0.050$	$p=0.121$	$p=0.153$	$p<0.050$	$p<0.050$	-
Eye Click Zoom	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$

Table A13.9 Enhanced device task quality

Device	*Metric*	*Quality*
Head Dwell Zoom	Overall quality (1-5)	4.38
Eye Dwell Zoom	"	4.10
Head Click Zoom	"	4.57
Eye Click Zoom	"	4.47
Head Dwell Zoom	Incorrect Commands Count/Task	0.012
Eye Dwell Zoom	"	0.0506
Head Click Zoom	"	0.004
Eye Click Zoom	"	0.006
Head Dwell Zoom	Target Misses Count/Task	0.093
Eye Dwell Zoom	"	0.184
Head Click Zoom	"	0.0508
Eye Click Zoom	"	0.064
Head Dwell Zoom	Control Corrections Count/Task	0.410
Eye Dwell Zoom	"	0.521
Head Click Zoom	"	0.307
Eye Click Zoom	"	0.402
Head Dwell Zoom	Zoom Level Corrections Count/Task	0.000
Eye Dwell Zoom	"	0.016
Head Click Zoom	"	0.005
Eye Click Zoom	"	0.003

Table A13.10 Statistics of enhanced device task quality

Device	*Task quality (1-5)*
Hand	4.90
Head Dwell	4.26 [a b c]
Eye Dwell	3.25 [e]
Head Click	4.23 [a d]
Eye Click	3.42 [e]
Head Dwell Zoom	4.38 [b f]
Eye Dwell Zoom	4.10 [c d]
Head Click Zoom	4.57 [g]
Eye Click Zoom	4.47 [f g]

Data with the same letter are not significantly different ($p > 0.050$)

Wilcoxon between-device comparisons								
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*	*Eye Click*	*Head Dwell Zoom*	*Eye Dwell Zoom*	*Head Click Zoom*
Head Dwell	$p<0.050$	-	-	-	-	-	-	-
Eye Dwell	$p<0.050$	$p<0.050$	-	-	-	-	-	-
Head Click	$p<0.050$	*p=0.194*	$p<0.050$	-	-	-	-	-
Eye Click	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	-	-	-	-
Head Dwell Zoom	$p<0.050$	*p=0.060*	$p<0.050$	$p<0.050$	$p<0.050$	-	-	-
Eye Dwell Zoom	$p<0.050$	*p=0.660*	$p<0.050$	*p=0.507*	$p<0.050$	$p<0.050$	-	-
Head Click Zoom	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	-
Eye Click Zoom	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$

Table A13.11 Statistics of enhanced device satisfaction

Device	*Satisfaction (1-7)*	*Device rank*
Hand	6.20	1
Head Dwell	4.36 [a]	6
Eye Dwell	2.93 [b]	9
Head Click	4.73 [d]	2
Eye Click	3.90 [a b c d]	7
Head Dwell Zoom	4.70 [a b d]	4
Eye Dwell Zoom	3.47 [c]	8
Head Click Zoom	4.83 [b d]	2
Eye Click Zoom	4.50 [c d]	5

Data with the same letter are not significantly different ($p > 0.050$)

Wilcoxon between-device comparisons								
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*	*Eye Click*	*Head Dwell Zoom*	*Eye Dwell Zoom*	*Head Click Zoom*
Head Dwell	$p<0.050$	-	-	-	-	-	-	-
Eye Dwell	$p<0.050$	$p<0.050$	-	-	-	-	-	-
Head Click	$p<0.050$	$p<0.050$	$p<0.050$	-	-	-	-	-
Eye Click	$p<0.050$	$p=0.521$	$p=0.148$	$p=0.093$	-	-	-	-
Head Dwell Zoom	$p<0.050$	$p=0.521$	$p=0.149$	$p=0.379$	$p=0.262$	-	-	-
Eye Dwell Zoom	$p<0.050$	$p=0.471$	$p<0.050$	$p<0.050$	$p=0.109$	$p<0.050$	-	-
Head Click Zoom	$p<0.050$	$p<0.050$	$p=0.873$	$p=0.749$	$p<0.050$	$p=0.378$	$p<0.050$	-
Eye Click Zoom	$p<0.050$	$p<0.050$	$p<0.050$	$p=0.471$	$p=0.521$	$p=0.936$	$p=0.065$	$p=0.471$

Table A13.12 Statistics of satisfaction categories - Workload

Device	*Workload (1-7)*
Hand	1.80
Head Dwell	4.10 [a]
Eye Dwell	5.80 [b]
Head Click	3.80 [c]
Eye Click	5.10 [b]
Head Dwell Zoom	3.80 [c]
Eye Dwell Zoom	5.60 [b]
Head Click Zoom	3.70 [c]
Eye Click Zoom	4.40 [a]

Data with the same letter are not significantly different ($p > 0.050$)

Wilcoxon between-device comparisons								
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*	*Eye Click*	*Head Dwell Zoom*	*Eye Dwell Zoom*	*Head Click Zoom*
Head Dwell	$p<0.050$	-	-	-	-	-	-	-
Eye Dwell	$p<0.050$	$p<0.050$	-	-	-	-	-	-
Head Click	$p<0.050$	$p<0.050$	$p<0.050$	-	-	-	-	-
Eye Click	$p<0.050$	$p<0.050$	$p=0.295$	$p<0.050$	-	-	-	-
Head Dwell Zoom	$p<0.050$	$p<0.050$	$p<0.050$	$p=0.753$	$p<0.050$	-	-	-
Eye Dwell Zoom	$p<0.050$	$p<0.050$	$p=0.893$	$p<0.050$	$p=0.208$	$p<0.050$	-	-
Head Click Zoom	$p<0.050$	$p<0.050$	$p<0.050$	$p=1.000$	$p<0.050$	$p=0.787$	$p<0.050$	-
Eye Click Zoom	$p<0.050$	$p=1.000$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$

Table A13.13 Statistics of satisfaction categories - Comfort

Device	*Comfort (1-7)*
Hand	6.50 [a]
Head Dwell	5.90 [b]
Eye Dwell	4.50 [c]
Head Click	6.30 [a]
Eye Click	5.20 [c]
Head Dwell Zoom	5.90 [b]
Eye Dwell Zoom	4.70 [c]
Head Click Zoom	6.10 [b]
Eye Click Zoom	5.20 [c]

Data with the same letter are not significantly different ($p > 0.050$)

Wilcoxon between-device comparisons								
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*	*Eye Click*	*Head Dwell Zoom*	*Eye Dwell Zoom*	*Head Click Zoom*
Head Dwell	$p<0.050$	-	-	-	-	-	-	-
Eye Dwell	$p<0.050$	$p<0.050$	-	-	-	-	-	-
Head Click	$p=0.834$	$p=0.281$	$p<0.050$	-	-	-	-	-
Eye Click	$p<0.050$	$p<0.050$	$p=0.281$	$p<0.050$	-	-	-	-
Head Dwell Zoom	$p<0.050$	$p=0.590$	$p<0.050$	$p=0.100$	$p<0.050$	-	-	-
Eye Dwell Zoom	$p<0.050$	$p<0.050$	$p=0.402$	$p<0.050$	$p<0.050$	$p<0.050$	-	-
Head Click Zoom	$p<0.050$	$p=0.273$	$p<0.050$	$p=0.855$	$p=0.208$	$p<0.050$	$p<0.050$	-
Eye Click Zoom	$p<0.050$	$p<0.050$	$p=0.675$	$p<0.050$	$p=1.000$	$p=0.249$	$p<0.050$	$p<0.050$

