Effects of Feedback on Eye Typing with a Short Dwell Time

Päivi Majaranta, Anne Aula, and Kari-Jouko Räihä
Human-Computer Interaction Unit (TAUCHI)
Department of Computer Sciences
FIN-33014 University of Tampere, Finland
{Paivi.Majaranta, Anne.Aula, Kari-Jouko.Raiha}@cs.uta.fi

Abstract

Eye typing provides means of communication especially for people with severe disabilities. Recent research indicates that the type of feedback impacts typing speed, error rate, and the user's need to switch her gaze between the on-screen keyboard and the typed text field. The current study focuses on the issues of feedback when a short dwell time (450 ms vs. 900 ms in a previous study) is used. Results show that the findings obtained using longer dwell times only partly apply for shorter dwell times. For example, with a short dwell time, spoken feedback results in slower text entry speed and double entry errors. A short dwell time requires sharp and clear feedback that supports the typing rhythm.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces – Evaluation/methodology; Input devices and strategies

Keywords: eye typing, text entry, feedback, disabled users

1 Introduction

Eye typing refers to the process of producing text by using the focus of the gaze as a means of input. For people with severe disabilities, controlling the eyes can be the only means of communication. Although research concentrating on the technical issues of eye typing extends over twenty years, to date, there is very little research on the design issues of eye typing systems [Majaranta and Räihä 2002].

A typical eye typing setup has an eye tracker and an on-screen keyboard (see Figure 1). During eye typing, the user first locates the letter by moving her gaze (focus) on it. To type (select) the letter, she continues to look at it, thus using dwell time as a selection command. After a predefined dwell time, the key is selected (the letter is typed). Feedback is typically shown on both focus and selection.

Well known guidelines (*e.g.*, Microsoft Windows User Experience [2002]) suggest that continuous feedback should be used for continuous input (*e.g.*, moving a cursor, dragging an object), and discrete feedback for discrete input (*e.g.,* highlighting the selected object). In eye typing, the action is a combination of continuous and discrete input. The user controls a visible or invisible cursor by moving her gaze (continuous input). When dwell time is used as the activation command, the user fixates at the desired target and waits for the action to happen. The typing action itself is a discrete selection task. Since eye typing requires both continuous and discrete actions, choosing the proper feedback is an interesting design question.

Figure 1. On-screen keyboard and eye-tracking device.

The current experiment is a follow-up to a previous study [Majaranta et al. 2003] on the effects of different feedback modes in eye typing. In that study, the results suggested that auditory feedback, either a "click" sound or a spoken letter, is an effective indication of selection resulting in smaller error rates than visual feedback alone. Visual feedback combined with a short "click" sound yielded the fastest typing speed, and it was also the users' preferred choice. The visual feedback was shown in two parts: the key was first highlighted, and then, on selection, additional feedback was given. The dwell time used was quite long, 900 ms in total (400 ms before showing feedback for the focus, and 500 ms before selection after focus).

In this paper, we will study how feedback affects typing speed, accuracy, gaze behavior, and the user experience when shorter dwell times are used. We assume that if the dwell time is substantially shorter, the results of the previous experiment [Majaranta et al. 2003] may no longer fully apply. For example, with longer dwell times, the 2-level feedback (focus + selection) is beneficial by giving the user a possibility to cancel the action before the actual selection occurs. However, with shorter dwell times, there might not be enough time to clearly give information on both focus and selection. Furthermore, the good results with the spoken feedback alone (with no visual feedback) were surprising and require further study. In this experiment, we use a shorter dwell time (450 ms instead of the 900 ms in the previous study) and simpler visual feedback.

2 Related Work

The effects of auditory and visual feedback in graphical user interfaces (GUI) have interested researchers since the late 1980's [Gaver 1989]. We describe some of the most relevant results, starting with examples from graphical user interfaces, and proceeding to examples in gaze-aware systems. Finally, we discuss research on the use of dwell time as a selection command in eye typing systems.

