Context Switching for Fast Key Selection in Text Entry Applications

Carlos H. Morimoto*
University of São Paulo
Rua do Matão 1010
São Paulo, SP 05508-090, Brazil

Arnon Amir†
IBM Almaden Research Center
650 Harry Road
San Jose, California 95120, USA

Abstract

This paper presents context switching as an alternative to selection by dwell time. The technique trades screen space for comfort and speed. By replicating the interface on two separate regions called contexts, the user can comfortably explore the whole content of a context without the effects of the Midas touch problem. Focus within a context is set by a short dwell time and fast selection is done by switching contexts. We present experimental results for a text entry application with 7 participants that show significant speed improvement over traditional fixed dwell time gaze controlled keyboards. After 8 sessions, 6 participants were able to type about 12 words per minute (wpm), and the fastest participant was able to type above 20 wpm with error rate under 2%.

CR Categories: H.5.2 [Information Systems]: Information Interfaces and Presentation—User Interfaces H.1.2 [Information Systems]: Models and Principles—User/Machine Systems

Keywords: Gaze Typing, Context Switching, Gaze Interfaces

1 Introduction

Recent reviews of text entry methods using eye gaze tracking (known as gaze typing or eye typing) are given in [Majaranta and Räihä 2007; Majaranta 2009]. Virtual keyboards and eye gestures constitute the current leading interaction paradigms.

Text entry using virtual keyboards controlled by gaze interaction requires a great number of design issues to be considered, such as keyboard layout, key size, and particularly, how the keys are activated ("pressed"). The most common activation method is **dwell time**, where the gaze must be fixated at the desired key for a predefined time interval to activate it. If the gaze is moved to a different key within the interval, no key is activated. Using fixed dwell time (typically, 500-1000ms) it has been reported that typing is limited to 5-10 words per minute (wpm).

Recently, it has been shown in [Majaranta et al. 2009] that adjustable dwell time can considerably improve typing speed without affecting accuracy. After 10 training sessions, each 15 minutes long, previously inexperienced users were able to achieve typing speeds of about 20 wpm.

An alternative to virtual keyboards activated by dwell time are gaze gesture based text entry methods. Gestures can be composed of a sequence of discrete eye movements [Huckauf and Urbina 2007; Wobbrock et al. 2008] or continuous gestures [Hansen et al. 2008; Ward and MacKay 2002]. The use of gestures avoids some of the problems with virtual keyboards. For example, discrete gestures can use small screen regions or even locations off screen as targets.

The most prominent gaze based text entry technique to date is Dasher [Ward and MacKay 2002], due to its availability (a free download), support on multiple platforms and high performance. Typing is achieved by zooming into characters flowing across the screen as they cross a vertical line. Early experiments reported in [Ward and MacKay 2002] show that an expert user of Dasher can achieve 25 wpm using its word completion feature. In [Urbina and Huckauf 2007] several gaze and gesture controlled text entry methods are compared, including Dasher and a traditional QWERTY virtual keyboard. Their preliminary results showed that, without character prediction, the fastest Dasher user was able to type 7.4 wpm, compared to 15 wpm using the QWERTY keyboard with 500 ms dwell time. However in a more recent longitudinal study reported in [Tuisku et al. 2008] ten Dasher users achieve an average 17.5 wpm after ten sessions of 15 minutes training.

2 Typing by Context Switching

In this paper we propose *Context Switching* (CS) as a new activation mechanism for gaze controlled interfaces. The CS paradigm attributes key-focus and key-selection to two different eye movements. The interface is replicated on two separate screen regions, called *contexts* (see Figure 1). Key-focus is activated by a short fixation (that in practice corresponds to short dwell times). Key selection (i.e., typing) is made by switching contexts (a saccade to the other context). At selection, the key which was last in focus in the previous context is selected. Saccades within the same context and detected key-focuses with no subsequent context switching are ignored. Hence, the user can comfortably explore the whole content of a context without the effects of the Midas Touch problem. The user alternates between the two contexts as he types. As a result, in CS-based input the traditional long dwell time is replaced by a short dwell followed by a context-switching saccade.

