Gaze and Speech: Pointing Device and Text Entry Modality

T. R. Beelders and P. J. Blignaut

1 Introduction

Communication between humans and computers is considered to be a two-way communication between two powerful processors over a narrow bandwidth (Jacob and Karn 2003). Most interfaces today utilize more bandwidth with computer-to-user communication than vice versa, leading to a decidedly one-sided use of the available bandwidth (Jacob and Karn 2003). An additional communication mode will invariably provide for an improved interface (Jacob 1993) and new input devices which capture data from the user both conveniently and at a high speed are well suited to provide more balance in the bandwidth disparity (Jacob and Karn 2003). In order to better utilize the bandwidth between human and computer, more natural communication which concentrates on parallel rather than sequential communication is required (Jacob 1993). The eye tracker is one possibility which meets the criteria for such an input device. Eye trackers have steadily become more robust, reliable and cheaper and, therefore, present themselves as a suitable tool for this use (Jacob and Karn 2003). However, much research is still needed to determine the most convenient and suitable means of interaction before the eye tracker can be fully incorporated as a meaningful input device (Jacob and Karn 2003).

Furthermore, the user interface is the conduit between the user and the computer and as such plays a vital role in the success or failure of an application. Modern-day interfaces are entirely graphical and require users to visually acquire and manually manipulate objects on screen (Hatfield and Jenkins 1997) and the current trend of Windows, Icons, Menu and Pointer (WIMP) interfaces have been around since the 1970s (Van Dam 2001). These graphical user interfaces may pose difficulties to users with disabilities and it has become essential that viable alternatives to mouse and keyboard input should be found (Hatfield and Jenkins 1997). Specially designed applications which take users with disabilities into consideration are available but these do not necessarily compare with the more popular applications. Disabled users

T. R. Beelders (✉) · P. J. Blignaut
University of the Free State, Bloemfontein, South Africa
e-mail: beelderstr@ufs.ac.za

M. Horsley et al. (eds.), *Current Trends in Eye Tracking Research,*
DOI 10.1007/978-3-319-02868-2_4,

should be accommodated in the same software applications as any other computer user, which will naturally necessitate new input devices (Istance et al. 1996) or the redevelopment of the user interface. Eye movement is well suited to these needs as the majority of motor-impaired individuals still retain oculomotor abilities (Istance et al. 1996). However, in order to disambiguate user intention and interaction, eye movement may have to be combined with another means of interaction such as speech. This study aims to investigate various ways to provide alternative means of input which could facilitate use of a mainstream product by disabled users. These alternative means should also enhance the user experience for novice, intermediate, and expert users. The technologies chosen to improve the usability of the word processor are speech recognition and eye tracking. The goal of this study is, therefore, to determine whether the combination of eye gaze and speech can effectively be used as an interaction technique to replace the use of the traditional mouse and keyboard within the context of a mainstream word processor. This will entail the development of a multimodal interface which will allow pointing-and-clicking, text entry, and document formatting capabilities.

The many definitions for multimodal interfaces (for example, Coutaz and Caelen 1991; Oviatt 1999; Jaimes and Sebe 2005; Pireddu 2007) were succinctly summarized for the purposes of this study as:

> A **multimodal interface** uses several human modalities which are combined in an effort to make human–computer interaction easier to use and learn by using characteristics of human–human communication.

Multimodal interfaces themselves date back to 1980, when Richard Bolt, in his seminal work entitled *Put That Here* (Bolt 1981), combined speech and gestures to select and manipulate objects. A distinct advantage of multimodal interfaces is that they offer the possibility of making interaction more natural (Bernhaupt et al. 2007). Furthermore, a multimodal interface has the potential to span across a diverse user group, including varying skill levels, different age groups as well as increasing accessibility for disabled users whilst still providing a natural, intuitive and pleasant experience for able-bodied users (Oviatt and Cohen 2000). For the purposes of this study, a multimodal interface was developed for a popular word processor application and tested as both a pointing device as well as for use as a text entry modality.

Both eye gaze (for example, Hansen et al. 2001; Wobbrock et al. 2008) and speech recognition (for example, Klarlund 2003) have been used in the past for the purpose of text entry. The current study will include eye gaze as an input technique but will require the use of an additional trigger mechanism, namely speech, in order to determine whether the accuracy and speed of the text entry method can be increased in this manner. The multimodal interface should also allow targets to be selected; thus, the viability of a number of pointing options was first investigated. Document formatting capabilities were provided through speech commands but the analysis thereof is beyond the scope of this chapter.

2 Background

Using a physical input device in order to communicate or perform a task in human–computer dialogue is called an interaction technique (Foley et al. 1990 as cited in Jacob 1995). However, for the purposes of this study, the definition will be modified and used in the following context:

> An **interaction technique** is the use of any means of communication in a human–computer dialogue to issue instructions or infer meaning.

Using eye gaze as an interaction device, specifically in the form of a pointing device, could seem natural as users tend to look at objects they are interacting with. However, the use of eye gaze does present some problems such as the Midas touch problem (Jacob 1991). Some of the associated problems of using gaze as a pointing device can be overcome through the use of an additional modality, such as speech.

Psycholinguistic studies have shown that there is a temporal relationship between eye gaze and speech (for example, Just and Carpenter 1976; Tanenhaus et al. 1995), often referred to as the eye–voice span. The eyes move to an object before the object is mentioned (Griffin and Bock 2000) with an approximate interval of 500 milliseconds between the eye movement and speech (Velichkovsky et al. 1997 as cited in Kammerer et al. 2008). However, recently it has been shown that these fixations on objects of interest could occur anywhere from the start of a verbal reference to 1500 milliseconds prior to the reference (Prasov et al. 2007). While the relationship between eye gaze and speech could be confirmed in a separate study, a large variance in the temporal difference between a fixation and a spoken reference to an object was also found (Liu et al. 2007) which could explain the various temporal differences reported on in different texts. This could lead to misinterpretation when attempting to react to verbal and visual cues in synchrony based on gaze position at the time a verbal command is uttered. However, eye gaze has been successful in resolving ambiguities when using speech input (Tanaka 1999) as it has been found that for the majority of verbal requests, users were looking at the object of interest when the command was issued. In order to maximize the disambiguation of both eye gaze and speech in this study, the user will be expected to maintain eye gaze on the desired object whilst issuing the verbal command to interact with that object.

The combination of eye gaze and speech has been used in the past for data entry purposes. For example, in a study conducted in the UK, eye gaze and speech could be used to complete a television license application (Tan et al. 2003a). In this instance, eye gaze was used to establish focus on a particular entry field and then dictation was used to complete the field which currently had focus. Users of the system much preferred using the eye gaze and speech to complete the application form even though it was neither the fastest nor the most accurate means of form completion tested.

The RESER and SPELLER (Tan et al. 2003b) systems used single-character entry mechanisms as opposed to dictation of complete words. The former application required users to gaze at the required key on a cluster keyboard and then to utter

the letter that they wished to type. Suggestions were given to complete the word currently being typed which the user could then accept or reject. The SPELLER application requires the entire word to be typed out character by character. For text entry, users preferred the mouse and the keyboard while speech and eye gaze were the preferred means of data recovery.

Dasher is a text entry interface which uses continuous pointing gestures to facilitate text entry (Ward et al. 2000). Speech Dasher extends the capabilities of Dasher even further by including speech recognition as well (Vertanen and MacKay 2010). Speech Dasher uses the same selection technique as the original Dasher but allows the user to zoom through entire words as opposed to single characters. The word set is obtained through speech recognition where the user speaks the text they would like to enter. With an error recognition rate of 22 %, users were able to achieve typing speeds of 40 WPM (Vertanen and MacKay 2010) which is similar to keyboard text entry. Speech Dasher is an example of a multimodal interface where gaze is used to enhance the capabilities of speech recognition.