Brewster and Crease [1999] suggested that interaction in conventional graphical user interfaces can be improved by using sound and graphics together. The results of their experiment showed that the usability of standard graphical menus can be improved by adding sound to them. Combined visual and auditory feedback significantly improved performance and reduced the subjective workload, compared to plain visual feedback.

Wolfson and Case [2000] found that in computer games, background color (red/blue) and sound (loud/quiet) not only affect the user's perceptions and physical reactions but also affect performance. Players using blue background gradually improved over time, while players with red screen peaked midway but then deteriorated. Bright colors and loud sounds are arousing and may temporarily improve performance but the effect decreases over time.

Non-speech sound has also been found to aid people who use so-called scanning as an input method (often people with severe disabilities, who cannot use mouse or keyboard). Scanning means that on-screen objects are highlighted one item at a time. For selecting an object, the user has to press a switch at the right time. Brewster et al. [1996] showed that added auditory feedback supported the scanning rhythm and helped the users in predicting the right time for pressing the switch for selection, thus improving the performance.

Seifert [2002] studied feedback in gaze interaction by comparing 1) continuously shown gaze cursor, 2) discrete feedback with highlighting the target under focus, and 3) no visible feedback for the gaze position. The task was to play a game where the user had to spot the target letter as soon as possible. Seifert found no differences between the gaze cursor and the highlight conditions in performance or in the perceived workload. However, the condition with no visible feedback caused significantly shorter reaction times, less false alarms and misses. Furthermore, the participants preferred the condition with no feedback and gave the worst ratings for the highlight condition. In Seifert's study, there were only three (large) letters on screen at a time. In eye typing where the on-screen targets are considerably smaller, the "no feedback" condition would require a very accurate eye tracker. For that reason, the "no feedback" condition was not considered in our study.

The fact that Seifert [2002] did not find any performance differences between cursor and highlight conditions is interesting, since previously it has been assumed that the constant movement of the gaze cursor distracts the user [Jacob 1993]. The distraction caused by movement is compounded by the problems caused by maintaining good calibration, which make the moving target gradually drift away from the focus of attention. This is also the reason why we did not show the cursor in the current study. The question of whether to show the cursor or not remains to be a subject for further studies.

Stampe and Reingold [1995] used a dwell time of 750 ms in their eye typing study. The dwell time was based on previous research and pilot tests. Typically, 1000 ms is a long enough dwell time duration to prevent false selections. For simpler tasks, 700 ms or less is enough. As Stampe and Reingold note, requiring the user fixate for a long time may be good for preventing false selections, but it is uncomfortable for the user.

In eye typing systems, people are usually asked to fixate long enough on a letter to select it. Typically, the fixations last from 200 to 600 ms [Jacob 1995]. As noted by Stampe and Reingold [1995], with long dwell times, a single fixation can not be used for selection, since gaze durations longer than 800 ms are often broken by blinks of corrective saccades. With long dwell times, it is better to use the total gaze duration spent on a region to indicate selection (in this case, the key on the on-screen keyboard).

3 Method

3.1 Feedback Modes

We carried out an experiment to study the effects of feedback modes with a short dwell time. Three different feedback modes were used:

Speech
The Speech mode does not use any visual feedback. The symbol on the key (a letter) is spoken on selection.

1-Level Visual
In the 1-Level Visual mode the background of the key briefly turns red on selection.

2-Level Visual
In the 2-Level Visual mode, the key is first highlighted on focus (the third row, second column in Figure 2). On selection the background of the key turns red.

Feedback mode	While focused	When selected
Speech	none h	letter spoken h
1-Level Visual	none h	red background h
2-Level Visual	highlight h	red background h

Figure 2. Tested feedback modes.

The dwell time for selection was the same, 450 ms, for all modes. 450 ms is exactly half of the time that was used in the previous experiment (900 ms). The dwell time for showing highlight on focus was 150 ms (used in 2-Level Visual only). The dwell time to re-select the current letter was increased by 120 ms to avoid erroneous double entries (*e.g.*, 'aa').

