AVIN (Assisted Visual Interactive Notepad): A Novel Interface Design to Expedite the Eye Writing Experience

Xianjun Sam Zheng[1], Stuart Goose[2], Joeri Kiekebosch[1], and James Jeng-Weei Lin[1]

[1] Siemens Corporate Research, Princeton, New Jersey, USA
[2] Siemens Corporate Research, Technology to Business Center, Berkeley, California, USA
`{sam.zheng,stuart.goose,joeri.kiekebosch.exe,`
`jameslin}@siemens.com`

Abstract. Eye typing, which utilizes eye gaze input to interact with computers, provides an indispensable means for people with severe disabilities to write, to talk, and to communicate. Despite more than two decades' of research into eye typing (which, for the most part, focused on the hardware/technical aspects associated with implementing a system), there lacks a well designed solution that has incorporated the key research findings and integrated them into a unified system. Hence, we designed and developed a novel user interface AVIN (Assisted Visual Interactive Notepad) for eye writing that expedites the writing workflow to enhance the overall user experience. Our preliminary user testing results showed that a novice user can achieve 7 wpm with an hour's typing practice, and a more experienced user can achieve 15 wpm with ten hours' practice; whereas an expert user can only reach 6-8 wpm using the standard QWERTY design.

Keywords: Eye typing, user interface (UI) design, eye tracking, user experience.

1 Introduction

Eye typing, which utilizes eye gaze input to interact with computers, provides an indispensable means for people with severe disabilities to write, to talk, and to communicate. Despite more than two decades' of research into eye typing (which, for the most part, focused on the hardware/technical aspects associated with implementing a system), there lacks a well designed solution that has incorporated the key research findings and integrated them into a unified system (Majaranta & Raiha, 2002, 2007). In this paper, we designed, prototyped, and tested a novel eye typing system, AVIN (Assisted Visual Interactive Notepad), to support the eye typing workflow and to enhance the overall eye writing user experience.

People direct and move their eyes to receive visual information from the environment. The two most typical eye movements are "fixation" and "saccade". Fixation is defined as the duration of time that the eye lingers at a location. In visual search or reading, the average fixation is about 200-500 milliseconds (ms). Saccade is defined as the rapid movement of eyes, lasting about 20-100 ms, with a velocity as high as 500 degree/sec (Fischer & Ramsperger, 1984).

C. Stephanidis (Ed.): Universal Access in HCI, Part II, HCII 2011, LNCS 6766, pp. 635–644, 2011.

It is natural to imagine using eye gaze as a computer input method for a variety of reasons. For example, research has shown that eye fixations are tightly coupled to an individual's focus of attention (Just & Carpenter, 1976; Rayner, 1998). Eye gaze input can potentially eliminate inefficiency associated with the use of an "indirect" input device (such as a computer mouse) that requires hand-eye coordination (e.g., looking at a target on a computer screen and then moving the mouse cursor to the target). Additionally, eye movements are much faster and require less effort than many traditional input methods, such as moving a mouse or joystick with your hand. Eye gaze input could be particularly beneficial for use with larger screen workspaces and/or virtual environments (Jacob, 1995). Last, perhaps the most important reason for considering and improving the utilization of eye gaze input, is that under some circumstances other control methods, such as using a hand or voice, might not be applicable. For example, for physically disabled people, their eyes may be the only available input channel for interacting with a computer.

Despite these benefits, eye gaze is not typically used as an input method for computer interaction. There remain critical design issues that need to be considered before eye gaze can be used an effective input method for eye typing.

2 Eye Typing Design

Even though eye typing has been studied for more than 20 years, previous research focused predominantly on the hardware design or technical design issues, such as how to achieve satisfying tracking accuracy and robustness, rather than on the user experience design aspects of the eye typing system (Majaranta & Raiha, 2002, 2007).