Table A13.14 Statistics of satisfaction categories - Ease of Use

Device	*Ease of Use (1-7)*
Hand	6.30
Head Dwell	3.50 [a]
Eye Dwell	2.60
Head Click	4.30 [b c]
Eye Click	4.40 [b c]
Head Dwell Zoom	3.90 [a c d]
Eye Dwell Zoom	3.40 [a c d]
Head Click Zoom	4.60 [b]
Eye Click Zoom	5.10 [b]

Data with the same letter are not significantly different ($p > 0.050$)

Wilcoxon between-device comparisons								
Device	*Hand*	*Head Dwell*	*Eye Dwell*	*Head Click*	*Eye Click*	*Head Dwell Zoom*	*Eye Dwell Zoom*	*Head Click Zoom*
Head Dwell	p <0.050	-	-	-	-	-	-	-
Eye Dwell	p<0.050	p =*0.100*	-	-	-	-	-	-
Head Click	p <0.050	p=*0.059*	p <0.050	-	-	-	-	-
Eye Click	p<0.050	p =*0.093*	p <0.050	p =*0.295*	-	-	-	-
Head Dwell Zoom	p <0.050	p =*0.281*	p <0.050	p =*0.345*	p =*0.675*	-	-	-
Eye Dwell Zoom	p<0.050	p=*1.000*	p <0.050	p =*0.106*	p<0.050	p=*0.208*	-	-
Head Click Zoom	p <0.050	p <0.050	p=*0.059*	p =*1.000*	p =*0.345*	p =*0.093*	p <0.050	-
Eye Click Zoom	p<0.050	p<0.050	p<0.050	p=*0.590*	p=*0.142*	p=*0.093*	p<0.050	p=*0.675*

Appendix Chapter 14

Table A14.1 Enhanced device task efficiency by target size

Device	*Metric*	*Efficiency (%)*
Head Dwell Zoom	Efficiency at Target Size 0.3° at 60cm	40.9
Eye Dwell Zoom	"	40.1
Head Click Zoom	"	44.1
Eye Click Zoom	"	54.1
Head Dwell Zoom	Efficiency at Target Size 0.6° at 60cm	48.4
Eye Dwell Zoom	"	43.9
Head Click Zoom	"	55.6
Eye Click Zoom	"	52.8
Head Dwell Zoom	Efficiency at Target Size 0.9° at 60cm	49.0
Eye Dwell Zoom	"	45.8
Head Click Zoom	"	55.2
Eye Click Zoom	"	57.3
Head Dwell Zoom	Efficiency at Target Size 1.2° at 60cm	58.9
Eye Dwell Zoom	"	53.4
Head Click Zoom	"	68.7
Eye Click Zoom	"	69.0

Table A14.2 Statistics of enhanced device efficiency by target size

Device	Efficiency (%) by Target Size (degrees at 60cm distance)			
	0.3°	*0.6°*	*0.9°*	*1.2°*
Head Dwell Zoom	40.9	48.4 [a]	49.0 [a]	58.9
Eye Dwell Zoom	40.1 [b]	43.9 [bc]	45.8 [c]	53.4
Head Click Zoom	44.1	55.6 [d]	55.2 [d]	68.7
Eye Click Zoom	54.1 [e]	52.8 [e]	57.3	69.0

Data within the same device with the same letter are not significantly different ($p > 0.050$)

Mann-Whitney within-device comparisons				
Device	*Target Size*	*0.3°*	*0.6°*	*0.9°*
Head Dwell Zoom	0.6°	$p<0.050$	-	-
	0.9°	$p<0.050$	*p=0.062*	-
	1.2°	$p<0.050$	$p<0.050$	$p<0.050$
Eye Dwell Zoom	0.6°	*p =0.045*	-	-
	0.9°	$p<0.050$	*p =0.825*	-
	1.2°	$p<0.050$	$p<0.050$	$p<0.050$
Head Click Zoom	0.6°	$p<0.050$	-	-
	0.9°	$p<0.050$	*p =0.784*	-
	1.2°	$p<0.050$	$p<0.050$	$p<0.050$
Eye Click Zoom	0.6°	*p =0.805*	-	-
	0.9°	$p<0.050$	$p<0.050$	-
	1.2°	$p<0.050$	$p<0.050$	$p<0.050$

Table A14.3 Device efficiency by interaction technique

Device	*Interaction Technique Efficiency (%)*				
	-------- Restricted --------			*----- Unrestricted ----*	
	Single	*Double*	*Drag*	*Single*	*Drag*
Hand	80.6 [a]	76.3	64.9 [b]	83.3 [a]	62.5 [b]
Head Dwell	48.2 [c]	48.6 [c]	36.2 [d]	60.3	33.7 [d]
Eye Dwell	47.4 [e]	42.2 [e]	12.1 [f]	45.4	8.5 [f]
Head Click	57.2 [g]	49.5 [g]	30.9 [h]	68.9	39.9 [h]
Eye Click	30.8 [i]	30.4 [i]	27.1 [j]	52.7	17.1 [j]
Head Dwell Zoom	54.1	48.0	33.7	60.5	32.9
Eye Dwell Zoom	59.1 [k]	52.0	28.5	59.8 [k]	36.1
Head Click Zoom	66.3	49.3	34.7 [l]	70.6	39.9 [l]
Eye Click Zoom	51.9 [m]	54.4 [m]	45.3	72.1	50.9 [m]

Data within the same device with the same letter are not significantly different ($p > 0.050$)

Table A14.4 Statistics of efficiency by interaction technique

Mann-Whitney within-device comparisons							
Device	*Interaction Technique*		*-------- Restricted --------*			*---- Unrestricted ----*	
			Single	*Double*	*Drag*	*Single*	*Drag*
Hand	Restrict.	*Single*	-	-	-	-	-
		Double	$p < 0.050$	-	-	-	-
		Drag	$p < 0.050$	$p < 0.050$	-	-	-
	Unrest.	*Single*	$p = 0.080$	$p < 0.050$	$p < 0.050$	-	-
		Drag	$p < 0.050$	$p < 0.050$	$p = 0.411$	$p < 0.050$	-
Head dwell	Restrict.	*Single*	-	-	-	-	-
		Double	$p = 0.515$	-	-	-	-
		Drag	$p < 0.050$	$p = 0.002$	-	-	-
	Unrest.	*Single*	$p < 0.050$	$p = 0.050$	$p < 0.050$	-	-
		Drag	$p < 0.050$	$p < 0.050$	$p = 0.773$	$p < 0.050$	-
Eye dwell	Restrict.	*Single*	-	-	-	-	-
		Double	$p = 0.350$	-	-	-	-
		Drag	$p < 0.050$	$p < 0.050$	-	-	-
	Unrest.	*Single*	$p = 0.370$	$p = 0.514$	$p < 0.050$	-	-
		Drag	$p < 0.050$	$p < 0.050$	$p = 0.757$	$p < 0.050$	-