3.2 Participants

18 volunteers participated in the experiment. Data from three participants had to be discarded due to technical problems with the eye tracking device. In the final sample, there were 10 males and 5 females (mean age of 25 years, range 21-31 years). The participants were university students (undergraduate or graduate). All participants had normal or corrected-to-normal vision. All were familiar with PC/Windows™ and QWERTY keyboard layout.

All participants had previous experience in eye typing because they all had participated in an earlier eye typing experiment. We used experienced users because we wanted to test a shorter dwell time than in earlier experiments (short dwell times can be stressing for inexperienced users [Hansen et al. 2003]) and we wanted to compare the results with the results from earlier experiments (including the users' opinions).

3.3 Apparatus

The experimental setup consisted of two desktop computers and an eye tracking device. Eye movements were collected using SensoMotoric Instruments (SMI) iViex X RED II remote eye tracking device with 50 Hz sampling rate and 1-degree gaze position accuracy. The eye tracking device automatically compensates for (slow) head movements. The eye tracker device was placed in front of the corner of the monitor (see Figure 1).

One of the computers (Subject PC, with 17" flat LCD monitor, 1280*1024 resolution) was used to run the experiment and the other (Operator PC) to collect the eye movement data. The eye coordinate data was transferred in real time from Operator PC to Subject PC. The data were saved into three separate log files: 1) raw data and 2) fixation data from the eye tracking device, and 3) event data logged by the experimental software.

We did not exploit fixations in our software but we calculated the gaze position directly from filtered raw data points. Filtered points were already mapped to screen coordinates, and blinks and erroneous data had been removed. The selection of a key occurred if the user's (measured) point of gaze stayed inside the key area for the pre-defined dwell time.

The experimental software had an on-screen keyboard, a typed text field (above the keyboard), where the text written by the user appeared, and a source text field (below the keyboard). For the experiment, we added a special "Ready" key into the lower right corner of the screen (see Figure 3).

For the spoken feedback, we used a Finnish speech synthesizer Mikropuhe (v. 4.2. by Timehouse Oy). We used its default parameters for speech.

3.4 Procedure and Design

Before the experiment, the participants were told that the purpose of this experiment was to study eye typing. They were also told that this was a follow-up to the earlier experiments. The task was to eye type short, given phrases. We instructed the participant to first read and memorize the source phrase and then eye type it as fast and as accurately as possible.

In the beginning of the experiment, the participant was seated in front of the Subject PC's monitor so that his or her eyes were about 70-80 cm from the eye tracker. The eye tracker was then calibrated (and later re-calibrated before a set of phrases was shown).

The participant had a chance to practice eye typing by typing three short (given) phrases of text. The feedback used during practice was somewhat different from the feedback modes used in the experiment. The feedback in the practice phase used 2-Level Visual feedback that first highlighted the key to indicate focus, and on selection, gave a short "click" sound plus had the key visually going down (no change in the background color). The dwell time used in the practice mode was 200 ms before showing feedback for focus and 500 ms before selection.

During the experiment, the participant was presented by short, easy to remember phrases of text, one at a time. All the phrases were in Finnish which was the native language of all the participants. When the participant had typed the given phrase (presented in the source text field below the virtual keyboard), she looked at the Ready key that loaded the next phrase.

The participant had a possibility to correct errors by looking at a Del key. The Del key deleted the last letter from the typed text field (acting similarly to the Backspace key). We had told the participants to correct the errors immediately as they were noticed during typing. However, if the phrase was fully typed and an error was noticed while reviewing the typed text, participants were instructed not to correct the error. In the analysis, both corrected errors and errors left into the final text are taken into account.

After typing ten phrases, the participant had a chance to rest for a few minutes. After that, the experiment continued and the participant was presented with another set of ten phrases. There were three sets of phrases (and two pauses between them). The feedback mode was changed after each set of phrases. After the experiment, we interviewed the participant.

Figure 3 illustrates a gaze path of a participant eye typing a phrase during the experiment. The participant first read the source text, eye typed the phrase by fixating on the letters on the virtual keyboard, corrected an error, and finally looked at the Ready key in the low right corner.