A typical eye typing system includes an eye tracking device and an on-screen keyboard interface (the graphical user interface, or GUI). The eye tracking device (referred to as an eye tracker) generally comprises an infrared camera, located near the computer, which monitors a user's eye movements. Typically, the device will track a user's point of gaze on the screen, and send this information to a computer application that analyzes the data and then determines the specific "key" of the on-screen keyboard at which the user is staring and wants to select. Thus, to start typing, a user will direct his/her gaze at the "key" of interest of the on-screen keyboard and confirm this selection by fixating on this key for a pre-determined time threshold (referred to as "dwell time").

In order to achieve the optimal eye typing performance and user experience, a number of design issues needed to be taken into consideration (Majaranta & Raiha, 2002).

2.1 Keyboard Layout Design

Most on-screen keyboards for eye typing utilize the standard QWERTY keyboard layout. While this layout is familiar to regular computer users, it may not be optimal for eye typing purpose. As some disabled users may not be adept at using a QWERTY keyboard in the first instance, modifying the keyboard layout to improve their user experience is considered to be a viable option.

Most contemporary eye typing systems are configured such that the on-screen keyboard occupies the majority of the central area of the screen. The typed content is displayed in a small region, typically above the on-screen keyboard along the upper part of the screen. This layout design does not consider a typical user's writing process. As illustrated in Figure 1, a typical writing process includes a first step of "thinking" about what to write, then selecting and typing a letter. After cycling through this process a number of times, a complete word is typed, and the process returns to think about the next word or words that need to be typed. Once the text is completed, the user will review and edit the typed content, the finally "finish" the typing process.

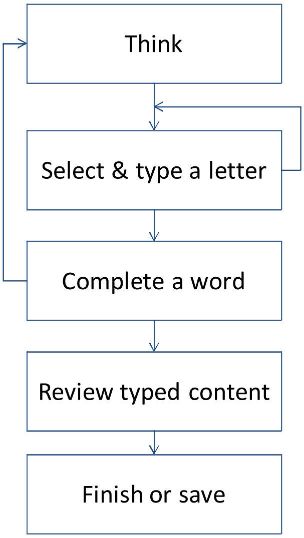

Fig. 1. A flowchart illustration of the conventional writing process

Traditional on-screen keyboard designs are configured to address the step of selecting and typing a letter, without considering the necessary support for other steps in the writing process, and/or the transitions between these steps. For instance, as the on-screen keyboard occupies the central area of the screen, it is difficult for the user to "think" about what to write next without unintentionally staring (gazing) at the keyboard. The user's eye gaze may accidently "select" a key, which must then be deleted before new letters are typed. Obviously, these tasks disrupt the natural flow of the thought process. Furthermore, the separation between the centrally-located on-screen keyboard and the 'text box' (generally in an upper corner of the screen) makes the transition to reviewing the typed content difficult, leading to eye fatigue on the part of the user.

2.2 Other Design Issues

Word prediction techniques have become popular in many text input methods. As a user starts typing a letter or letters, a list of words associated with the typed letter(s) will be suggested and presented in an order typically by their frequency of use. The user can either search the suggested list of words to find the target word they want to type or can continue to type the next letter, or letters, to complete the word. Research has shown the benefits of using the word prediction, such as reducing the number of keystrokes per character of text, and increasing the eye typing speed (e.g., Majaranta & Raiha, 2002).

As the QWERTY on-screen keyboard occupies the majority of the central area of the screen, this leaves limited screen real estate for the suggested word list. Usually only a small number of words are displayed, e.g. 5-8. As a result, users tend to type more keystrokes to reduce the risk of searching the word list, but without finding the target word. Increasing the size of the word list can reduce keystroke per word, but needs more screen real estate with the potential to increase the users' visual scanning time. There has not been a good design solution that balances these tradeoffs.

Other design factors, such as feedback and visualization, can also have effects on typing performance. Research (Majaranta, MacKenzie, Aula, & Raiha, 2006) has shown the positive effects of proper audio and visual feedback, which can improve both the eye typing performance and subjective experience. Effective visualizations can improve visual search performance. For example, studies show that people can perform subset search based on color, luminance etc. (e.g, D'Zmura,1991).