The current study built on the idea that eye gaze can be used to establish which keyboard button is required by the user. However, instead of relying on the inaccurate or time-consuming methods of eye gaze only, an additional modality is suggested. The use of the look-and-shoot method with a physical trigger assumes that the user may have some mobility although it may be possible to use a triggering mechanism such as blowing in a pipe. Instead, this study will remove the reliance on physical dexterity and will build on the idea proposed by Tan et al. (2003b) that speech could be used to activate the focused key. However, it also assumes that some users may have limited vocabularies and may not be able to vocalize all alphabetic letters. Therefore, a single command, which can be customized to meet the abilities of the user, will be used to activate the key which currently has focus. Through this means, it will be possible to provide text entry capabilities using eye gaze and speech.

3 Eye Gaze and Speech as a Pointing Device

In order to use the proposed modality for text entry, it must first be established how eye gaze and speech can best be used as a pointing device, since in this context pointing forms the basis of text entry. Furthermore, if the interface is to be used within a word processor, the user must be able to select targets such as buttons.

The most commonly used metrics to evaluate pointing devices are speed and accuracy (MacKenzie et al. 2001) which give a good indication as to whether there is a difference between the performance of pointing devices (Hwang et al. 2004). ISO ratified a standard, ISO 9241-9, for determining the speed and accuracy of pointing devices for comparison and testing purposes. The ISO standard uses a throughput metric which encapsulates both speed and accuracy (ISO 2000) in order to compare pointing devices and is measured using any one of six tasks including three point-and-click tasks which conform to Fitts' law (Carroll 2003).

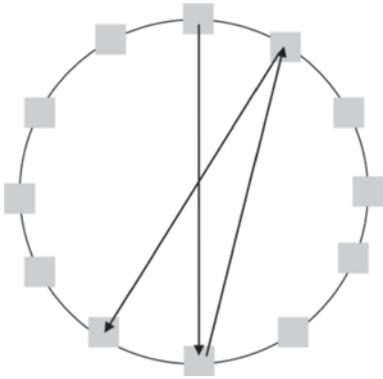

Fig. 1 Multidirectional tapping task

The one-directional tapping test requires the participant to move from a home area to a target and back. In contrast, the multidirectional tapping test (Fig. 1) consists of 24 boxes placed around the circumference of a circle. The participant is then required to move from the centre of the circle to a target box. From there the participant must move to and click in the box directly opposite that box and then proceed in a clockwise direction around the circle until all the targets have been clicked and the user is back at the first selected target box.

The ISO standard has been used to test eye tracking as an input device (Zhang and MacKenzie 2007). This test used the multidirectional tapping test across four conditions, namely (a) a dwell time of 750 ms, (b) dwell time of 500 ms, (c) look-and-shoot method which required participants to press the space bar to activate the target they were looking at and (d) the mouse (Zhang and MacKenzie 2007). A head-fixed eye-tracking system with an infrared camera and a sampling rate of 30 Hz was used for the study. The look-and-shoot method was the best of the three eye-tracking techniques with a throughput of 3.78 bps compared to the mouse with 4.68 bps.

The fact that the look-and-shoot method is the most efficient activation mechanism is not surprising since the selection time of a target is not dependent on a long dwell time and theoretically target acquisition times for all interaction techniques should be similar. Target acquisition in this chapter refers to when the target receives focus to such an extent that visual feedback is given. This does not imply that the target has been selected yet. Therefore, when a fixation is detected on a target or when the mouse enters the bounds of a target, the target is said to be acquired. The time required to press the space bar, particularly if users can keep their hand on it, should be shorter than the dwell time, which was confirmed by the results of the aforementioned study (Zhang and MacKenzie 2007). Recommendations stemming from the study included that a dwell time of 500 ms seemed the most appropriate so as to avoid the Midas touch problem whilst simultaneously ensuring that participants did not get impatient waiting for system reaction (Zhang and MacKenzie 2007). Increasing the width of the target reduced the number of errors made but had no effect on the throughput.

In a comparable study, the ISO standard was used to compare four pointing devices which could serve as a substitute mouse for disabled users (Man and Wong 2007). The four devices tested were the (1) CameraMouse, which was activated by body movements captured via a Universal Serial Bus (USB) webcam, (2) a Head-Array Mouse Emulator, an Adaptive Switch Laboratories, Inc. (ASL) mouse emulator that can provide solutions for power mobility, computer interfacing and environmental control for people with severe disabilities, (3) a CrossScanner, which has a mouse-like pointer activated by a single click and an infrared switch and (4) a Quick Glance Eye Gaze Tracker which allows cursor placement through the use of eye movement (Man and Wong 2007). Targets had a diameter of 20 pixels and the distance between the home and the target was 40 pixels. Two disabled participants, both with dyskinetic athetosis and quadriplegia, were tested over a period of eight sessions with two sessions per week. Each participant was analyzed separately and it was found that the CrossScanner was suitable for both participants although the ASL Head-Array was also suitable for use by one of the participants.

While ISO9241-9, similar to Fitts' law, is undoubtedly a step in the right direction, allowing researchers to establish whether there are differences in speed and accuracy between various pointing devices, it does, however, fail to determine why these differences exist (Keates and Trewin 2005). MacKenzie et al. (2001) propose seven additional measures which will provide more information as to why differences are detected between performance measures of pointing devices. These measures are designed to complement the measures of speed, accuracy and throughput and to provide more insight into why differences exist between pointing devices. The seven measures as proposed by MacKenzie et al. (2001) are as follows:

1. **Target re-entry**
   a. If the pointer enters the area of the target, leaves it and then re-enters it, a target re-entry has occurred.
2. **Task axis crossing**
   a. A task axis crossing is recorded if the pointer crosses the task axis on the way to the target. The task axis is normally measured as a straight line from the centre of the home square to the centre of the target (Zhang and MacKenzie 2007).
3. **Movement direction change**
   a. Each change of direction relative to the task axis is counted as a movement direction change.
4. **Orthogonal direction change**
   a. Each change of direction along the axis orthogonal to the task axis is counted as an orthogonal direction change.
5. **Movement variability**
   a. This "represents the extent to which the sample points lie in a straight line along an axis parallel to the task axis".
6. **Movement error**
   a. This is measured as the average deviation of the sample points from the task axis, regardless of whether these sample points are above or below the task axis.

7. **Movement offset**
   a. This is calculated as the mean deviation of sample points from the task axis.

The ISO9241-9 multidirectional tapping task was used to verify these metrics with 16 circular targets, each 30 pixels in diameter and placed around a 400-pixel-diameter outer circle (MacKenzie et al. 2001). These seven metrics, as well as throughput, movement time and missed clicks were used in a study to determine the difference in cursor movement for motor-impaired users (Keates et al. 2002).

A further six metrics, which could assist in determining why a difference exists, were specifically designed for use with disabled users and were proposed by Keates et al. (2002). These measures were not used during this study as they were not considered relevant. An additional metric measuring the number of clicks outside the target is also suggested in order to measure the performance of pointing devices (Keates et al. 2002).

4 Methodology

4.1 Experimental Design

The ISO test requires that the size of the targets and the distance between targets be varied in order to measure the throughput. In this study, however, variable size targets were used, but in order to reduce the time required to complete a test the distance between targets was not adjusted during this testing.

Standard Windows icons are 24×24 (visual angle $\approx 0.62°$) pixels in size. This was, therefore, used as the base from which to start testing target selection with speech recognition and eye gaze. Miniotas et al. (2006) determined that the optimal size for targets when using speech recognition and eye gaze as a pointing device was 30 pixels. This was determined using a 17″ monitor with a resolution of $1{,}024 \times 768$. Participants were seated at a viewing distance of 70 cm. This translated into a viewing angle of 0.85°. The eye tracker used in the current study was a Tobii T120 with a 17″ monitor where the resolution was set to $1{,}280 \times 1{,}024$. In order to replicate the viewing angle of 0.85° obtained by Miniotas et al. (2006), a 30-pixel target could be used but at a viewing distance of 60 cm from the screen. Therefore, the next size target to be tested in the trials was determined to be a 30×30 pixel button. It was decided to also test a larger target than that established by Miniotas et al. (2006). Following the example set by Miniotas et al. (2006) of testing target sizes in increments of 10 pixels, the final target size to be used was 40 pixels (visual angle $\approx 1.03°$).