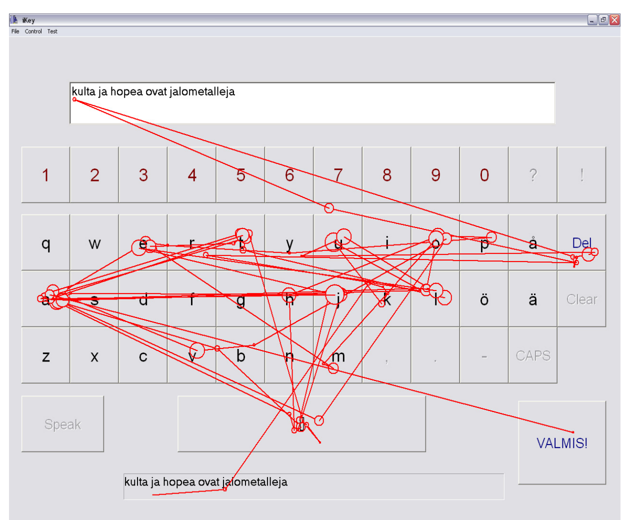

Figure 3. A gaze path of a participant eye typing a phrase.

The experiment was a repeated measures design, with three feedback modes: Speech, 1-Level Visual, and 2-Level Visual. The order of the feedback modes was counter-balanced; the participants were divided into three groups according to the feedback mode they started with. Each participant eye typed thirty phrases in total, ten with each feedback mode.

4 Results

The results are based on a total of 450 phrases (15 participants * 3 feedback modes * 10 phrases). We analyzed the typing speed, accuracy, gaze behavior, and the data from the interview. If, for some reason, the typing was interrupted (*e.g.,* due to a loss of calibration caused by a burst of cough), the experiment was briefly paused. Before typing resumed, the typed text field was cleared, and the participant had to re-start typing the current phrase. Analysis of a phrase started from the press of the first character and ended in the press of the Ready key (press here meaning a successful selection of the key by gaze).

The statistical analysis were done using repeated measures ANOVA and Bonferroni corrected t-tests.

4.1 Typing Speed

The grand mean for the text entry speed was 9.89 words per minute (wpm). A word is defined as 5 characters, including space [1]. The differences between participants were high; the typing rate varied considerably during the experiment, from below 7 wpm to over 14 wpm.

ANOVA showed a significant effect of feedback type on entry speed, $F_{2,28} = 6.54$, $p < .01$. Pairwise t-tests showed that the difference in text entry speed between Speech and 1-Level Visual was significant, $t = 2.72$, $df = 14$, $p < .05$ (see Figure 4). Similarly, the difference between Speech and 2-Level Visual was significant, $t = 2.87$, $df = 14$, $p < .05$. The difference between 1-Level Visual and 2-Level Visual was not significant. The Speech mode was significantly slower than either of the two visual feedback modes with the mean of 9.22 wpm. The means for the visual feedback modes were 10.17 wpm (1-Level Visual) and 10.27 wpm (2-Level Visual).

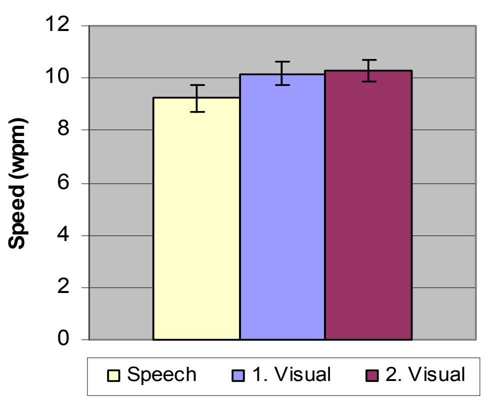

Figure 4. Text entry speed (wpm).
The error bars show the standard error of the mean (SEM).