Head click	Restrict.	*Single*	-	-	-	-	-
		Double	*p =0.528*	-	-	-	-
		Drag	p =0.005	p <0.050	-	-	-
	Unrest.	*Single*	p <0.050	p <0.050	p <0.050	-	-
		Drag	p <0.050	p <0.050	*p =0.183*	p <0.050	-
Eye Click	Restrict.	*Single*	-	-	-	-	-
		Double	*p =0.912*	-	-	-	-
		Drag	*p =0.358*	*p =0.245*	-	-	-
	Unrest.	*Single*	p <0.050	p <0.050	p =0.050	-	-
		Drag	*p =0.060*	*p =0.057*	*p =0.449*	p <0.050	-
Head Dwell Zoom	Restrict.	*Single*	-	-	-	-	-
		Double	p <0.050	-	-	-	-
		Drag	p <0.050	p <0.050	-	-	-
	Unrest.	*Single*	p <0.050	p =0.050	p <0.050	-	-
		Drag	p <0.050	p <0.050	p <0.050	p <0.050	-
Eye Dwell Zoom	Restrict.	*Single*	-	-	-	-	-
		Double	p <0.050	-	-	-	-
		Drag	p <0.050	p <0.050	-	-	-
	Unrest.	*Single*	*p =0.374*	p <0.050	p <0.050	-	-
		Drag	p <0.050	p <0.050	p <0.050	p <0.050	-
Head Click Zoom	Restrict.	*Single*	-	-	-	-	-
		Double	p <0.050	-	-	-	-
		Drag	p <0.050	p <0.050	-	-	-
	Unrest.	*Single*	p <0.050	p <0.050	p <0.050	-	-
		Drag	p <0.050	p <0.050	*p =0.536*	p <0.050	-
Eye Click Zoom	Restrict.	*Single*	-	-	-	-	-
		Double	*p =0.384*	-	-	-	-
		Drag	*p =0.358*	p <0.050	-	-	-
	Unrest.	*Single*	p <0.050	p <0.050	p =0.050	-	-
		Drag	*p =0.388*	p =0.034	p <0.050	p <0.050	-

Table A14.5 All devices efficiency by subject experience

Device	*Efficiency (%) by subject experience (Low, Medium, High)*		
	L	*M*	*H*
Hand	-	-	83.3
Head Dwell	51.6	54.5	63.9[a]
Eye Dwell	19.7[c]	31.8	61.1[a]
Head Click	55.0[e]	66.9	73.0[b]
Eye Click	21.2[c]	48.4	73.5[b]
Head Dwell Zoom	49.4[f]	59.7	62.0[d]
Eye Dwell Zoom	44.1	57.1	66.2[d]
Head Click Zoom	55.0[e]	70.2	75.3
Eye Click Zoom	48.6[f]	74.1	78.9

Data within the same device with the same letter are not significantly different ($p > 0.050$)

Table A14.6 Statistics of enhanced device efficiency by subject experience

Wilcoxon within-device comparisons			
Device	*Experience*	*H*	*M*
Head Dwell	M	$p < 0.050$	-
	L	$p < 0.050$	$p < 0.050$
Eye Dwell	M	$p < 0.050$	-
	L	$p < 0.050$	$p < 0.050$
Head Click	M	$p < 0.050$	-
	L	$p < 0.050$	$p < 0.050$
Eye Click	M	$p < 0.050$	-
	L	$p < 0.050$	$p < 0.050$
Head Dwell Zoom	M	$p < 0.050$	-
	L	$p < 0.050$	$p < 0.050$
Eye Dwell Zoom	M	$p < 0.050$	-
	L	$p < 0.050$	$p < 0.050$
Head Click Zoom	M	$p < 0.050$	-
	L	$p < 0.050$	$p < 0.050$
Eye Click Zoom	M	$p < 0.050$	-
	L	$p < 0.050$	$p < 0.050$

Table A14.6 Statistics of enhanced device efficiency by subject experience continued...

Device	Hand	Head Dwell	Eye Dwell	Head Click	Eye Click	Head Dwell Zoom	Eye Dwell Zoom	Head Click Zoom
Mann-Whitney between-device, within experience comparisons								
	High Experience							
Head Dwell	$p<0.050$	-	-	-	-	-	-	-
Eye Dwell	$p<0.050$	$p=0.142$	-	-	-	-	-	-
Head Click	$p<0.050$	$p<0.050$	$p<0.050$	-	-	-	-	-
Eye Click	$p<0.050$	$p<0.050$	$p<0.050$	$p=0.215$	-	-	-	-
Head Dwell Zoom	$p<0.050$	$p=0.188$	$p<0.050$	$p<0.050$	$p<0.050$	-	-	-
Eye Dwell Zoom	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p=0.023$	-	-
Head Click Zoom	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	-
Eye Click Zoom	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p=0.004$
	Medium Experience							
Head Dwell	-	-	-	-	-	-	-	-
Eye Dwell	-	$p<0.050$	-	-	-	-	-	-
Head Click	-	$p<0.050$	$p<0.050$	-	-	-	-	-
Eye Click	-	$p<0.050$	$p<0.050$	$p<0.050$	-	-	-	-
Head Dwell Zoom	-	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	-	-	-
Eye Dwell Zoom	-	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	-	-
Head Click Zoom	-	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	-
Eye Click Zoom	-	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$
	Low Experience							
Head Dwell	-	-	-	-	-	-	-	-
Eye Dwell	-	$p<0.050$	-	-	-	-	-	-
Head Click	-	$p=0.050$	$p<0.050$	-	-	-	-	-
Eye Click	-	$p<0.050$	$p=0.445$	$p<0.050$	-	-	-	-
Head Dwell Zoom	-	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	-	-	-
Eye Dwell Zoom	-	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	-	-
Head Click Zoom	-	$p<0.050$	$p<0.050$	$p=0.635$	$p<0.050$	$p<0.050$	$p<0.050$	-
Eye Click Zoom	-	$p<0.050$	$p<0.050$	$p<0.050$	$p<0.050$	$p=0.322$	$p<0.050$	$p<0.050$

Table A14.7 All device zoom level by original target size

Device	*Original Target Size*	*Zoom level (Visual Angle at 60cm)*	
		Mean	*SD*
Head Dwell Zoom	0.3°	0.95	0.65
Eye Dwell Zoom	0.6°	0.88	0.32
Head Click Zoom	0.9°	0.97	0.24
Eye Click Zoom	1.2°	1.20	0.050
Head Dwell Zoom	0.3°	1.56	0.75
Eye Dwell Zoom	0.6°	1.81	1.26
Head Click Zoom	0.9°	1.88	0.98
Eye Click Zoom	1.2°	1.81	0.70
Head Dwell Zoom	0.3°	0.73	0.45
Eye Dwell Zoom	0.6°	0.88	0.41
Head Click Zoom	0.9°	1.02	0.31
Eye Click Zoom	1.2°	1.22	0.16
Head Dwell Zoom	0.3°	1.61	0.75
Eye Dwell Zoom	0.6°	1.67	0.85
Head Click Zoom	0.9°	1.82	1.02
Eye Click Zoom	1.2°	1.72	0.62