3 AVIN Design

We proposed and designed a novel user interface AVIN (Assisted Visual Interactive Notepad), a three-layer interface that allows for controlling computer input with eye gaze to expedite the writing process and typing experience (See Figure 2 for an example).

3.1 Novel Three-layer Layout to Support the Writing Workflow

It was our explicit design goal with this unique three-layer user interface to address the remaining limitations of conventional eye typing on-screen QWERTY keyboards, with the intended benefits of expediting the natural writing workflow and enhancing the overall user experience.

As described in detail below, the novel arrangement comprises a three-layer disposition of functionality – (1) letters, (2) words, and (3) typed text – that supports improved transitions between the various activities that occur during eye typing, as discussed above and shown in Fig. 1. The letters are selected from the outer ring, allowing for frequently-used words to be scanned in the inner ring, with the selected letter (word) displayed in the center text box.

As letters and words are arranged alphabetically, a natural spatial proximity between the letters and words is created, allowing for a more efficient visual search for a target word. As explained in more detail below, visual and audio feedback may be used to supplement the typing process, enhancing the overall eye typing experience.

Figure 2 is a screenshot of the three-layer interactive on-screen keyboard formed in accordance with the present design concept. A first layer, the *outer ring*, shows the standard 26-letter English alphabet, arranged alphabetically and moving clockwise from the upper left-hand corner. In this example, the letters "A", "V", "I", "N" form

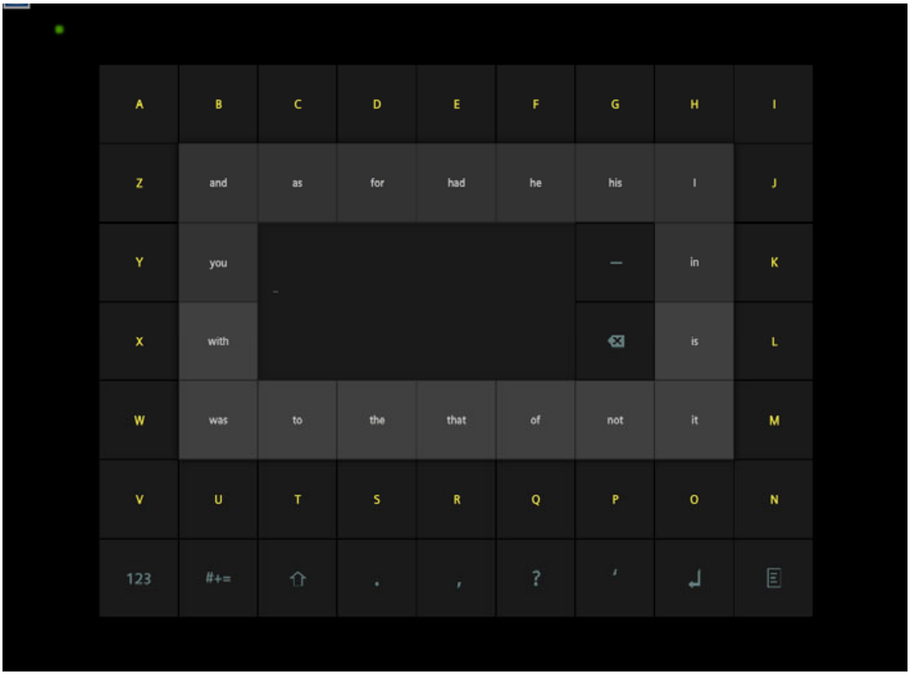

Fig. 2. A screen example of the AVIN embodying our three-layered design

the four corner letter, creating a rectangular "ring" structure. It is to be understood that in regions of the world where other alphabets are utilized, the keys would be modified to fit the alphabet (including the total number of alphabet/character keys included in the display).