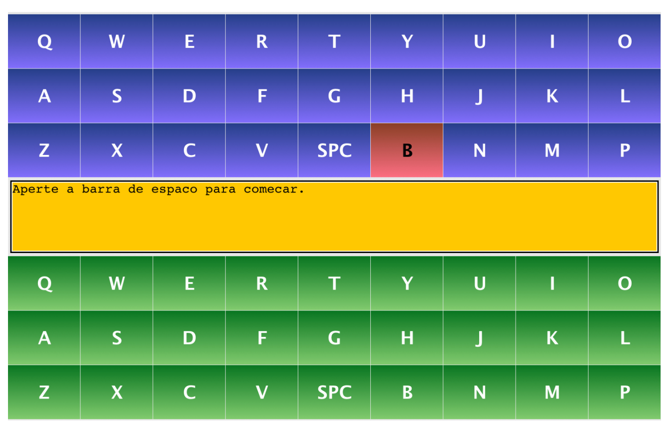


Figure 1: *The KKboard screen layout, utilizing the new context switching paradigm*

*e-mail: hitoshi@ime.usp.br
†e-mail: arnon@almaden.ibm.com

The use of discrete gestures to a predesignated button or to other screen locations for selection was suggested by [Ohno 1998; Isokoski 2000]. Their method also uses two different eye movements for focus and selection, however it requires two saccades for each selection — slower and more tiring. Instead, CS replicates the complete virtual keyboard in two separate contexts, thus a single saccade is required to switch contexts, trading space for speed. Theoretically, an experienced user may achieve continuous typing by focusing directly to the characters to be typed (while alternating contexts), thus avoiding any extra saccades.

The elimination of the Midas touch problem in CS is particularly helpful for apprentices that are not familiar with the keyboard layout and need time to search for keys. Here there is no need to adjust for longer dwell time intervals, and the user can naturally select his or her own typing speed.

Because two contexts (keyboards in this example) are required for CS, we call this layout *KKBoard* (key-key board). An obvious disadvantage of CS is the extra space needed to replicate the interface. Nonetheless, our experiments show that using KKBoard is very easy to learn, and people familiar with the QWERTY layout can quickly achieve entry rates faster than 10 wpm without character prediction.

3 KKBoard Design Issues

A low cost eye tracker was used in this study, based on the pupil center corneal reflection technique [Morimoto et al. 2000]. The accuracy of the tracker is about 1 degree in the central area dropping down to 2 degrees near the screen edges (compared to 0.5 degrees in commercial systems). Hence we had to use larger displayed keys, which limit the number of keys to three rows and nine columns, enough for only the 26 modern Latin alphabet ('A' to 'Z') and the Space key. Figure 1 shows the KKBoard layout used in our experiments. Low gaze tracking accuracy was the main limitation in our current experimentation design.

A short dwell time of 150 ms was used to detect key-focus. The focused key changes color (red in Figure 1) to provide feedback to the user. Selection is made if context switching is completed within 450 ms. The center part of the screen was reserved for typed text display. A saccade to the text display area does not invoke key selection. The font size at the text display area was selected to allow 5 lines of text to be displayed.

4 Experimental Procedure

The experimentation goal was to evaluate the performance and effectiveness of context switching typing with KKboard. The experimental procedure is similar to those in [Tuisku et al. 2008; Majaranta et al. 2009]. Six native Portuguese speakers, 5 men and 1 woman, ages from 23 to 46 years old, volunteered for the experiment. They were all able-bodied graduate students and researchers in our laboratory. Just one of them had some previous experience with eye trackers, and all of them were familiar with the QWERTY keyboard layout, with at least 8 years of typing experience on such keyboards. No participant had previous experience with any eye typing application.

During the first session, all participants received a brief explanation of the experiment and their task. They were instructed to type phrases as quickly and accurate as possible within a 10 minutes session. Because the current KKBoard layout has no backspace, they were also told to ignore all errors and continue typing.

A 300 Portuguese phrase set, adapted from the 500 English phrase set from [MacKenzie and Soukoreff 2003] was used in this experiment. The phrases are easy to remember everyday sentences. No punctuation, capital letters or accents were used. The users were asked to memorize each phrase before typing.

The space bar of a physical keyboard was used as a start/stop switch during the experiment and was controlled by the participant. After pressing the space bar for the first time, the first phrase was displayed on the KKBoard text area. Each participant would take a few seconds to read and memorize the phrase and then look at the top keyboard context to start typing. The clock was started when the first character was typed, i.e., when the first context switching from the top to the bottom keyboard was detected. The clock was stopped when the participant pressed the space bar again to indicate the end of typing that phrase. The time spent typing the phrase and the total elapsed time were then displayed to the participant.

The text display area was used to display the typed text as well as study-specific information. The 1st line was used to display the phrase to be typed. The 2nd line showed the text typed by the participant. The 3rd and The 4th lines showed the participant performance information and session elapsed time. The 5th line display instructions for the next step in the experiment.