The multidirectional tapping task used in this study had 16 targets situated on a circle with a diameter of 800 pixels. The square targets were positioned on the edges of the circle—thereby creating an inner circle with a diameter of 800 pixels.

Target acquisition was either via eye tracking and speech recognition (denoted by ETS for the purpose of this chapter) or the mouse (M). The mouse was used to

establish a baseline for selection speed. When using a verbal command to select a target, the subjects had to say "go" out loud in order to select the target that they were looking at. This method of pointing can, therefore, be considered analogous to look and shoot.

Magnification (ETSM) and the gravitational well (ETSG) were used to combat various shortcomings of using eye gaze for target selection, namely the instability of the eye gaze and the difficulties experienced in selecting small targets. The default zoom factor for the magnification enlarged the area to double its actual size within a 400×300 window while the gravitational well was activated within a 50-pixel radius around each button. The target button which had to be clicked was denoted by an "X".

This resulted in a total of 14 trials per session, the number of which served as motivation for not adding more trials for the mouse as this would simply prolong the session time and might cause participants to become irritable and fatigued during the session.

A balanced Latin square for all trial conditions was obtained by following the instructions provided by Edwards (1951). Participants were randomly assigned to a Latin square condition for each session.

Together with the throughput measure of the ISO standard, additional measurements were analyzed in an effort to explain the difference in performance if such a difference exists between the interaction techniques. To this end, the total task completion time was measured as well as the task completion time from when the target was highlighted to when it was clicked, the number of target re-entries, the number of incorrect targets which were acquired during task completion and the number of incorrect clicks. This will allow efficiency and effectiveness of each interaction technique to be tested.

4.2 Participants

Participants, who were senior students at the university at which the study was conducted, volunteered to participate in the study. For each session completed, the participant received a small cash amount.

Each participant completed three sessions and each session consisted of all 14 trials. In total there were 15 participants who completed all three sessions.

Eleven of the participants were male and four were female. The average age of the participants was 22.3 (standard deviation = 1.9). The only selection criteria was that the participant have normal or corrected-to-normal vision, that they were proficient with the mouse and that they had no prior experience with either eye tracking or speech recognition.

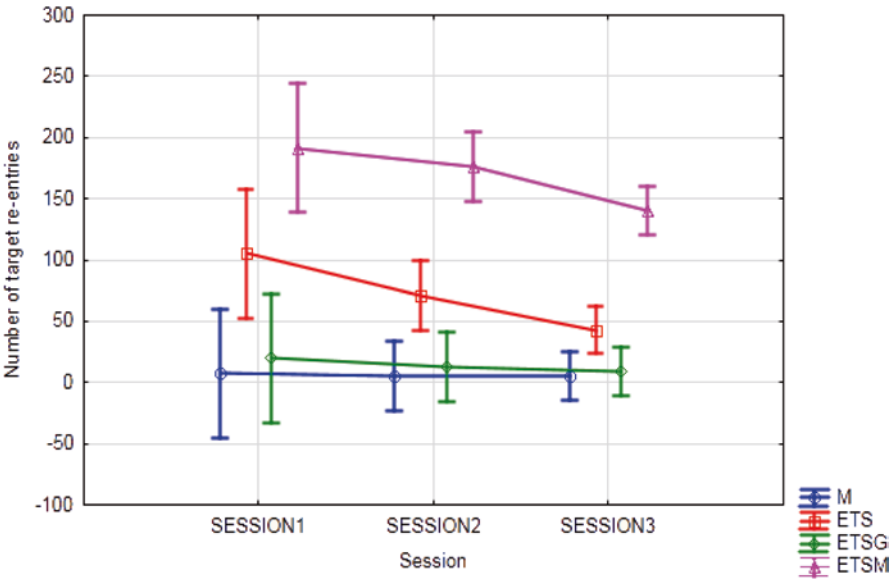

Fig. 2 Target re-entries for all interaction techniques

5 Results

5.1 Throughput and Time to Complete a Trial

As stipulated in the ISO test, the throughput was measured and analyzed for each of the interaction techniques. The results of this analysis, as well as the time to complete a trial, are discussed in detail in Beelders and Blignaut (2012). In summary, the mouse had a much higher throughput than the other interaction techniques. In terms of the time to complete a trial, the mouse and the use of the gravitational well had comparable selection times. Interestingly, when using a gravitational well, the time to select a target was much faster than when using any other interaction technique, including the mouse.

5.2 Target Re-Entries

Target re-entries were defined as the number of times the designated target was gazed upon before the user was able to click on it.

The graph in Fig. 2 plots the number of target re-entries for all interaction techniques over the three sessions.

At an α-level of 0.05, there was a significant difference between the number of target re-entries for the different interaction techniques ($F(3, 56) = 32.071$). Post-hoc

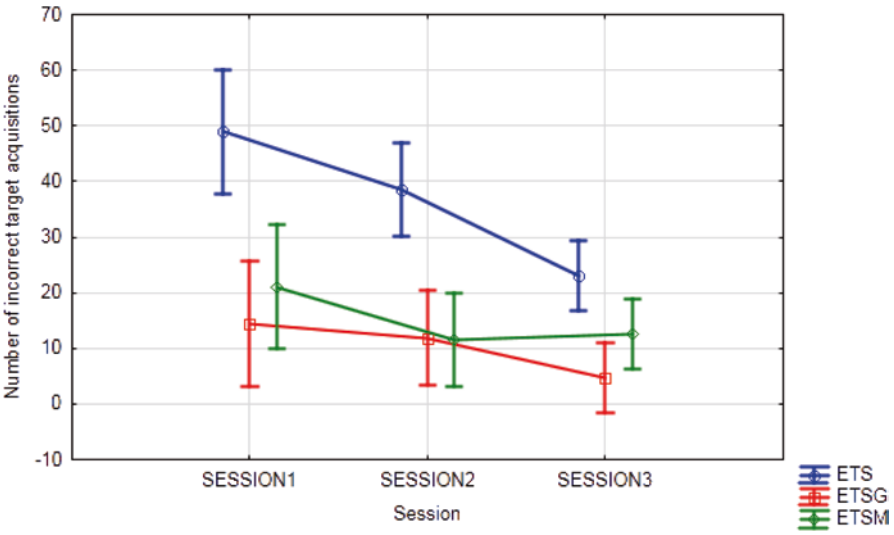

Fig. 3 Incorrect target acquisitions for all interaction techniques

tests indicated that ETSM had a significantly higher incidence of target re-entries than the other interaction techniques. This would imply that it was much more difficult to achieve a prolonged stable gaze on a button such that the required verbal command can be issued when the magnification tool was activated than for any other interaction technique. ETS also differed significantly from the mouse and ETSG. ETSG did not differ significantly from the mouse, which means that ETSG is able to perform comparably with the mouse in terms of target re-entries.

There was also a significant difference between the sessions ($F(2, 112) = 4.249$).

5.3 Incorrect Target Acquisitions

Incorrect target acquisitions were defined as the number of times a target, which was not the designated target, was acquired. This means that in the event of the eye tracker and speech being used, each time a button received enough focus to give visual feedback, the incorrect target acquisitions were incremented, provided that the focused button was not the designated target. The number of incorrect target acquisitions was counted as those targets which were acquired *after* the designated target had been acquired. Therefore, the incorrect targets that were acquired could not be attributed to normal searching for the designated target. For the purposes of this measurement, only the eye gaze and speech interaction techniques will be included in the analysis as the number of incorrect target acquisitions for the mouse interaction techniques was always zero.

The graph in Fig. 3 plots the number of incorrect target acquisitions for all included interaction techniques over all sessions.