The second tier of our on-screen keyboard, the *inner ring*, is a set of constantly-updated "frequently used" words. In this particular example, a group of eighteen words is displayed, again in the alphabetical order starting from the top, left-hand corner. The screenshot shown in Fig. 2 is an "initial" screen, before any typing has begun, thus a general set of frequently-used words is displayed. The use of eighteen terms is considered optimal, striking the balance to offer an abundance of word choices without being overwhelming. As elaborated upon below, the word list comprising inner ring is constantly updated; as letters are typed, the word set will be updated to reflect the actual letters being typed.

The third layer of on-screen keyboard comprises a central/inner region, which is the area where the typed letters will appear (referred to below as "text box"). A limited set of frequently-used function keys is included within inner region. In the specific example illustrated in Fig. 2 & 4, a "space" key and a "backspace" key are shown. By placing the typed content in the central area of the screen, the user may easily review the content and ponder what is to be typed next without fear of "accidently" inadvertently selecting a key by gazing at the screen for an extended period of time (as was the case for prior art on-screen keyboard arrangements).

Our on-screen keyboard comprises also a row of function keys, including a mode-switching functionality key (upper case vs. lower case), a numeric key, punctuation keys, etc. Again, the specific keys included in this row of function keys may be adapted for different situations. In the specific arrangement shown in Fig. 2, the row is positioned below the outer ring.

3.2 Visual and Audio Feedback to Confirm User Action or Command

Similar to previous eye typing arrangements, the current system uses dwell time to confirm a key selection. In AVIN, "dwell time" is visualized by using a running circle over the selected key. Fig. 3 illustrates this aspect of the present design concept, where the user has gazed at the letter "A". When the user fixates on this key, the circle will start (shown as circle on letter "A"). The user can easily cancel this action before the circle is completed by moving his/her gaze to another key before the circle is completed. Presuming in this case that the user desires to select the letter "A", the circle will run until completed, based upon a predetermined dwell time threshold. When the circle has completed, additional confirmation of the selection of this letter can be provided by the "A" block changing color (visual confirmation), and/or a "clicking" (i.e., audio confirmation) may be supplied. The selected letter will then "fly" to central region (text box) for display.

3.3 Visualizations and Animation to Facilitate Users' Visual Search

The addition of visual confirmation (such as color change) for a selected letter, with or without the utilization of an audio confirmation, is considered to enhance the user's experience, providing feedback and an affirmation to the user.

As shown in Fig. 4, the selection of the letters "uni" has caused the frequently-used words within inner ring to change, in this example, to frequently-used words beginning with the letter "uni". Again, the words are arranged alphabetically, starting from the upper left-hand corner. Thus, the user can quickly scan these words and see if any are appropriate for his/her use. Since the letters "uni" have already been typed, these are dimmed in the display of the frequently-used words. This feature can be further modified by using two different luminance contrast levels for the words, based on their absolute frequency of use. The leading letters in all the words that are redundant with the already-typed text may be "dimmed" to provide an additional visual aid.

Fig. 3. (a) The border for "E" is highlighted showing the eye-over effect, and (b) the circle shows the selection in progress (See text for details.)

Fig. 4. Dimming the typed letter(s)

Once a particular letter has been selected (in this example, “i”), a subset of letters predicted to be the next most likely to be typed are highlighted around the outer ring (or change in color – generally, made visually distinctive) to aid the user more rapidly and easily identify the next letter. Research has shown the positive effect of letter prediction on typing performance.

4 User Studies

4.1 Method – Part I

Participants. Six full time staff and student interns (5 men and 1 woman) from Siemens Corporate Research naïve to the goals of the study volunteered to participate. All participants had normal or corrected-to-normal vision, and were free of color blindness.

Implementation. For the purpose of prototyping and evaluation of the AVIN design, we utilized a SMI iView X Red eye tracker. The SMI tracker is a video-based contact-free eye tracker with the sampling rate of 50 Hz. The AVIN prototype was implemented using ActionScript 3.0, and can be run in Adobe Flash Player or Adobe Air environment. A 21-inch Dell monitor was used to display the AVIN interface with a resolution of 1280 x 1024 pixels.