Table A14.8 All device efficiency by high subject experience and target size

Device	*High experience only Efficiency (%) by Target Size (Visual Angle degrees at 60cm distance)*			
	0.3°	*0.6°*	*0.9°*	*1.2°*
Hand	66.6 [1, 2, 3]	78.1 [a, n]	78.1 [a, m]	83.3
Head Dwell	33.2 [g]	59.9 [b, h]	62.3 [b, c, j]	63.9 [c, l, 3]
Eye Dwell	34.2 [g]	50.9 [d, i]	53.6 [d]	65.9 [4]
Head Click	45.8 [g]	58.9 [e, h]	70.1 [e, k, l]	75.2 [l, 5]
Eye Click	25.2 [g]	56.3 [f, h, i]	61.3 [f, j, k, 2]	78.1 [m, n]
Head Dwell Zoom	47.9	58.4	62.0	64.4
Eye Dwell Zoom	56.9	62.5	62.9	67.1
Head Click Zoom	60.7	66.3	71.0	77.7
Eye Click Zoom	64.7	69.1	74.2	81.5

Data within the same device with the same letter or number are not significantly different ($p > 0.050$)

Table A14.9 Statistics of efficiency by high subject experience and target size

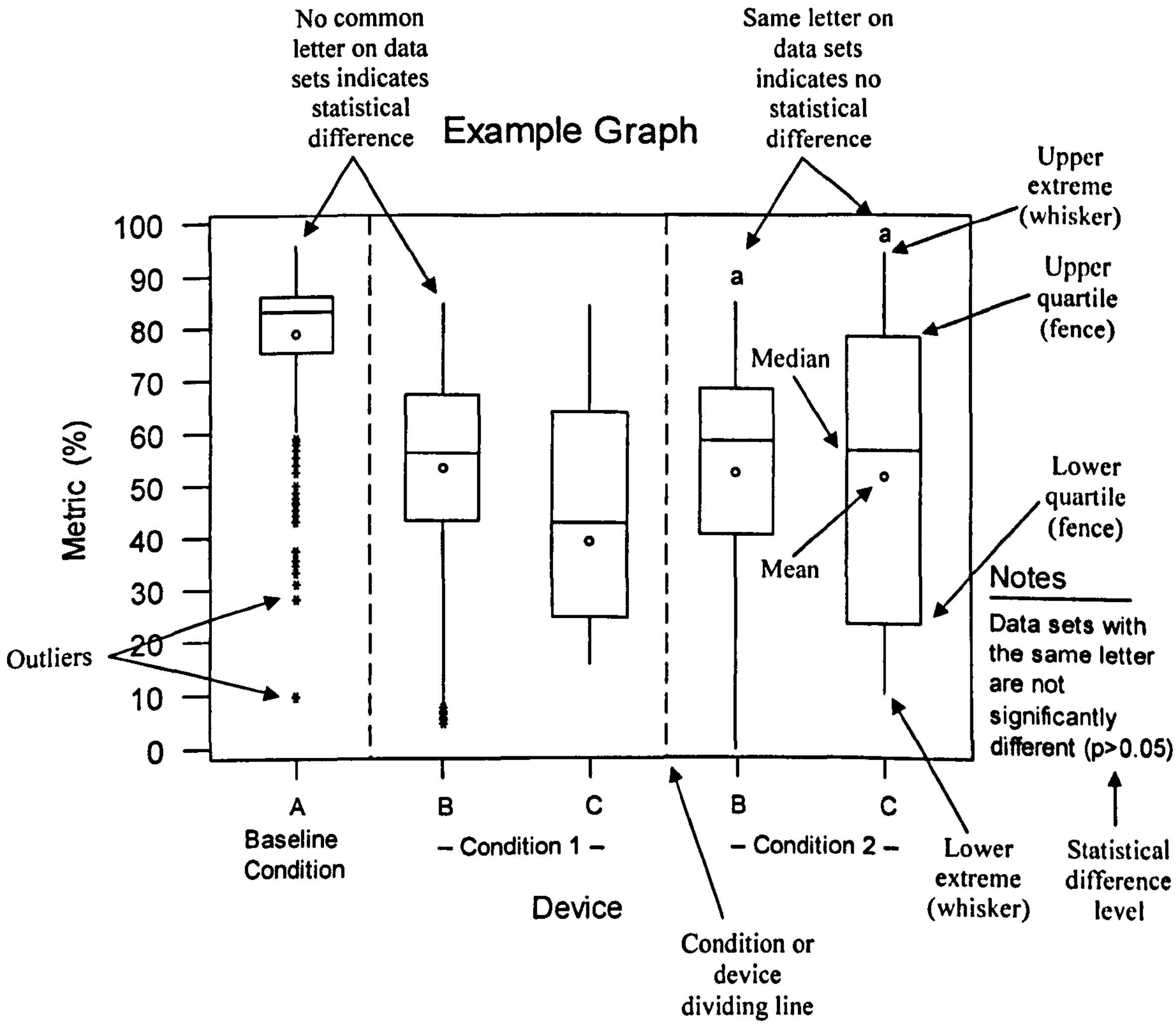
Mann-Whitney within-device comparisons				
Device	*Target Size*	*0.3°*	*0.6°*	*0.9°*
Hand	0.6°	$p = 0.742$	-	-
	0.9°	$p < 0.050$	$p < 0.050$	-
	1.2°	$p < 0.050$	$p < 0.050$	$p < 0.050$
Head Click Zoom	0.6°	$p < 0.050$	-	-
	0.9°	$p < 0.050$	$p < 0.050$	-
	1.2°	$p < 0.050$	$p < 0.050$	$p < 0.050$
Eye Click Zoom	0.6°	$p < 0.050$	-	-
	0.9°	$p < 0.050$	$p < 0.050$	-
	1.2°	$p < 0.050$	$p < 0.050$	$p < 0.050$

Mann-Whitney between-device, within target size comparisons		
Device	*Hand*	*Head Click Zoom*
	Target Size 0.3°	
Head Click Zoom	$p < 0.050$	-
Eye Click Zoom	$p = 0.575$	$p < 0.050$
	Target Size 0.6°	
Head Click Zoom	$p < 0.050$	-
Eye Click Zoom	$p < 0.050$	$p < 0.050$
	Target Size 0.9°	
Head Click Zoom	$p < 0.050$	-
Eye Click Zoom	$p < 0.050$	$p < 0.050$
	Target Size 1.2°	
Head Click Zoom	$p < 0.050$	-
Eye Click Zoom	$p < 0.050$	$p < 0.050$

Mann-Whitney Hand mouse baseline to device and target size comparisons				
Device	***Hand mouse 0.3° VA at 60cm***			
	0.3°	0.6°	0.9°	1.2°
Head Click Zoom	$p < 0.050$	$p = 0.712$	$p < 0.050$	$p < 0.050$
Eye Click Zoom	$p = 0.575$	$p < 0.050$	$p < 0.050$	$p < 0.050$
	Hand mouse 0.6° VA at 60cm			
Head Click Zoom	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p = 0.396$
Eye Click Zoom	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$
	Hand mouse 0.9° VA at 60cm			
Head Click Zoom	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p = 0.174$
Eye Click Zoom	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$
	Hand mouse 1.2° VA at 60cm			
Head Click Zoom	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$
Eye Click Zoom	$p < 0.050$	$p < 0.050$	$p < 0.050$	$p < 0.050$

Appendix Notes

Notes Figure N1. Explanation of graph symbols



Notes:
- Lower quartile to median = 25th to 50th percentile = range of upper 25% of data from median.
- Median to upper quartile = 50th to 75th percentile = lower 25% range of data from median.
- Range between lower and upper quartiles = interquartile range (IQR) = range of 50% of data.
- Whiskers = indicate full range of data, not including outliers, extend up to 1.5 times IQR.
- Outlier = data beyond 1.5 times IQR, indicate values greater than 1.5 times IQR from median.