[1] "In computing text entry throughput in 'words per minute' it is customary to consider a 'word' any sequence of 5 characters, including letters, spaces, punctuation, etc. " [MacKenzie 2003]

One possible reason for the slower typing speed with the Speech mode is revealed by inspecting the gaze paths. The participants spent time listening to the speech synthesizer speaking the letter, thus not leaving the key as soon as they could have. (The key was selected as soon as the dwell time had elapsed, and the dwell time for the next key started running instantly after the previous selection.) By studying the audio (wav) file recorded from the speech synthesis we found that it took typically at least 200 ms for the speech synthesizer to speak out the letter (*e.g.*, ~200 ms for 'a', and ~350 ms for 'm', with a soft fading in the end). Compared to the short (70 ms) red flash that was used for selection in the visual feedback modes, the spoken feedback was quite long.

By reversing the wpm measure back to search + dwell times, we get the average time spent to type a character in each feedback mode: 1300 ms for Speech, 1180 ms for 1-Level Visual, and 1170 ms for 2-Level Visual. The difference between Speech and 1-Level Visual is 120 ms, and 130 ms between Speech and 2-Level Visual. Since the difference is less than the time required by spoken feedback (typically more than 200 ms), the participants left the key before the spoken feedback ended. Nevertheless, the spoken feedback consumed more time. The extra time spent on listening to the spoken feedback also caused a decrease in accuracy, as discussed below.

4.2 Accuracy

The error rate was counted by comparing the transcribed text (text written by the participant) with the presented text (stimulus); it does not take into account the errors the users corrected. The grand mean for the error rate was 1.20%. The 1-Level Visual mode had the lowest error rate of 0.57%. Percentages for the other two were 1.36% (2-Level Visual) and 1.69% (Speech). ANOVA showed that the differences were not statistically significant ($F_{2,28} = 2.00$, $p > .05$).

Keystrokes per character (KSPC) [Soukoreff and MacKenzie 2001] measures the average number of keystrokes used to enter each character of text. Ideally, KSPC = 1.00 indicating that each key press produces one character of text. If participants correct mistakes during entry, KSPC will be greater than 1.00. For example, if in entering "hello", the user types h e l x [del] l o, the final result is correct (error rate is 0%), but KSPC is 7 / 5 = 1.4 since seven keystrokes were used to enter five characters. KSPC is a measure of the overhead incurred in correcting mistakes.

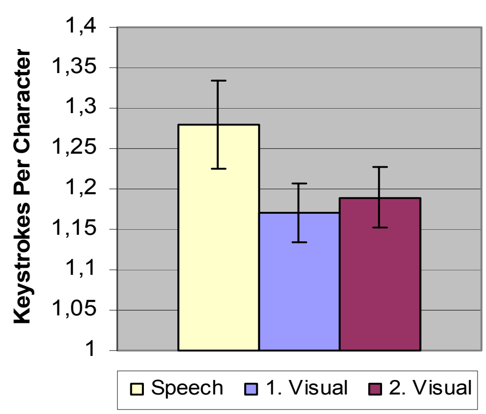

Figure 5. Keystrokes Per Character (and SEM).

The grand mean for KSPC was 1.21 meaning that there was about 21% overhead in keystrokes associated with correcting mistakes. Mean KSPC values for each feedback mode were 1.17 (1-Level Visual), 1.19 (2-Level Visual), and 1.28 (Speech). ANOVA showed that the effect of the feedback mode on KSPC was statistically significant ($F_{2,28} = 9.83$, $p < .005$). KSPC for Speech was significantly higher compared to 1-Level Visual ($t = 3.74$, $df = 14$, $p < .01$) and 2-Level Visual ($t = 3.05$, $df = 14$, $p < .05$, see Figure 5).

The higher KSPC for Speech is again due to the users' tendency of pausing to listen to the speech synthesizer. If the user spent more time on the key (listening to the spoken output) than the defined dwell time, the key was re-typed, causing an unintended "double-click".

A closer examination of the error types affirmed the problem of double entries. There were about three times more (corrected) double entry errors in the Speech mode (see Figure 6) than in the other two modes (the differences are statistically significant, $F_{2,28} = 19.12$, $p < .001$). Other kinds of errors were, for example, the user leaving the key before it was selected (missing character), or typing a wrong character (substitution).