Testing materials. The 500 phrases produced by MacKenzie & Soukoreff (2003) were selected as the testing materials.

Experiment design. There were two experiment conditions: the AVIN interface and the standard QWERTY interface, which was used as a control condition. A within-in subject design was used, where each participant was tested in both experiment conditions. The order of the experiment condition was counterbalanced across all the participants.

Testing procedure. Participants were tested individually. After reading an instruction sheet and providing demographic information, each participant was seated in front of the SMI eye tracker at a distance to the display of about 70 cm. A five points-calibration is needed before starting the eye typing application. The participants received a practice session, which required them to type letters A-Z twice. After the participants became familiarized with the keyboard layout, they would start practice writing sentences, and the audio was presented to them through the computer speakers.

When the participant was ready to listen to a sentence, he/she simply pressed the "Enter" key on the regular keyboard. A sentence was then randomly picked from the 500 phrases database (MacKenzie & Soukoreff, 2003) and read out to the user. The user can listen to the sentence as many times as he/she needs: the sentence will be repeated by itself if any key input is detected. Once the user started typing any key, then he/she would not be able listen to the sentence again.

For each experiment condition, there were two sessions, and each took about 30 minutes. Participants received breaks in between.

4.2 Results & Discussions – Part I

Fig. 5 shows that, after one hour of practice, the users can type on average 7.08 words/min using the AVIN design vs. 7.49 words/min using the QWERTY design. ANNOVA test show no difference between QWERTY and AVIN conditions (F (1,10)=0.29, p=0.60). This result is encouraging for two reasons. First, past research (e.g., Majaranta & Räihä, 2002) has shown that an expert user can only reach 6-8 wpm using the familiar QWERTY design, and using the AVIN design a novice user can achieve 7 wpm within only one hour of typing practice. Second, almost all users expressed that because they are familiar QWERTY layout, it is much quicker for them to learn to type. However, once they become more familiar with the AVIN design, their eye-typing performance can be improved significantly.

To test this hypothesis, a second part the study was conducted, where one user was invited back to do the eye-typing using the AVIN design for ten days (one hour per day).

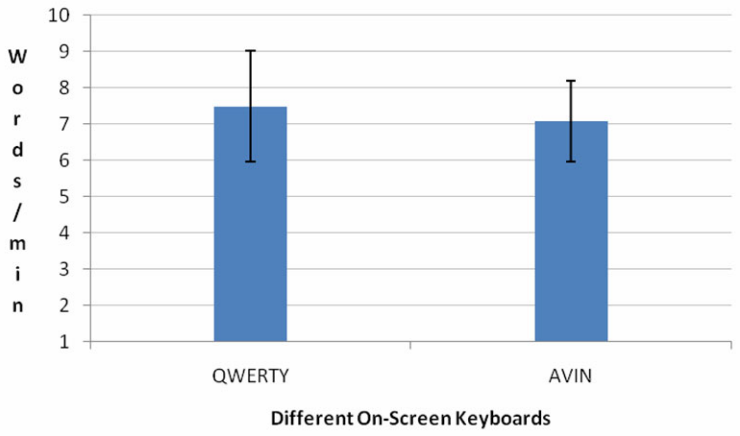

Fig. 5. Eye typing performance for different on-screen keyboard designs after one hour of practice