Notes Discussion N1. Non-parametric tests in this work

The t-test and other similar parametric tests

The t-test is probably the most widely used statistical test of all time, and certainly the most widely known. It is simple, straightforward, easy to use, and adaptable to a broad range of situations. The t-test, and other similar parametric tests, are the basic statistical tests for comparing two sets of data.

For a t-test:

1. The two samples are independently and randomly drawn from the source population(s).

2. The scale of measurement for both samples has the properties of an equal interval scale.

3. The source population(s) can be reasonably supposed to have a normal distribution.

Why a t-test is not suitable:

In this work, many of the measurements are either of task time or a quantity based on time. Time measurement during tasks tends not to have a normal distribution as zero time is not possible, and tasks tend to have an upper time limit before being failed. Thus a normal distribution is unlikely.

In addition, in this work many of the measurements are either of task quality or a quantity based on task quality. Here this is measured on a scale. A rating scale cannot be assumed to possess the properties of an equal interval scale hence a normal distribution cannot be achieved.

Thus the data in this work cannot fulfil conditions 2) and 3) for the t-test, so a t-test and similar parametric tests cannot be used and alternative non-parametric tests must be found.

The Mann-Whitney two-sample rank test

The Mann-Whitney two-sample rank test is a non-parametric equivalent to a two-sample t-test based on the ranks of the data. It is particularly useful for data with non-normal distributions where data sets have different sample numbers and the *data do not have something in common* – such as they come from different test domains, and each data sample from one device *cannot be paired* with a corresponding sample from the another device

The only assumptions of the Mann-Whitney test are:

1. That the two samples are randomly and independently drawn.

2. That the dependent variable (e.g., efficiency) is intrinsically continuous, capable in principle, if not in practice, of producing measures carried out to the nth decimal place.

3. That the measures within the two samples have the properties of at least an ordinal scale of measurement, so that it is meaningful to speak of "greater than," "less than," and "equal to."

The Wilcoxon matched-pairs signed rank test

The Wilcoxon matched pairs signed rank non-parametric test is used when distributions are not normal, sample sizes are equal and both data sets have commonality – for example *both sets of data originate from the same set of test tasks* and each data sample from one device *can be paired* with a corresponding sample from the another device.

The only assumptions of the Wilcoxon matched-pairs signed rank test are:

1. That the paired values of A and B are randomly and independently drawn (i.e., each pair is drawn independently of all other pairs).

2. That the dependent variable (e.g., efficiency) is intrinsically continuous, capable in principle, if not in practice, of producing measures carried out to the nth decimal place.

3. That the measures of A and B have the properties of at least an ordinal scale of measurement, so that it is meaningful to speak of "greater than," "less than," and "equal to."

Notes Discussion N2. Multiple comparisons in this work

Multiple tests

In comparisons of multiple permutations, p-values (alpha) may require correction for the familywise error rate. For example, suppose that we have four groups and we want to carry out all pairwise comparisons of the group means. There are six such comparisons: 1 with 2, 1 with 3, 1 with 4, 2 with 3, 2 with 4 and 3 with 4. Such set of comparisons is called a family. If we use, for example, a t-test to compare each pair at a certain significance level alpha, then the probability of Type I error (incorrect rejection of the null hypothesis of equality of means) can be guaranteed not to exceed alpha only individually, for each pairwise comparison separately, but not for the whole family. To ensure that the probability of incorrectly rejecting the null hypothesis for any of the pairwise comparisons in the family does not exceed alpha, multiple comparisons methods that control the familywise error rate (FEW) may need to be used.

A Bonferroni multiple comparison test is often used for alpha correction. This is regarded as a conservative test, that is, the FWE is not exactly equal to alpha, but is less than alpha in most situations. Even though the Bonferroni test controls the FEW rate, in many situations it may be too conservative and not have enough power to detect significant differences.

Another area where the Bonferroni correction becomes useful is with comparisons across multiple groups of subjects. If you have four treatment groups (e.g., A, B, C, and D), then there are six possible pairwise comparisons (n=6?) among these groups (A vs B, A vs C, A vs D, B vs C, B vs D, C vs D). If interested in all possible pairwise comparisons, the Bonferroni correction provides a simple way to ensure that making these comparisons does not lead to some of the same problems as testing multiple outcome measures.

Some scientists dislike the use of the Bonferroni correction; they prefer instead that researchers clearly label any results from a fishing expedition as preliminary and/or exploratory. Furthermore, the Bonferroni correction can cause a substantial loss in the precision of research findings.

The tests in this work are not aimed as being multiple comparisons, but are rather comparisons either between devices or between conditions. However, this results in the tables of comparisons shown in this work. The reader is advised that these comparisons are not corrected for alpha in the cases where multiple comparisons may be felt occur, as this would result in a substantial loss of precision in the findings, and the reader is advised that the findings are exploratory.

References

Abrams, R. A., Meyer, D. E., & Kornblum, S. (1989). The speed and accuracy of saccadic eye movements: Characteristics of impulse variability in the oculomotor system. Journal of Experimental Psychology: Human Perception and Performance, 15, 529-543.

Accot, J., and Zhai, S. (1997) Beyond Fitts Law: Models for Trajectory-Based HCI Tasks, CHI 1997.

Albinsson, P. A. and Zhai, S. (2003) High Precision Touch Screen Interaction. Proceedings of CHI2003 Vol 5, Issue1, pp 105-112, April 5-10 2003, Florida, USA. ACM Press.

Bass, B. M., and Cascio, W. F. (1974) Magnitude estimations of expressions of frequency and amount. *Journal of Applied Psychology* 59(3), 313-320.

Bates, R. (1999) Multimodal Eye-based Interaction for Zoomed Target Selection on a Standard Graphical User Interface Proceedings of Interact'99 vol II p7-8. British Computer Society.

Bates, R. (2000b) Using Soft Devices as Enabling Translators - Standard GUI Interaction For All! Proceedings of Human Computer Interaction 2000 vol II p53-54. British Computer Society.

Bates, R. (2002a) A Computer Input Device Selection Methodology for Users with High-Level Spinal Cord Injuries - Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Trinity Hall, University of Cambridge 25th-27th March 2002.

Bates, R. (2002a) Have patience with your eye mouse! Eye-gaze interaction with computers can work- Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Trinity Hall, University of Cambridge 25th-27th March 2002.

Bates, R. (2002c) Eye-gaze interaction with computers can work - Proceedings of the Scientific meeting of Institute of Physics in Engineering and Medicine (IPEM) September 9 - 12. 2002, Durham.