For this measure, ETSG had the best performance, although all interaction techniques exhibited some degree of improvement, most notably that of ETS.

At an α-level of 0.05, there was a significant difference between the interaction techniques ($F(2, 42) = 19.327$) as well as the sessions ($F(2, 84) = 12.046$, $p < 0.05$).

All the sessions differed significantly from one another. Since only ETSM actually increased slightly in session 3, it can be surmised that the incorrect target acquisitions lessened at a significant rate over time. ETS, in particular, had a sharp decrease and it may be beneficial to increase the number of sessions so that it can be properly analyzed whether it can ever reach the low values of ETSG or ETSM. In terms of the interaction techniques, ETS differs significantly from both ETSG and ETSM. ETSG and ETSM do not differ significantly from each other.

Observations made of the participants while they were completing the tasks could provide an explanation for this. Many participants soon realised that when struggling to focus on a button it was sometimes easier to focus on another button at a suitable distance from the designated one. It was not necessary to focus on this other button for a protracted time. Participants would then look back at the designated button and the extended movement seemed to provide more accuracy in focusing on the desired target rather than trying to "fine-tune" the selection within a small area around the designated button. The smoothing algorithm could have contributed to this as small movements within a certain radius are interpreted as a single fixation. Since the gravitational well effectively pulls the selection onto the nearest target once the "pointer" is within a certain distance, it becomes easier to focus on a target and no fine-tuning is required. This could explain the reason why ETSG has such a low number of acquisitions compared to ETS.

ETSM also has a lower rate and this could possibly be attributed to participants rather trying to fine-tune the selection when using the magnification. Since the buttons appear larger, participants may have perceived the fine-tuning process to be easier since a larger target could create the impression that it can be easily acquired. The high incidence of target re-entries coupled with the low number of incorrect target acquisitions may serve to substantiate the suspicion that fine-tuning was the preferred method for ETSM.

The similar pattern for ETS, regarding target re-entries and target acquisitions, also corroborates the claim that the participants preferred to employ the use of a shifting of their eye gaze to focus on another button and then returning to the designated button. Closer inspection of the averages for ETS shows that incorrect target acquisitions constituted approximately half the number of re-entries for each session. This could indicate that participants would attempt to re-acquire the designated target and, when they were unable to achieve a stable selection, they resorted to focusing on another target before attempting to select the designated target—in contrast to the strategy employed with ETSM.

The reason for this could be that the magnification disturbs the users while they adjust their gaze and they are unwilling to move their gaze substantially because they perceive this to require more effort when magnification is activated. Another reason for the different strategies could be attributed to the fact that the magnification tool that was used has in-built visual feedback which allows the user to get an

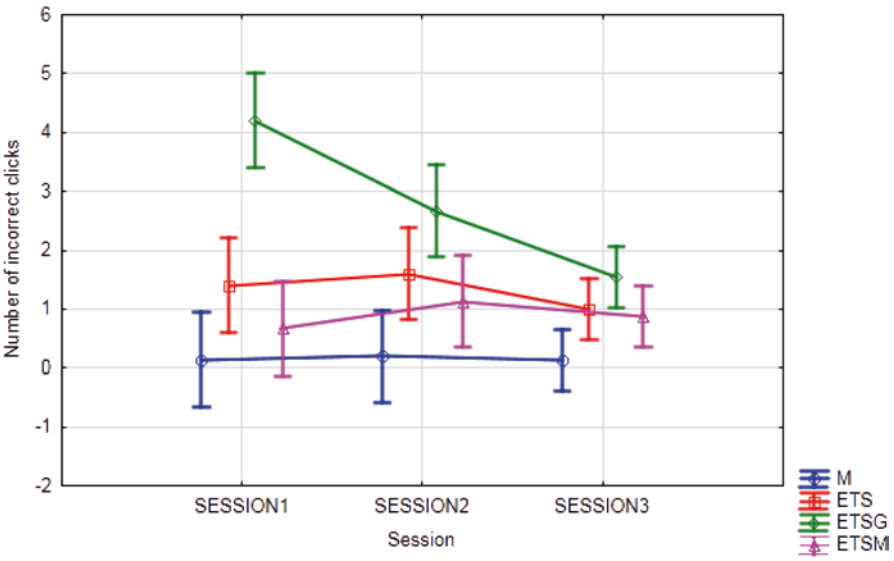

Fig. 4 Incorrect clicks for all interaction techniques

approximation of their eye gaze position, which is centred in the magnified area. Since this feedback is present, the user may feel that fine-tuning is a better option since they can determine how close they are to the target, which is not the case with ETS. With ETS, they will know they have lost the target but not how close they are to re-acquiring it; hence, they feel more secure glancing at another target, establishing position and then looking at the required target again until they can maintain a stable eye gaze. Therefore, to slave a cursor to the eye gaze may be disruptive but in this instance it could tentatively be said that it may have provided useful information to the participants. However, the evidence suggests that it in no way increased the efficiency or effectiveness of target selection and, therefore, it is not recommended for use.

The average number of target re-entries for ETSG was roughly the same as the average incorrect target acquisitions for ETSG. This could provide evidence that when using ETSG, the target was easier to acquire and it was easier to keep the focus long enough to issue the required command. Since the buttons were effectively larger, it would make sense that they were easier to focus on for a prolonged period of time.

5.4 Incorrect Clicks

Incorrect clicks were determined as the number of times a target that was not the designated target was clicked during a trial.

The graph in Fig. 4 plots the number of incorrect clicks for all interaction techniques.

Owing to the fact that there was significant interaction between the factors, each session was analyzed separately to determine whether there was a difference between the interaction techniques. It was found that in the first session, ETSG differed significantly from all other techniques; in the second session ETSG differed significantly from the mouse and ETSM and in the third session only from the mouse.

These results clearly show that ETSG resulted in the highest number of incorrect clicks. Although continued practice allowed ETSG to have a comparable number of incorrect clicks to ETS and ETSM, its performance could not match that of the mouse over the three sessions. This indicates that some learning did take place over the three sessions.

Natural eye movement may provide an explanation for the observed difference. Participants could acquire the target and then issue a verbal command while already starting to look at the next target (for all eye gaze and speech interaction techniques). Since the use of the gravitational well increases the speed with which a target can be acquired, this often meant that by the time the speech engine recognized the command, the next target had already been acquired. This could account for the high number of incorrect targets for ETSG. These findings also confirm previous findings that the fixation immediately prior to the action or command being issued usually occurs on the object of interest (Land and Tatler 2009; Maglio et al. 2000).

Since ETSG had significantly lower incorrect target acquisitions, coupling it with this finding of more incorrect clicks creates the following dilemma. The use of the gravitational well increases the possibility of correctly acquiring a target and maintaining a stable gaze on the target. This is evidenced by the fact that other eye gaze and speech interaction techniques caused participants to first glance away, acquire another target and then glance back. However, the fact that a gravitational well is present together with human tendency to start glancing at the next object of interest whilst still issuing a command to the current target, means that the next target is acquired far quicker than when no gravitational well is present. This causes the next target to be incorrectly clicked on with higher frequency for ETSG. Since participants started moving their eye gaze away from the buttons before the speech command had been executed for all eye gaze interaction techniques, it would be assumed that for ETS and ETSM, which pose greater difficulty in target acquisition, the participant would inadvertently have caused a click somewhere on the application form which was not a clickable area. Unfortunately, this measurement was not captured during these tests. Further research must be done in order to determine if this proposition is true.

6 Discussion

Incorrect clicks were experienced with all eye gaze interaction techniques although more so with ETSG. Nevertheless, this finding corresponds with the finding of Kaur et al. (2003) that the target which was acquired a certain amount of time prior

to command execution is the target that must be selected. Although the interval was found to be 630 ms (Kaur et al. 2003), this interval will have to be confirmed for use with eye gaze and speech. While natural eye gaze movement appears to dictate that the target prior to command utterance must be selected, it must still be determined whether this will appear natural to the user or whether they would prefer to adapt to the use of ETSG as it was tested in this study. Clearly, practice allows them to adjust their natural behaviour to a degree to compensate for the interaction technique as is evidenced by the improvement over the sessions. However, requiring users to change their natural behaviour is not the aim of a multimodal interface. Therefore, it becomes necessary to establish the interval required for target selection and test the usability of that compared to the standard gravitational well employed in this study.