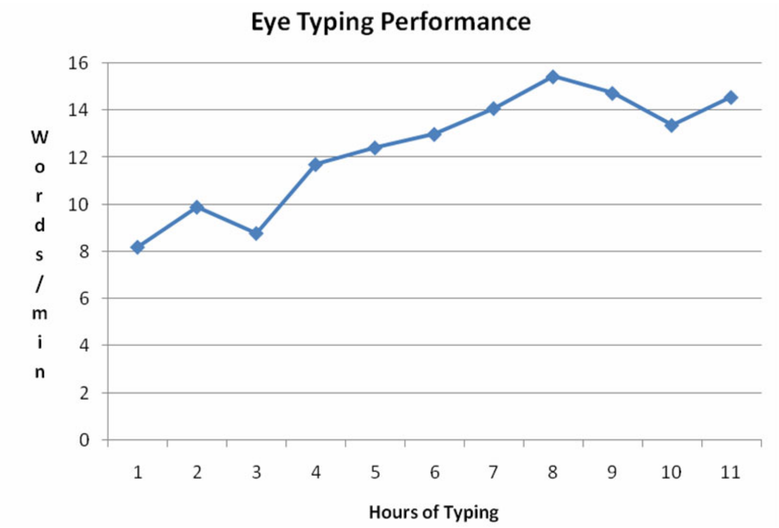

Fig. 6. Eye typing performance increasing with hours of practice

4.3 Method & Results – Part II

One of the previously tested six subjects volunteered to participate. The same materials were used as previous experiment. The participant was tested every weekday for an hour for 10 hours using the AVIN system.

Fig 6 shows that the eye typing performance increases almost linearly as the user spent more hours of practice using the AVIN system. After about 10 hours' practice, a more experienced user can achieve about 15 wpm.

5 Discussions and Conclusion

Reported in this paper was the design, prototype, and testing of a novel eye typing system, AVIN (Assisted Visual Interactive Notepad), to expedite the eye typing workflow and enhance the overall eye writing user experience. Our preliminary user testing results showed that a novice user can achieve 7 wpm after one hour of typing practice, and a more experienced user can achieve 15 wpm after ten hours of practice. In contrast, an expert user can only reach 6-8 wpm using the standard QWERTY design.

More user validation data, in particular the longitudinal data, are needed to prove the effectiveness of the AVIN design.

Acknowledgments. We thank Geoffrey Cooper, Dr. Yakup Genc, Dr. Hari Sundar for their feedback on earlier versions of the design. We thank Christine Wu and Jessica Downey for assisting with the experiment setup. Thanks also to Ren-Yi Lo for brainstorming the name, AVIN. Last, but not the least, we thank Dr. Chenyang Xu's support for this project.

References

1. Duchowski, A.T.: Eye Tracking Methodology. Theory and Practice. Springer, London (2003)
2. D'Zmura, M.: Color in visual search. Vision Research 31, 951–966 (1991)
3. Fischer, B., Ramsperger, E.: Human express saccades: extremely short reaction times of goal directed eye movements. Experiment Brain Research 57, 191–195 (1984)
4. Jacob, R.J.K.: Eye tracking in advanced interface design. In: Barfield, W., Furness, T.A. (eds.) Virtual Environments and Advanced Interface Design, pp. 258–288. Oxford University Press, New York (1995)
5. Just, M.A., Carpenter, P.A.: Eye fixations and cognitive processes. Cognitive Psychology 8, 441–480 (1976)
6. MacKenzie, I.S., Soukoreff, R.W.: Phrase sets for evaluating text entry techniques. In: MacKenzie, I.S., Soukoreff, R.W. (eds.) Extended Abstracts of the ACM Conference on Human Factors in Computing Systems – CHI 2003, pp. 754–755. ACM, New York (2003)
7. Majaranta, P., MacKenzie, I.S., Aula, A., Räihä, K.-J.: Effects of feedback and dwell time on eye typing speed and accuracy. Universal Access in the Information Society 5(2), 199–208 (2006)
8. Majaranta, P., Räihä, K.-J.: Twenty years of eye typing: Systems and design issues. In: Proceedings of the Symposium on Eye Tracking Research & Applications - ETRA 2002, pp. 15–22. ACM, New York (2002)
9. Majaranta, P., Räihä, K.-J.: Text entry by gaze: Utilizing eye-tracking. In: MacKenzie, I.S., Tanaka-Ishii, K. (eds.) Text Entry Systems: Mobility, Accessibility, Universality, pp. 175–187. Morgan Kaufmann, San Francisco (2007)