Bates, R. Istance, H.O. (2002a) Why are eye mice unpopular? A detailed comparison of head and eye controlled assistive technology pointing devices- Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Trinity Hall, University of Cambridge 25th-27th March 2002.

Bates, R. Istance, H.O. (2002b) Zooming interfaces! Enhancing the performance of eye controlled pointing devices - Proceedings of ASSETS 2002, The Fifth

International ACM SIGCAPH Conference on Assistive Technologies, July 8 - 10, 2002, Edinburgh, Scotland.

Bates, R. Istance, H.O. (2003) Why are eye mice unpopular? A detailed comparison of head and eye controlled assistive technology pointing devices - Journal of the Association for the Advancement of Assistive Technology in Europe (AAATE): Universal Access in the Information Society Volume 2, Number 3, October 2003 Pages: 280 - 290 Special Issue on "Countering Design Exclusion", guest-edited by Simeon Keates and John Clarkson, Springer-Verlag Heidelberg ISSN: 1615-5289.

Bates, R. Istance, H.O. (2004) Towards eye based virtual environment interaction for users with high-level motor disabilities. International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT), pages 275-282, 20th-22nd September 2004, New College, Oxford, UK, ISBN 0704911442.

Bates, R., Istance, H.O., Donegan, M., Oosthuizen, L. (2005) Fly Where You Look: Enhancing Gaze Based Interaction in 3D Environments. In Proceedings of HCI International 2005, Volume 7 - Universal Access in HCI: Exploring New Interaction Environments, Caesars Palace, Las Vegas, USA, July 22-27 2005.

Bates, R., and Bierton, R. (2000) Experimental Determination of Quantifiers for Usability Questionnaire Design. Proceedings of Human Computer Interaction 2000 vol II p57-58. British Computer Society.

Betke M, Gips J, Fleming P. The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities. IEEE Trans Neural Syst Rehabil Eng. 2002 Mar;10(1):1-10.

Beukelman, D. R. and Mirenda, P. (1992) Augmentative and Alternative Communication: Management of Severe Communication Disorders in Children and Adults. Brookes Publishing Co., Baltimore, USA.

Bevan, N., Kirakowski, J., and Maissel, J. (1991) What is usability?, in Human Aspects in Computing: Design and Use of Interactive Systems and Work with Terminals, Elsevier Science.

Bevan, N., Rengger, R., Macleod, M., Bowden, R., Drynan, A., Blayney, M., and Cooper, D. (1995) The Performance Measurement Handbook Version 3, National Physical Laboratory, Teddington, Middlesex, United Kingdom, TW11 0LW, NPLUS/PMH/v3.0/Dec 95.

Bierton, R. (1999) The Prediction of Task Domain Performance Using Domain Representative Tests for Eye-Tracking and Mouse Control. MSc Thesis, De Montfort University, UK.

Bierton, R., and Bates, R. (2000) Experimental Determination of Optimal Scales for Usability Questionnaire Design. Proceedings of Human Computer Interaction 2000 vol II p55-56. British Computer Society.

Bolt, R. A. (1981) Gaze-Orchestrated Dynamic Windows, Computer Graphics Vol 15, 3, August 1981, pp 109-119, ACM Press.

Brewster, S. A., Raty, V. P. and Kortekangas, A. (1996) Enhancing scanning input with non-speech sounds, Proceedings of the second annual ACM conference on Assistive technologies, p.10-14, April 11-12, 1996, Vancouver, British Columbia, Canada.

Brewster, S.A. (1994) Providing a structured method for integrating non-speech audio into human-computer interfaces. PhD Thesis, University of York, UK.

Byers, J.C., Bittner, A.C., and Hill, S.G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisions necessary? In A. Mital (Ed.), Advances in industrial ergonomics, pp. 481-485. Taylor & Francis.

Byrne, M. D., Anderson, J. A., Douglass, S., & Matessa, M. (1999). Eye tracking the visual search of click-down menus. In ACM CHI'99 Conference on Human Factors in Computing Systems (pp. 402-409). NewYork: ACM.

Card, S. K., English, W. K., and Burr, B. J. (1978) Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT, Ergonomics 21 601-613.

Card, S. K., Moran, T. P. and Newell, A. (1983) The Psychology of Human-Computer Interaction, Lawrence Erlbaum.

Carpenter, R.H.S. (1991) Eye movements, MacMillan.

Casali, S. P. and Chase, J. D. (1995) The Effects of Physical Attributes of Computer Interface Design on Novice and Experienced Performance of users with Physical Disabilities, International Journal of Industrial Ergonomics 15, 237-245.

Chapman, J. E. (1991) The Use of Eye-Operated Computer System in Locked-In Syndrome. Proceedings of the Sixth Annual International Conference on Technology and Persons with Disabilities (CSUN'91), Los Angeles, CA.

Chen YL, Chen WL, Kuo TS, Lai JS. A head movement image (HMI)-controlled computer mouse for people with disabilities. Disabil Rehabil. 2003 Feb 4;25(3):163-7.

Demasco, PW. McCoy, KF. (1992) Generating Text from Compressed Input: An Intelligent Interface for People with Severe Motor Impairments. Communications of the ACM vol35 no5 May 1992 p69-78 1992, ACM Press.

Douglas, S.A., Kirkpatrick, A.E., and MacKenzie, I.S. (1999) Testing Pointing Device Performance and use Assessment with the ISO9241, Part 9 Standard. Proceedings of CHI'99, ACM Press.

Duchowski, A. (2000) Eye-Based Interaction in Graphical Systems: Theory and Practice. SigGraph 2000, New Orleans, 23 – 28th July 2000. Course Notes, ACM press.

Edwards, A. L. (1957) Techniques of attitude scale construction, Appleton-Century-Crofts, Inc. New York.

Edwards, G. (1998) New Software makes eye tracking viable: You can control computers with your eyes, in Proceedings of CSUN Technology and Persons with Disabilities Conference 1998, California State University Northridge.

Epps, B. W. (1986) Comparison of six cursor control devices based on Fitts' Law models. Proceedings of the 30th Annual Meeting of the Human Factors Society, 327-331, Human Factors Society.

EUROSTAT (2002) http://europa.eu.int/comm/eurostat.

Evans, G., Drew, R., Blenkhorn, P. (2000) Controlling mouse pointer position using an infrared head-operated joystick. IEEE Trans Rehabil Eng. 2000 Mar;8(1):107-17.

Evans, G., and Blenkhorn, P. (1999) A Head Operated Joystick – Experience with Use, in Proceedings of the CSUN Conference on Technology and Persons with Disabilities 1999, California State University Northridge.

Fernstrom, E., and Ericson, M.O. (1997) Computer mouse or Trackpoint - effects on muscular load and operator experience, Applied Ergonomics vol 28 no 5/6 347-354, Elsevier Science Ltd

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

Frey, L. A., White, K. P. and Hutchinson, T. E. (1990) Eye-Gaze Word processing, IEEE Transactions on Systems, man and Cybernetics 20 (4), 944-950.

Frisbie, D. A., and Brandenburg, D. C. (1979) Equivalence of questionnaire items with varying response formats. Journal of Educational Measurement 16(1), 43-48.