Previous studies such as the touch-sensitive mouse (Drewes and Schmidt 2009) and MAGIC pointing (Zhai et al. 1999) warped the mouse pointer to the position of the eye gaze and then users were required to use the mouse to click on the desired target. Although this exploits the high speed of eye gaze and also reduces incidences of incorrect clicks since users are not likely to click on the incorrect target when having to manually manipulate a mouse pointer, some physical dexterity is required. The solution may lie in a combination of this technique and speech. Eye gaze could be used to establish intent, a single voice command could be issued to warp the pointer to the selectable target closest to the current eye gaze, and once the user has verified that the correct target is acquired, a second command can be issued to click on the target. For fine-tuning purposes of the mouse cursor, direction- or target-based navigation can also be provided.

7 Multimodal Word Processor

The next step was to test the modality when used for text entry. For these purposes, it was decided to use the familiar environment of Microsoft Word® and to simply develop a multimodal interface for Word. Visual Studio Tools for Office (VSTO), which allows developers to create extensions to the Office Suite with customized functionality (Anderson 2009), was used to add multimodal functionality (Fig. 5).

An on-screen keyboard was available which was overlaid on the bottom of the current document. Users could then type in the Word document by focusing their gaze on the desired character on the keyboard and issuing a verbal command to trigger the keyboard key. Auditory feedback, in the form of a beep, was given to alert the user that the character had been typed. This should allow them to continue typing without having to glance back at the document for confirmation.

As can be seen in Fig. 5, a magnification tool was available which allowed the area directly under the gaze of the user to be enlarged. Typing tasks using the magnification tool were not required during this study and will, therefore, not be discussed in this chapter. Figure 5 also shows a number of other customizations which were available in the multimodal interface which was developed, most of which are beyond the scope of this chapter but it is interesting to note their inclusion.

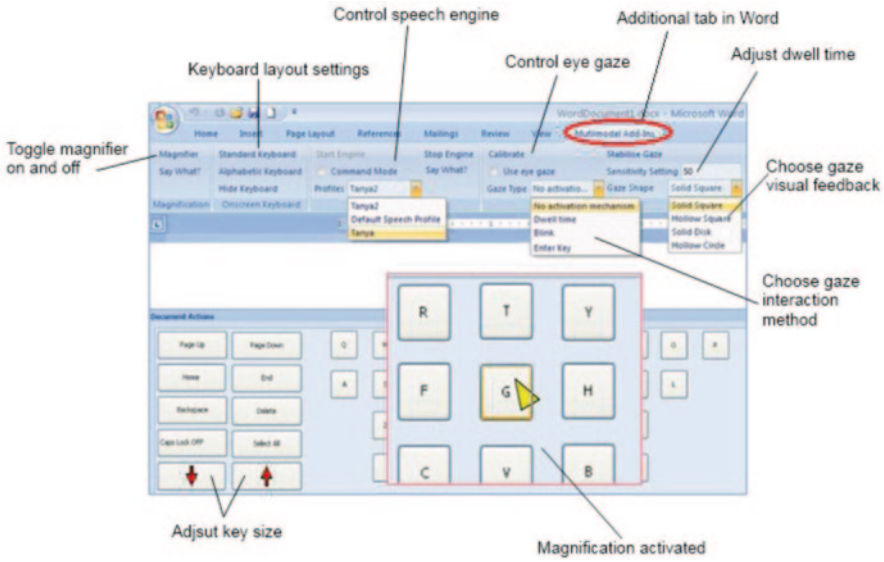

Fig. 5 Multimodal add-in for Microsoft Word

8 Analysis

8.1 Participants

A total of 25 participants participated in the 10-week long study. A prerequisite for participation in the study was sufficient computer literacy as well as word processor expertise. Forty percent of the sample was drawn from second-year computer science students who were registered for a community service module. The other 60 % of the sample was drawn from the student assistants for the computer literacy course of the university, with the proviso that they were not studying for a computer science or related degree. These students all had to complete the literacy course prior to becoming an assistant and they had to achieve at least 70 % for a competency test of Microsoft Office applications. Their proficiency with Word was verified through the completion of a questionnaire before the commencement of the study. The questionnaire evaluated the duration and frequency of use in order to determine an expertise measurement.

The first week was simply an introductory session and the data collected there were not included in the analysis. Furthermore, the data of three participants had to be discarded from the sample due to various reasons. Of the remaining 22 participants, only 8 completed all sessions on the on-screen keyboard and 14 with the traditional keyboard. These were the participants who were included in the analysis.

8.2 Tasks

Each participant had one session per week during which they were expected to complete a series of tasks on the adapted word processor. Three of these tasks were typing tasks with the on-screen keyboard and two with the traditional keyboard. For each session, the buttons of keyboard were sized at 60 × 60 pixels (≈1.55° visual angle). Buttons were spaced 60 pixels apart with a gravitational well of 20 (≈0.52° visual angle) pixels on all sides of each button. The gravitational well effectively increased the selection area of each button since once the gaze was detected within the bounds of the gravitational well it was pulled onto the button. Participants were not aware of the gravitational well as it was not visible.

Additional typing tasks were added from the fifth session onwards in order to test varying sizes and spacing between buttons. These additional tasks were added to the end of the existing task list. By then the majority of the participants were completing the current task list in less than 30 min. No pressure was placed on the participants to complete all tasks within their scheduled time so it was felt that adding additional tasks to the end of the test would not unduly cause any more anxiety or place more strain on the participants. Within these additional typing tasks, the first one had to be completed using the originally sized and spaced buttons. The next two had to be completed with buttons that were 50 × 50 (≈1.29° visual angle) pixels in size and spaced 70 (≈1.80° visual angle) pixels apart. Following this, there were another two tasks which had to be completed using buttons that were 50 × 50 pixels in size but were spaced 60 pixels apart. The original configuration of button will henceforth be referred to as speech-SC (small, closely spaced), the larger button configuration as speech-L and the smaller more widely spaced configuration will be referred to as speech-SW.

The typing tasks required the participant to type a phrase that was randomly selected from a set of 35 phrases. The phrase set used was a subset of the 500 as determined by MacKenzie and Soukoreff (2002) to be everyday phrases which are commonly used. The results of the typing tests from session 5 onwards will be the focus of this chapter.

8.3 Measures

The Levenshtein distance (Levenshtein 1965) between two strings measures how many insertions, deletions and substitutions have taken place between presented text and transcribed text. The sum of these errors can then divided by the number of characters to give a character error rate (CER) (Read 2005). Since there are multiple ways in which the presented text can be transformed into the transcribed text, the possible transformations or optimal alignments were identified, their mean length was calculated and then the Levenshtein distance was divided by this mean length to give an error rate (MacKenzie and Soukoreff 2003). This was how the CER was calculated for text entry in this study.