Before each session, participants were asked to sit down at about 60 cm from the 17" monitor, and adjust the chin rest until they find a comfortable position. The 10 minutes eye typing sessions started soon after the calibration of the eye tracker. Participants were encouraged to calibrate themselves before each session, and were assisted by the study instructor when needed. In a few occasions calibration was lost during a session. In those cases the lost phrase was discarded and the system was re calibrated.

Before the very first session, users were allowed 5 minutes to practice with KKBoard, during which its basic functionality was explained, as well as a few tips to avoid common mistakes, such as fast context switching before key-focus and the need to completely cross the central text region. To motivate the participants, movie tickets were offered as a prize to the person who achieved the best performance during the trials.

All 8 sessions were recorded along one week. A participant could take several sessions in one day with at least 15 minutes interval between sessions. One of the participants were only able to complete 7 sessions during that week. In total, participants had gaze typed for 85 minutes including the initial 5 minute training session. After the last session, they were interviewed and answered a brief questionnaire about their experience with KKBoard.

5 Experimental Results and Discussion

We use error rate and text entry rate to evaluate the performance of KKBoard. The error rate (ER) based on the Minimum String Distance (MSD) proposed in [Soukoreff and MacKenzie 2001] is a common metric for quantifying text-entry error rates and is defined as:

$$\mathrm{ER}_{\mathrm{MSD}} = \frac{\mathrm{MSD}(A, B)}{max(|A|, |B|)} \tag{1}$$

where $|A|$ denotes the length of string A.

$\mathrm{ER}_{\mathrm{MSD}}$ represents the smallest proportion of characters considered errors between the strings A and B. The algorithm to compute the MSD is described in [Soukoreff and MacKenzie 2001].

Gaze typing speed is measured in words per minute (WPM) defined by the following equation [Wobbrock 2007]:

$$\mathrm{WPM} = \frac{|S| - 1}{5 \times T} \tag{2}$$

where $|S|$ corresponds to the length of the final transcribed string including spaces, and T is the time in minutes measured from the entry of the first character to the entry of the last. The constant 5 is adopted as the average number of characters per word, including spaces.

It is clear that errors made and corrected during text entry reduce WPM since a person may type faster when leaving more errors. Because no error correction was possible during the experiment, we also provide results using the Adjusted Words per Minute (AWPM) entry rate, as suggested in [Matias et al. 1996] . AWPM is computed as follows:

$$\mathrm{AWPM} = \mathrm{WPM} \times (1 - \mathrm{ER}_{\mathrm{MSD}}) \quad (3)$$

Figure 2 shows the results of the average $\mathrm{ER}_{\mathrm{MSD}}$ for each participant in each session. For most participants, the error rate drops consistently over the 8 sessions.

Figure 3a shows the results of the average text entry speed (WPM). The adjusted AWPM is shown in Figure 3b.

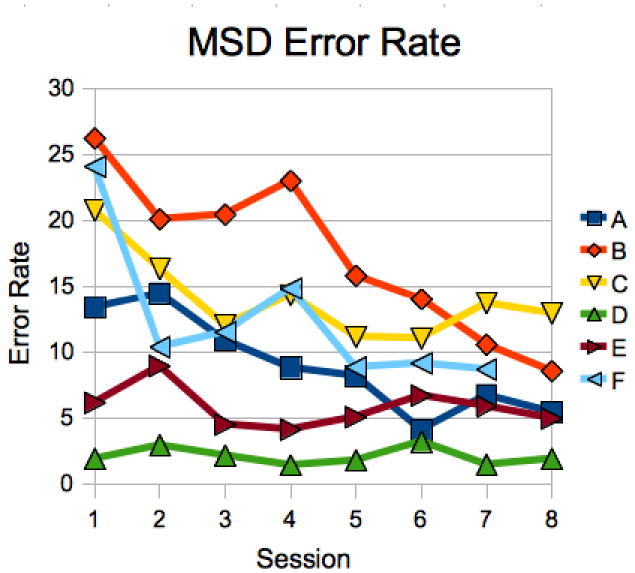


Figure 2: *Minimum String Distance error rate results.*

From Figures 2 and 3 it is noticeable that participant B has the largest error rate and also the lowest overall speed. This participant had problems calibrating and at times used to review what was typed in the central text display area. Due to poor calibration, when the user compared the given phrase with the typed one, several other keys were selected. The user also had problems with the small dwell time and at times, while transitioning between contexts, a different key was selected. All other participants, but participant D, were able to achieve the speed of about 12 wpm, with under 8% error rate. Participant E had previous eye tracking experience. His performance (error rate and speed) varies very little from the first to the last session.