Gips, J. DiMattia, P. Curran, FX. Oliveri, P. (1996) Using EagleEyes: An Electrodes based Device for Controlling the Computer with Your Eyes to Help People with Special Needs. In Interdisciplinary Aspects on Computers Helping People with Special Needs ed. Klaus, J. 1996. R. Oldenberg, Vienna.

Gips, J. Olivien, P. Tecce, J. (1993) Direct Control of the Computer Through Electrodes Placed Around the Eyes. In Human-Computer Interaction: Applications and Case Notes, Elsevier.

Glenstrup, A. J. and Engell-Nielsen, T. (1995) Eye controlled media: Present and future state. Degree thesis. www.diku.dk/panic/eyegaze/

Hancock, G. R. and Klockars, A. J. (1991) The effect of scale manipulations on validity: targeting frequency rating scales for anticipated performance levels, *Applied Ergonomics vol.22 no.3* pp147-154.

Hansen, J. P., Hansen, D. W. and Johansen, A. S. (2001) Bringing Gaze-based Interaction Back to Basics. In Universal Access In HCI, C. Stephanidis (ed.), Lawrence Erlbaum Associates. 2001, pp.325-328.

Hansen, J. P., Johansen, A. S., Hansen, D. W., Itoh, K. and Mashino, S. (2003) Language Technology in a Predictive, Restricted On-screen Keyboard with Ambiguous Layout for Severely Disabled People. Paper submitted to the EACL 2003 Workshop on Language Modeling for Text Entry Methods, April 13, 2003, Budapest, Hungary.

Hansen, J. P., Torning, K., Johansen, A. S., Itoh, K. and Aoki, H. (2004) Gaze typing compared with input by head and hand. Eye Tracking Research & Applications Conference (ETRA 2004), ACM Press.

Hart, S. G., and Staveland, L. E. (1988) Development of the NASA-tlx (Task Load Index): Results of empirical and theoretical research, in Human Mental Workload, 139-183, Elsevier.

Hartson, H. R., Siochi, A. C. and Hix, D. (1990) The UAN: A User Oriented Representation for Direct Manipulation Interface Design. ACM Transactions on Information Systems Vol8 No3, 181-203, ACM Press.

Heuvelmans, A. M. F. Melotte, H. E. M. Neve, J. J. (1990) A typewriting system operated by head movements, based on home-computer equipment. In applied ergonomics 21.2 pp115-120, Human Factors Society.

Howarth, P. A., Istance, H. O., Paterson, S. J. and O'Connor, D. N. (1992) WYLAIWYG: Controlling Computers by Eye. Proceedings of the Third International Conference on Visual Search. Nottingham University, August 1992.

Hutchinson, T. E., White, K. P., martin, W. N., Reichert, K. C. and Frey, L. A. (1989) Human-Computer Interaction Using Eye-Gaze Input. IEEE Transactions on Man, Machine and Cybernetics 19 (6), 1527-1534.

International Alliance of Amyotrophic Lateral Sclerosis and Motor Neurone Disease (ALS/MND) Associations (2005), http://www.alsmndalliance.org

ISO (1998) ISO/TC 159/SC4/WG3 N147: *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9 - Requirements for non-keyboard input devices*, International Organisation for Standardisation, May 25, 1998.

Isokoski, P. (2000) Text Input Methods for Eye Trackers Using Off-Screen Targets, in Proceedings of the Eye Tracking Research and Applications Symposium 2000, (ETRA'00), palm Beach Gardens, FL. ACM Press, p15 - 21.

Istance, H.O.I. (1997) Eye-Based Interaction for Users with Special Needs. Interfaces No35 1997, BCS.

Istance, H.O.I. and Howarth, P. (1993) Performance Modelling and Investigation of User Performance using an Eyetracker as an Input Device, Polish-English Meeting on Information Systems, Technical University of Bialystock, Poland

Istance, H.O.I., and Howarth, P.A. (1994) Keeping an Eye on Your Interface: The Potential for Eye-based Control of Graphical User Interfaces (GUIs) Proceedings of HCI'94 People and Computers IX ed. Cockton, G.p195-209, Cambridge University Press

Istance, H.O.I., Spinner, C., and Howarth, P.A. (1996a) Eye-based control of standard GUI software Proceedings of people and Computers XI HCI '96, Springer

Istance, H.O.I., Spinner, C., and Howarth, P.A. (1996b) Providing Motor-Impaired users with Access to Standard Graphical User Interfaces (GUI) Software via Eye-Based Interaction. Proceedings of the 1st European Conference on Disability, Virtual Realityu and Associated Technologies (ECDVRAT'96), Maidenhead, UK.

Jacob, R.J.K. (1990) What You Look At is What You Get: Eye Movement-Based Interaction Techniques, CHI90

Jacob, R.J.K. (1991) The use of eye movements in human-computer interaction techniques: What you look at is what you get, ACM Transactions on Information Systems 9, 3, 152-169.

Jacob, R.J.K. (1995) Eye tracking in advanced interface design, in Advanced Interface Design and Virtual Environments, 258-288, Oxford University Press.

Jagacinski, R. J., and Monk, D. L. (1985). Fitts' law in two dimensions with hand and head movements. Journal of Motor Behavior, *17,* 77-95.

Lankford (2000) Effective eye-gaze input into Windows. In Proceedings of the symposium on Eye tracking research & applications, Palm Beach Gardens, Florida, United States, p23-27, ACM Press.

Leventhal, L. M. (1991) Screen keyboards: an empirical study of the effects of shape and character layout. In Human aspects in computing:design and use of interactive systems and work with terminals (ed) Bullinger, H. J., Elsevier.

LoPresti, E., Brienza, D. M., Angelo, J., and Gilbertson, L. (2000a) Computer Head Controls: Ergonomics and Effect of Neck Movement Limitations, in CSUN's 15th

Annual Conference 'Technology and Persons with Disabilities', California State University Northridge.

LoPresti, E., Brienza, D.M., Angelo, J., Gilbertson, L., and Sakai, J. (2000b) Neck Range of Motion and Use of Computer Head Controls. Proceedings of Assets 2000: The Fourth International ACM Conference on Assistive Technologies. pp. 121-128.

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

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. Human-Computer Interaction, 7, 91-139.

MacKenzie, I.S. (1992) Fitts Law as a Research and Design Tool in Human-Computer Interaction, Human-Computer Interaction 1992 vol7 p99-139, Lawrence Erlbaum Assoc.

MacKenzie, I.S., and Buxton, W. (2000) Extending Fitts Law to Two-Dimensional Tasks, CHI 1992

MacKenzie, I.S., and Oniszczak, A. (1998) A Comparison of Three Selection Techniques for Touchpads In Touch with Interfaces Proceedings of ACM CHI 98 Conference on Human Factors in Computing Systems 1998 v.1 p.336-343

MacKenzie, I.S., Kauppinen, T., and Silfverberg, M. (2001) Accuracy Measures for Evaluating Computer Pointing Devices. Proceedings of CHI2001, ACM Press.

MacKenzie, I.S., Sellen, A., and Buxton. W. (1991) A comparison of input devices in elemental pointing and dragging tasks, CHI 1991.