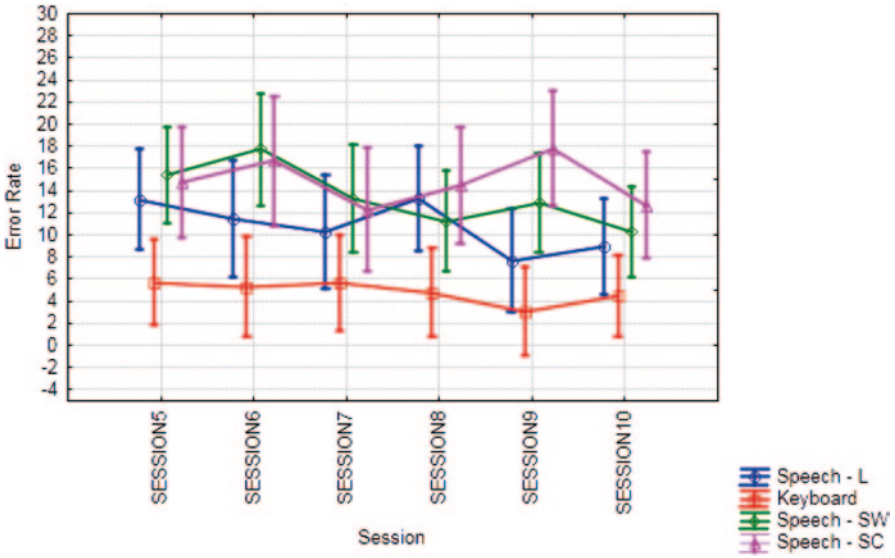

Fig. 6 Character error rate for secondary study

The second measurement analyzed during the study was characters per second (CPS). This measures the number of characters that were typed and then divides it by the time taken to type the characters, measured in seconds. In order to ensure that the time it takes to read a phrase does not unduly influence the results, the time taken to type the phrase was measured from when the first character was typed to when the last character was typed.

9 Results of Study

9.1 Character Error Rate

The graph in Fig. 6 plots the mean error rate for all sessions and for all interaction techniques.

From the graph, it can be surmised that the keyboard had the lowest error rate of all interaction techniques for all sessions. Thereafter, speech-L had the next lowest error rate while the smaller buttons, both speech-SW and speech-SC, caused the highest error rates for all sessions. The latter two seem to cause approximately the same error rates while typing; however, the widely spaced buttons have an improved error rate during the later sessions while the error rates for the closely spaced buttons increased over the same period.

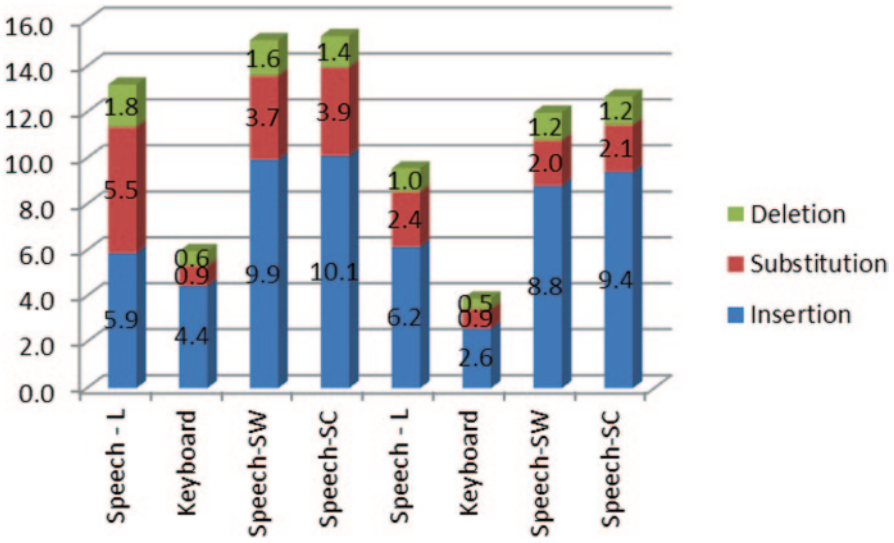

Fig. 7 Breakdown of **character error rate** (CER) for secondary study

At an α-level of 0.05, it was found that there was a significant difference between the error rate of the different interaction techniques ($F(3, 43) = 7.303$). Post-hoc tests indicate that the keyboard differed significantly from both speech-SW and speech-SC. In this instance, it is encouraging to determine that speech-L does not differ significantly from the keyboard in these later sessions. This would seem to indicate that after some practice with the larger buttons, the number of errors made decreases. The same cannot be said of the smaller buttons. There was also a significant difference between the CER for the sessions ($F(5, 215) = 2.530$), where session 6 differed significantly from session 10.

The errors were then categorized as insertions, deletions or substitutions and further analyzed as such. The bar graph in Fig. 7 shows the breakdown for the first session (first four stacks) using all interaction techniques and the last session (last four stacks). Deletions are on top of each stack, substitutions are in the middle and insertions constitute the lower part of each stack.

The percentage of insertion errors was the highest for all interaction techniques for both of these sessions. The interaction techniques of speech-SW and speech-SC have very similar distributions over the number of insertions, substitutions and deletions.

At an α-level of 0.05, there was a significant difference between the interaction techniques in terms of the number of insertions ($F(3, 44) = 4.100$), but not for deletions ($F(3, 39) = 1.638$). Owing to significant interaction between the factors, separate analyses had to be performed for the substitution errors where it was found that there was a significant difference between the interaction techniques for all sessions.

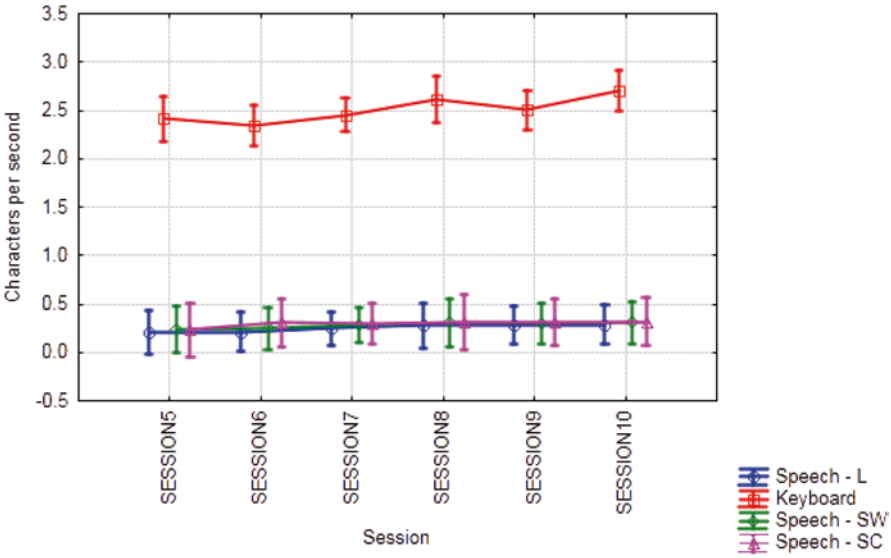

Fig. 8 Characters per second for secondary study

Post-hoc tests indicated that the use of speech-SW resulted in significantly more insertions than the keyboard. For the substitution errors, the keyboard generally had significantly less errors than a variety of the speech interaction techniques depending on the session.

9.2 Characters per Second

The CPS were measured for each interaction technique and for each session. The graph in Fig. 8 plots this measure for each session and each interaction technique.

When using the keyboard, participants were clearly able to type at a much faster rate than when using eye gaze and speech with the on-screen keyboard, which remained fairly consistent regardless of the keyboard settings. As can be expected, there was a significant difference between the interaction techniques at an α-level of 0.05 ($F(3, 44) = 148.369$). There was a significant difference between sessions ($F(5, 15) = 3.002$); in particular, session 10 differed significantly from sessions 5 and 6.

10 Discussion

It was found that the eye gaze and speech interaction technique had a significantly higher error rate than that of the keyboard, undoubtedly as a result of a higher number of insertions and substitutions. This may serve as confirmation that even when

using eye gaze and speech as a text input mechanism, the user is inclined to glance away before completing the issuing of the verbal command. The average insertions are generally higher than the substitutions which would seem to indicate that users are aware that they have activated the incorrect character and attempt to correct it by inserting the correct character. This is encouraging as it indicates that users become familiarized with the system such that they can interpret the selection (indicated by audio feedback) and are able to make corrections to text entry. Further research could confirm these suppositions by capturing the correction of the text input as well so that it can be analyzed to determine whether incorrect inputs are reversed/erased before text input is continued. Whether the buttons are large, small and widely spaced or small and closely spaced seems to be of little consequence. There was no difference between the error rates of these three interaction techniques and they all differed from the keyboard at some stage. However, the interaction technique of speech-L did seem to offer the most improved error rate as it did not differ from the keyboard when analyzed for the later sessions only. In some instances, there was improvement over the sessions, which indicates some measure of learning when using the interaction technique. If the learning effect can be maintained then more practice with the eye gaze and speech could eventually lead to an effectiveness measurement which is comparable to that of the keyboard.