Participant D was clearly much faster than all the other participants, and committed very few errors since the first session. After the final session, we asked D about D's strategy. D notice that the system responded very quickly to D's eye movements, and found out that it was not necessary to wait for the key to change color. D simply moved from one key to the next, waited a bit to focus on the desired key, and switched contexts. Participant C also commented trying similar strategies at the final sessions, but that resulted in higher error rates. All other participants waited for the key to change color, and then switched contexts, which might explain why they perform similarly at the last session.

All participants found the technique very easy to learn, and the use of the QWERTY layout help them reducing the search time for the next key. They mention that after the second session, it was already natural to switch contexts but, because no feedback was given about the last key typed, they had to look at the text area and sometimes got confused about which keyboard had their focus.

They were also asked about eye fatigue during the sessions. Using a Likert scale from 1 (not fatigued) to 5 (very fatigued), one of them report 1, three reported 2 (low fatigue), and the others reported 3 (medium fatigued). For all that reported medium fatigued, they said that at lower speeds they would feel more comfortable and probably not fatigued. Despite the fatigue, all subjects reported that they could type even faster and with less errors if they had more practice and a more accurate eye tracker. Indeed the learning curve in Figure 3 does not reach a plateau after eight sessions. This is similar to the learning curve for Dasher as shown in [Tuisku et al. 2008]. It suggests that better performance can be achieved with additional sessions. We plan to conduct more experiments in the future, with a more accurate EGT and a larger number of sessions.

6 Conclusion

We introduce *Context Switching* (CS) as a new activation mechanism for gaze controlled interfaces. By replicating interface components in two different contexts, short eye fixations can be used to control target focus and saccades between contexts to control selections. We demonstrated the CS paradigm in the design of KKBoard, a dual keyboard eye typing application. KKBoard eliminates the Midas Touch problem, allowing users to comfortably look at any key for any period of time. Because the keyboard is replicated in each context, only a single saccade is required, minimizing eye strain and saving time.

The use of a QWERTY layout facilitated quick learning. The fastest participant used a different strategy and was able to achieve over 20 wpm, with low error rate (under 2%). Compared to dwell time keyboards, KKBoard trades screen space for speed and comfort with favorable performance compared to fixed dwell time applications. The method naturally adjusts to typing speed, so it is possible to slow down or speed up typing without the use of an explicit button.

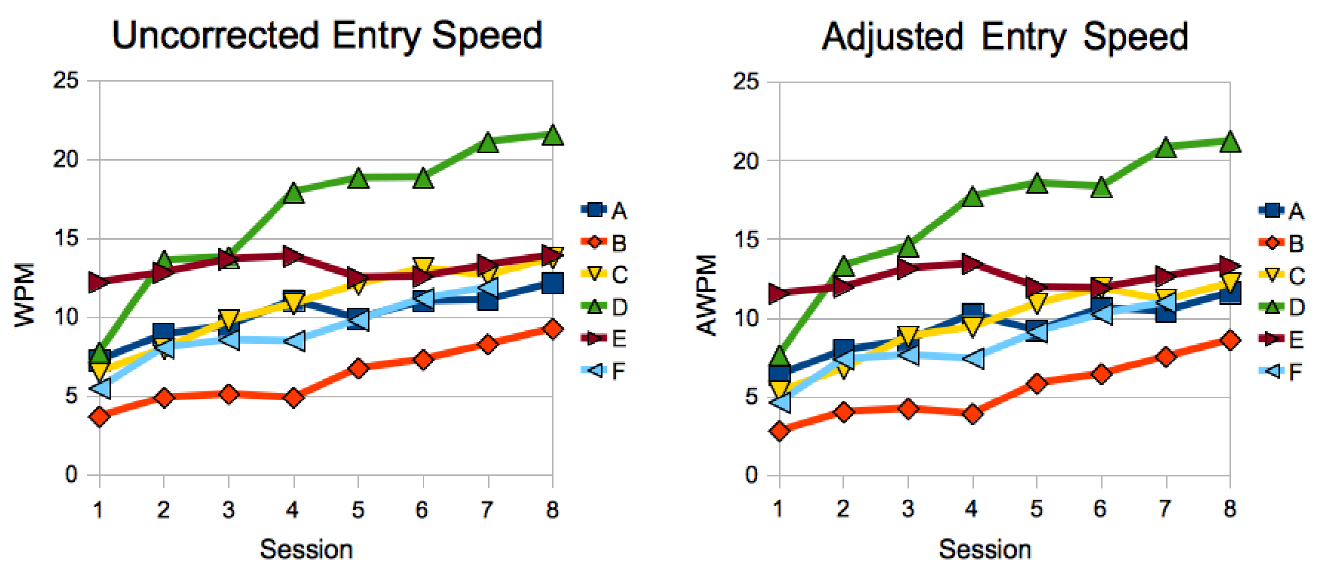
In future work we will test the CS paradigm with a more accurate commercial eye tracker, richer keyboard layouts, and with smaller interfaces, such as numeric keypads and menus.