MacKenzie, S. Zhang, S. Soukoeff, RW. (1999) Text Entry Using Soft Keyboards. In Behaviour and Information Technology 1999 vol18 number4 pp235-244.

Macleod, M., Bowden, R., Bevan, N., and Curson, I. (1997) The MUSiC performance measurement method. Behaviour and Information Technology vol 16, no. 4/5 pages 279-293, Crown.

Majaranta, P., Aula, A. and Räihä, K. J. (2004) Effects of feedback on eye typing with a short dwell time Proceedings of the Eye tracking research & applications symposium on Eye tracking research & applications San Antonio, Texas, 139 - 146.

Majaranta, P., Räihä, K. J. (2002) Twenty years of eye typing: systems and design issues, Proceedings of the symposium on ETRA 2002: eye tracking research & applications symposium, March 25-27, 2002, New Orleans, Louisiana, ACM Press.

Mathy-Laikko, P. West, C. Jones, R. (1993) Development and Assessment of a Rate Acceleration Keyboard for Direct-Selection Augmentative and Alternative Communication Users. In Technology and Disability 1993 2(3) p57-67, Andover Medical.

Miniotas, D. (2000) Application of Fitts' law to eye gaze interaction. In proceedings of the Conference on Human Factors in Computing Systems (CHI '00) extended abstracts on Human factors in computing systems, 339 – 340, ACM Press.

Miniotas, D., Spakov, O., MacKenzie, I. S. (2004). Eye gaze interaction with expanding targets. In Extended Abstracts of the ACM Conference on Human Factors in Computing Systems,Vienna, Austria, 1255-1258, ACM Press.

Murata, A. (1991) An experimental evaluation of mouse trackball joystick joycard lightpen trackball and touchscreen for pointing - basic study on human interface design, Human aspects in computing:design and use of interactive systems and work with terminals (ed) Bullinger, H. J. Elsevier

Nantais, T. Shein, F. Treviranus, J. (1994) A Predictive Selection Technique for Single-Digit Typing with a Visual Keyboard. IEEE Transactions on Rehabilitation Engineering vol2 no3 sept 1994, IEEE Press.

Oppenheim, A. N. (1992) Questionnaire design, interviewing and attitude measurement, Pinter Press, London.

Partala, T., Aula, A. and Surakka, V. (2001) Combined Voluntary Gaze Direction and Facial Muscle Activity as a New Pointing Technique. In M. Hirose (Ed.) Proceedings of the Eighth IFIP TC13 Conference on Human-Computer Interaction (INTERACT 2001), IOS Press.

Phillips, C.H.E., and Apperley, M.D. (1991) Direct Manipulation Interaction Tasks: A Macintosh-based Analysis. Interaction with Computers, **3** (1), 9-26.

Preece, J. (1994) Human-Computer Interaction. Addison-Wesley, p290-302.

Radwin RG, Vanderheiden GC, Lin ML. A method for evaluating head-controlled computer input devices using Fitts' law. Hum Factors. 1990 Aug;32(4):423-38.

Rasmusson, D., Chappell, R. and Trego, M. (1999). Quick Glance: Eye-tracking Access to Windows95 Operating Environment. Proceedings of the Fourteenth International Conference on Technology and Persons with Disabilities (CSUN'99), Los Angeles, CA.

Rayner, K. (1995) Eye Movements and Cognitive Processes in Reading, Visual Search and Scene Perception. In J.M. Findlay, R.W. Kentridge and R. Walker (Eds.), Eye Movement Research: Mechanisms, Processes, and Applications. Amsterdam: North-Holland, pp-3-22.

Saito, S. (1992) Does fatigue exist in a quantitative measurement of eye movements? Ergonomics 35 5/6, 607-615.

Salvucci, D. D. (1999) Inferring Intent in Eye-Based Interfaces: Tracing Eye Movement with Process Models. Human Factors in Computing Systems: CHI 99 Conference Proceedings, ACM Press.

Salvucci, D.D., and Anderson, J.R. (2000) Intelligent gaze-added interfaces, in Proceedings of CHI 2000, 2, ACM Press, 273-280.

Schriesheim, C.A., and Schriesheim, J.F. (1974) Development and empirical verification of new response categories to increase the validity of multiple response alternative questionnaires, *Educational and Psychological Measurement, 34,* pp877-884.

Schriesheim, C.A., and Schriesheim, J.F. (1978) The invariance of anchor points obtained by magnitude estimation and pair-comparison treatment of complete ranks scaling procedures: an empirical study comparison and implications for validity measurement, *Educational and Psychological Measurement, 38,* pp977-983.

Shein, F. Hamann, G. Brownlow, N. Treviranus, J. Milner, M. Parnes, P. (1991) WiViK: A Visual Keyboard for Windows 3.0. RESNA 14th Annual Conference 1991.

Shein, F. Treviranus, J. Hamann, G. Galvin, R. Parnes, P. Milner, M. (1992) New Directions in Visual Keyboards for Graphical User Interfaces. Proceedings of the 7th Annual Conference on Technology and Persons with Disabilities, ACM Press.

Sibert, L.E., and Jacob, R.J.K. (2000) Evaluation of eye gaze interaction, in Proceedings of CHI 2000, 2, ACM Press, 281-288.

Sibert, L.E., Templeman, J.N. and Jacob, R.J.K. (2001) Evaluation and Analysis of Eye Gaze Interaction, NRL Report NRL/FR/5513--01-9990, Naval Research Laboratory, Washington, D.C.

Smith, W. J. (1996) ISO and ANSI Ergonomic Standards for Computer Products: A Guide to Implementation and Compliance. Prentice Hall, New York.

Spector, P. E. (1976) Choosing response categories for summated rating scales, *Journal of Applied Psychology, vol.61,* pp374-375.

Spinal Injuries Association Company Limited (2005), Newpoint House, 76 St James Lane, London.

Sprent, P. (1993) Applied Non-Parametric Statistical Methods, 2nd Edition. Chapman & Hall.

Stampe, D. M. (1993) Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. Behavior Research Methods, Insturments and Computers 1993, 25, 2, 137-142, Psychonomic Society Inc.

Stampe, D.M. Reingold, E.M. (1995) Selection by Looking: A Novel Computer Interface and its Application to Psychological Research. Findlay, J.M. Kentridge, R.W. Walker, R. (Eds.) Eye Movement research: Mechanisms, processes and applications, Elsevier.

Symonds, P. M. (1924) On the loss of reliability in ratings due to coarseness of the scale. Journal of Experimental Psychology 17, 456-461.

Velichkovsky, B., Sprenger, A., and Unema, P. (1997) Towards gaze mediated interaction: Collecting solutions of the "midas touch" problem, in Human-Computer Interaction: INTERACT'97, Chapman & Hall.

Ware, C., and Mikaelian, H.H. (1987) An evaluation of an eye tracker as a device for computer input, in Proceedings of CHI 1987, 183-188, ACM Press.

Yarbus, A.L. (1967) Eye Movements and Vision, Plenum, New York.

Zhai, S., Morimoto, C., and Ihde, S. (1999) Manual and Gaze Input Cascaded (MAGIC) Pointing, in Proceedings of CHI 1999, ACM Press, 246-253.