In terms of efficiency, the keyboard also outperformed all the eye gaze and speech interaction techniques with significantly higher numbers of CPS which could be typed. The typing speed of the eye gaze and speech also did not improve as exposure increased. This could indicate that either more practice is needed to achieve increased speeds or the typing speed quickly reaches the fastest achievable rate. Neither the size of the buttons nor the spacing between buttons affected the efficiency of the eye gaze and speech.

Therefore, in terms of effectiveness and efficiency, the three eye gaze and speech interaction techniques seem fairly interchangeable as they perform on comparable levels to each other. The keyboard is far more effective and efficient than any of the eye gaze and speech interaction techniques when used for text input.

No similar studies were found with which these results could be compared. However, the fact that speech outperforms keyboard input for young children (Read et al. 2001) indicates that the learning curve for keyboard entry is fairly steep. This could be the same for text entry with eye gaze and speech. Although there was no significant improvement in the speed of the text entry, participants clearly became more comfortable with the use of the interaction technique. Therefore, extended practice may be required to improve speeds.

The mean entry rate of eye gaze and speech fell within the range between 0.2 and 0.3 CPS. Considering that the entry rate was relatively low for context switching at 12 WPM (Morimoto and Amir 2010) and 9 WPM for symbol creator (Miniotas et al. 2003), the range in this study was much lower than these previous studies. A previous study (Majaranta 2009) showed that the use of both visual and auditory feedback increased the entry speed to 7.55 WPM which is still faster than the speeds achieved in this study. Speech Dasher achieved much higher speeds (40 WPM), while using Dasher with eye gaze also resulted in higher speeds (17 WPM).

Therefore, when comparing the text entry method to studies using only eye gaze without text predictors, speech and eye gaze performs slightly better. However, the speeds are still lower than using text prediction methods and when using speech as an activator. While these comparisons are promising since they indicate that speech and eye gaze could facilitate faster entry speeds than using eye gaze only, they are discussed with caution since the text entered in the current study required only a few short phrases to be entered and more prolonged use could have an impact on the entry speed.

11 Conclusion

As evidenced by the incorporation of the technologies used in the multimodal interface of this study, the time has perhaps dawned when they should be exploited as replacement interaction techniques. Speech recognition has become a standard feature in personal computers and is often available for dictation purposes. Similarly, there are packages available for purchase which can react to spoken commands (cf. Dragon n.d.). Furthermore, the first fully integrated eye-controlled laptop has recently been showcased at exhibitions (Tobii 2011) and bodes well for the adoption of eye tracking as a standard feature in personal computers. Cheaper, accurate eye trackers (cf. Haro et al. 2000) are also available which could function just as well as a standard interaction technique.

Therefore, the fact that a popular mainstream application can be adapted to include a highly customizable, multimodal interface could be a step in the right direction for the next generation of interfaces. The multimodal user interface displays great potential and test results indicate that the interaction techniques can be used for pointing and selecting tasks and common word processing tasks. Moreover, it has been proven that speech recognition can indeed be used for editing commands in a word processor which was contrary to theoretical beliefs (Klarlund 2003). This could mean that in the future a more diverse group of users can be accommodated and disabled users may no longer have to be relegated to using specialized applications.

The findings, therefore, suggest that the word processor is well placed to include such an interface in future developments as the technology is rapidly becoming available. As it is foreseen that access to the technologies by mainstream users is imminent, future word processors could be developed with multimodal interfaces incorporated.

The combination of eye gaze and speech could successfully be used to fulfil the needs of a pointing device, particularly when employed with a gravitational well. While text entry was slower than using a keyboard, indications are that there was an overall positive response to the interface and that it may well herald a suitable multimodal interface. The ease with which participants became accustomed to the interface is further proof of the naturalness and intuitiveness provided by speech and eye gaze. With constant progress being made in the development of the hardware required by such an interface, the proposed multimodal interface may well

lay the foundation for a word processor to continue its exploitation of emerging technologies and remain a forerunner in the establishment of trends. While there is undoubtedly room for improvement and expansion, the use of eye gaze and speech has proven to be very promising.

References

Anderson, T. (2009). *Pro office 2007 development with VSTO*. United States of America: APress

Beelders, T. R., & Blignaut, P. J. (2012). Using eye gaze and speech to simulate a pointing device. In *Proceedings of the symposium on eye-tracking research and application (ETRA)*, Santa Barbara, California

Bernhaupt, R., Palanque, P., Winkler, M., & Navarre, D. (2007). Usability study of multi-modal interfaces using eye-tracking. In *Proceedings of INTERACT 2007*, 412–424

Bolt, R. (1981). Gaze-orchestrated dynamic windows. *Computer Graphics*, *15*(3), 109–119.

Carroll, J. M. (2003). *HCI models, theories, and frameworks: Towards a multidisciplinary science*. San Francisco: Morgan Kaufmann.

Coutaz, J., & Caelen, J. (1991). A taxonomy for multimedia and multimodal user interfaces. In *Proceedings of the Second East-West HCI conference*, St Petersburg, Russia, 229–240

Dragon Naturally Speaking. (nd). History of speech and voice recognition and transcription software. Retrieved 13 Feb 2009 from http://www.nuance.com

Drewes, H., & Schmidt, A. (2009). The MAGIC touch: Combining MAGIC-pointing with a touch-sensitive mouse. In *Human-Computer Interaction—INTERACT 2009. 12th IFIP TC 13 International Conference, Part II*, Uppsala, Sweden, 415–428

Edwards, A. L. (1951). Balanced Latin-square designs in psychological research. *The American Journal of Psychology*, *64*(4), 598–603.

Foley, J. D., Van Dam, A., Feiner, S. K., & Hughes, J. F. (1990). *Computer graphics: Principles and practice*. Reading, Massachusetts: Addison-Wesley.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274–279.

Hansen, J. P., Hansen, D. W., & Johansen, A. S. (2001). Bringing gaze-based interaction back to basics. In C. Stephanidis (Ed.), *Universal Access in HCI (UAHCI): Towards an Information Society for All—Proceedings of the 9th International Conference on Human-Computer Interaction (HCII'01)*, 325–328. Mahwah: Lawrence Erlbaum Associates

Haro, A., Essa, I., & Flickner, M. (2000). A non-invasive computer vision system for reliable eye tracking. In *Proceedings of CHI '00*, The Hague, Netherlands, 167–168

Hatfield, F., & Jenkins, E. A. (1997). An interface integrating eye gaze and voice recognition for hands-free computer access. In *Proceedings of the CSUN 1997 Conference*, 1–7

Hwang, F., Keates, S., Langdon, P., & Clarkson, J. (2004). Mouse movements of motion-impaired users: A submovement analysis. In *Proceedings of ASSETS '04*, Atlanta, Georgia, United States of America, 102–109

Istance, H. O., Spinner, C., & Howarth, P. A. (1996). Providing motor impaired users with access to standard Graphical User Interface (GUI) software via eye-based interaction. In *Proceedings of 1st European Conference on Disability, Virtual Reality and Associated Technology*, Maidenhead, United Kingdom, 109–116

ISO. (2000). *ISO 9241–9: Ergonomic requirements for office work with visual display terminals (VDTs)—Part 9: Requirements for non-keyboard input devices*. International Organization for Standardization

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*(2), 152–169.