Acknowledgements

We would like to thank Simone Hanazumi, Flavio Coutinho, and Filipe Campos for their help with coding KKBoard and translating the list of phrases used in the experiment. We also thank FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) for the financial support.

References

HANSEN, D., SKOVSGAARD, H., HANSEN, J., AND MOLLENBACH, E. 2008. Noise tolerant selection by gaze-controlled pan and zoom in 3d. In *Proceedings of Eye Tracking Research & Applications, ETRA 2008*, ACM Press, Savannah,GA, 205–212.



a

b

Figure 3: *a) Uncorrected, and b) adjusted speed text entry rate results in words per minute.*

HUCKAUF, A., AND URBINA, M. 2007. Gazing with peye: new concepts in eye typing. In *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization (APGV'07)*.

ISOKOSKI, P. 2000. Text input methods for eye trackers using off-screen targets. In *ETRA '00: Proceedings of the 2000 symposium on Eye tracking research & applications*, ACM, New York, NY, 15–21.

MACKENZIE, I. S., AND 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*, ACM Press, 754–755.

MACKENZIE, I. S., AND TANAKA-ISHII, K. 2007. *Text entry systems: mobility, accessibility, universality*. Morgan Kaufmann Academic Press.

MAJARANTA, P., AND RÄIHÄ, K. J. 2007. Text entry by gaze:utilizing eye tracking. In *Text entry systems: Mobility, accessibility, universality*, I. MacKenzie and K. Tanaka-Ishii, Eds. Morgan Kaufmann, ch. 9, 175–187.

MAJARANTA, P., AHOLA, U., AND SPAKOV, O. . 2009. Fast gaze typing with an adjustable dwell time. In *Procedings of the 27th international conference on human factors in computing systems - CHI´09*, 357–360.

MAJARANTA, P. 2009. *Text Entry by Eye Gaze*. PhD thesis, University of Tampere, Department of Computer Science, Tampere, Finland.

MATIAS, E., MACKENZIE, I., AND BUXTON, W. 1996. One-handed touch-typing on a qwerty keyboard. *Human-Computer Interaction 11*, 1–27.

MORIMOTO, C., KOONS, D., AMIR, A., AND FLICKNER, M. 2000. Pupil detection and tracking using multiple light sources. *Image and Vision Computing 18*, 4, 331–335.

OHNO, T. 1998. Features of eye gaze interface for selection tasks. In *Proc. of the Third Asia Pacific Computer Human Interaction-APCHI'98*, IEEE Computer Society, 1–6.

SOUKOREFF, R. W., AND MACKENZIE, I. S. 2001. Measuring errors in text entry tasks: An application of the levenshtein string distance statistic. In *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems - CHI 2001*, ACM Press, 319–320.

TUISKU, O., MAJARANTA, P., ISOKOSKI, P., AND RÄIHÄ, K.-J. 2008. Now dasher! dash away! longitudinal study of fast text entry by eye gaze. In *Proceedings of Eye Tracking Research & Applications, ETRA 2008*, ACM Press, Savannah,GA, 19–26.

URBINA, M., AND HUCKAUF, A. 2007. Dwell time free eye typing approaches. In *Proceedings of the 3rd Conference on Communication by Gaze Interaction (COGAIN 2007)*, Available at http://www.cogain.org/cogain2007/COGAIN2007Proceedings.pdf, 65–70.

WARD, D., AND MACKAY, D. 2002. Fast hands-free writing by gaze direction. *Nature 414*, 838.

WOBBROCK, J. O., RUBINSTEIN, J., SAWYER, M. W., AND DUCHOWSKI, A. T. 2008. Longitudinal evaluation of discrete consecutive gaze gesture for text entry. In *Proceedings of Eye Tracking Research & Applications, ETRA 2008*, ACM Press, Savannah,GA, 11–18.

WOBBROCK, J. O. 2007. Measures of text entry performance. In *Text entry systems: Mobility, accessibility, universality*, I. MacKenzie and K. Tanaka-Ishii, Eds. Morgan Kaufmann, ch. 3, 47–74.