Jacob, R. J. K. (1993). Eye movement-based human-computer interaction techniques: Toward non-command interfaces. In H. R. Hartson & D. Hix (Eds), *Advances in human-computer interaction, 4, 151–190. Norwood*. New Jersey: Ablex Publishing.

Jacob, R. J. K. (1995). Eye tracking in advanced interface design. In W. Barfield & T. A. Furness (Eds.), *Virtual environments and advanced interface design* (pp. 258–288). New York: Oxford University Press.

Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises (Section Commentary). In J. Hyona, R. Radach & H. Deubel (Eds), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573–605). Amsterdam: Elsevier Science.

Jaimes, A., & Sebe, N. (2005). Multimodal human computer interaction: A survey. *IEEE workshop on human computer interaction*, Las Vegas, Nevada, United States of America, 15–21

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*, 441–480.

Kammerer, Y., Scheiter, K., & Beinhauer, W. (2008). Looking my way through the menu: The impact of menu design and multimodal input on gaze-based menu selection. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA)*, Savannah, Georgia, United States of America, 213–220

Kaur, M., Tremaine, M., Huang, N., Wilder, J., Gacovski, Z., Flippo, F., & Mantravadi, S. (2003). Where is "it"? Event synchronization in gaze-speech input systems. In *Proceedings of ICIM '03*, Vancouver, Canada, 151–158

Keates, S., Hwang, F., Langdon, P., Clarkson, P. J., & Robinson, P. (2002). Cursor movements for motion-impaired computer users. In *Proceedings of ASSETS '02*, Edinburgh, Scotland, 135–142

Keates, S., & Trewin, S. (2005). Effect of age and Parkinson's Disease on cursor positioning using a mouse. In *Proceedings of ASSETS '05*, Baltimore, Maryland, United States of America, 68–75

Klarlund, N. (2003). Editing by voice and the role of sequential symbol systems for improved human-to-computer information rates. In *Proceedings of ICASSP*, Hong Kong, 553–556

Land, M. F., & Tatler, B. W. (2009). *Looking and acting: Vision and eye movements in natural behaviour*. United States of America: Oxford University Press.

Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Doklady Akademii Nauk*, *163*, 845–848.

Liu, Y., Chai, J, Y., & Jin, R. (2007). Automated vocabulary acquisition and interpretation in multimodal conversational systems. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*

MacKenzie, I. S., Kauppinen, T., & Silfverberg, M. (2001). Accuracy measures for evaluating computer pointing devices. In *Proceedings of SIGCHI '01*, Seattle, Washington, United States of America, 9–16

MacKenzie, I. S., & Soukoreff, R. W. (2002). A character-level error analysis technique for evaluating text entry methods. In *Proceedings of NordiCHI 2002*, Aarhus, Denmark, 243–246

MacKenzie, I. S., & Soukoreff, R. W. (2003). Phrase sets for evaluating text entry techniques. In *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems—CHI 2003*, Fort Lauderdale, Florida, United States of America, 754–755

Maglio, P. P., Matlock, T., Campbell, C. S., Zhai, S., & Smith, B. A. (2000). Gaze and speech in attentive user interfaces. In *Proceedings of the Third International Conference on Advances in Multimodal Interfaces*, Vancouver, Canada, 1–7

Majaranta, P. (2009). Text entry by eye gaze. Dissertations in Interactive Technology, number 11, University of Tampere

Man, D. W. K., & Wong, M.-S., L (2007). Evaluation of computer-access solutions for students with quadriplegic athetoid cerebral palsy. *American Journal of Occupational Therapy*, *61*, 355–364.

Miniotas, D., Špakov, O., & Evreinov, G. (2003). Symbol Creator: An alternative eye-based text entry technique with low demand for screen space. In *Proceedings of Human Computer Interaction—INTERACT '03*, Zurich, Switzerland, 137–143

Miniotas, D., Špakov, O., Tugoy, I., & MacKenzie, I. S. (2006). Speech-augmented eye gaze interaction with small closely spaced targets. In *Proceedings of the 2006 Symposium on Eye Tracking Research and Applications (ETRA)*, 67–72

Morimoto, C. H., & Amir, A. (2010). Context switching for fast key selection in text entry applications. In *Proceedings of the 2010 Symposium on Eye Tracking Research and Applications (ETRA)*, 271–274

Oviatt, S. (1999). Mutual disambiguation of recognition errors in a multimodal architecture. In *Proceedings of the ACM SIGCHI 99*, Pittsburgh, Pennsylvania, United States of America, 576–583

Oviatt, S., & Cohen, P. (2000). Multimodal interfaces that process what comes naturally. *Communications of the ACM*, *43*(2), 45–53.

Pireddu, A. (2007). Multimodal Interaction: An integrated speech and gaze approach. Thesis, Politecnico di Torino

Prasov, Z., Chai, J. Y., & Jeong, H. (2007). Eye gaze for attention prediction in multimodal human-machine conversation. In *Proceedings of AAAI Spring Symposium on Interaction Challenges for Intelligent Assistants*

Read, J. (2005). On the application of text input metrics to handwritten text input. Text Input Workshop, Dagstuhl., Germany

Read, J., MacFarlane, S., & Casey, C. (2001). Measuring the usability of text input methods for children. In *Proceedings of Human-Computer Interaction (HCI) 2001*, New Orleans, United States of America, 559–572

Tan, Y. K., Sherkat, N., & Allen, T. (2003a). Eye gaze and speech for data entry: A comparison of different data entry methods. In *Proceedings of the International Conference on Multimedia and Expo*, Baltimore, Maryland, United States of America, 41–44

Tan, Y. K., Sherkat, N., & Allen, T. (2003b). Error recovery in a blended style eye gaze and speech interface. In *Proceedings of ICMI '03*, Vancouver, Canada, 196–202

Tanaka, K. (1999). A robust selection system using realtime multi-modal user-agent interactions. In *Proceedings of IUI'99*, 105–108

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science*, *268*, 1632–1634.

Tobii. (2011). Tobii unveils the world's first eye-controlled laptop. Retrieved 14 March 2011 from http://www.tobii.com/en/ eye-tracking-integration/global/news-and-events/press-releases/tobii-unveils-the-worlds-first-eye-controlled-laptop/

Van Dam, A. (2001). Post-Wimp user interfaces: The human connection. In R. Earnshaw, R. Guedj, A. van Dam & J. Vince (Eds), *Frontiers of human-centred computing, online communities and virtual environments* (pp. 163–178). London: Springer-Verlag.

Velichkovsky, B. M., Sprenger, A., & Pomplun, M. (1997). Auf dem Weg zur Blickmaus: Die Beeinflussung der Fixationsdauer durch kognitive und kommunikative Aufgaben. In R. Liskowsky, B. M. Velichkowsky & W. Wünschmann (Eds), *Software-Ergonomie* (pp. 317–327)

Vertanen, K., & MacKay, D. J. C. (2010). Speech Dasher: Fast writing using speech and gaze. In *Proceedings of CHI 2010*, Atlanta, Georgia, United States of America, 595–598

Ward, D. J., Blackwell, A. F., & MacKay, D. J. C. (2000). Dasher—a data entry interface using continuous gestures and language models. In *Proceedings of UIST 2000: The 13th Annual ACM Symposium on User Interface Software and Technology*, San Diego, California, United States of America, 129–137

Wobbrock, J. O., Rubinstein, J., Sawyer, M. W., & Duchowski, A. T. (2008). Longitudinal evaluation of discrete consecutive gaze gestures for text entry. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA)*, Savannah, Georgia, United States of America, 11–18

Zhai, S., Morimoto, C., & Ihde, S. (1999). Manual And gaze input cascaded (MAGIC) pointing. In *Proceedings of CHI '99: ACM Conference on Human Factors in Computing Systems*, Pittsburgh, Pennsylvania, United States of America, 246–253

Zhang, X., & MacKenzie, I. S. (2007). Evaluating eye tracking with ISO 9241– Part 9. In J. Jacko (Ed.), *Human Computer Interaction*, 779–788