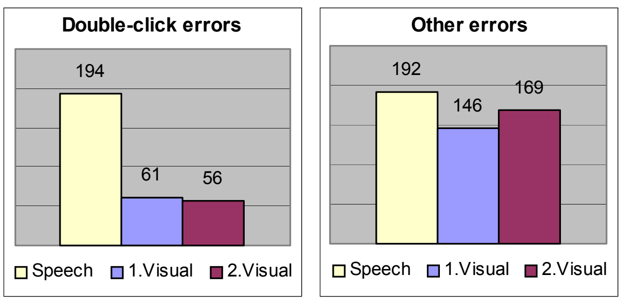

Figure 6. Double entry errors (left), and other errors.

We anticipated the double entry problem based on the pilot tests, therefore we increased the dwell time for repeated key press by 120 ms (for all feedback modes). Thus, the total dwell time for re-entering the same character was 450 + 120 = 570 ms. However, even though the pilot tests indicated that the 120 ms increase to the normal dwell time was adequate, the experiment proved it insufficient for many participants, especially with the Speech mode.

4.3 Gaze Behavior

In addition to typing speed and accuracy, we studied various aspects of gaze behavior. There were no significant differences in the number of fixations between the feedback modes. However, there were significant differences in the participants' gaze path behavior showing that the problems of using long feedback (Speech) with the rather short dwell time further cumulated in the gaze behavior.

Our experimental software logged various events of interest. One such event was "read text", referring to a participant switching her point of gaze from the virtual keyboard to the typed text field (above the virtual keyboard) to review the text written so far (see Figure 3). Typically, inexperienced users review the written text more than experienced users [Bates 2002] but the type of feedback may also have an effect, as discussed below.

For studying gaze path characteristics, the mean number of Read Text Events (RTE) was analyzed. Instead of reporting raw counts, RTE is normalized and reported on a "per character" basis. The ideal value is 0, implying participants were confident enough to proceed without verifying the transcribed text.

The feedback mode had a significant effect on the number of Read Text Events ($F_{2,28} = 4.50$, $p < .05$) with means of 0.139, 0.087, and 0.140 for the Speech, 1-Level Visual, and 2-Level Visual modes, respectively. The 1-Level Visual mode had significantly less Read Text Events than 2-Level Visual ($t = 2.92$, $df = 14$, $p < .05$). Due to greater variation in the Speech mode, the difference between Speech and 1-Level Visual was not significant (after the Bonferroni correction, $t = 2.70$, $df = 14$, $p < .1$). However, there was a trend towards Speech having more Read Text Events (see Figure 7).

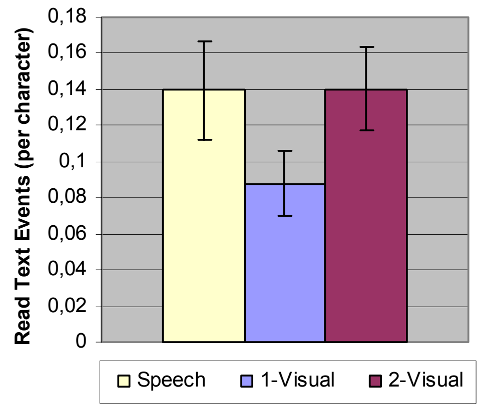

Figure 7. Number of Read Text Events per character (and SEM).

There were significantly fewer Read Text Events in the 1-Level Visual mode than in the Speech or 2-Level Visual modes. The increased need to review the typed text in the Speech mode is explained by the need to correct more errors. When correcting errors, the user deleted the last character (pressed Del), briefly glanced the typed text field to see if the deletion was successful, typed the right letter, and then again reviewed the typed text

The reason for the increased need to review the typed text in the 2-Level Visual mode is probably explained by the nature of the feedback. Since the visual feedback was shown in two parts (focus + selection), it may have caused extra confusion about whether the key was already selected or not. This assumption is affirmed by the comments from the participants (reported below).

4.4 Subjective Satisfaction

47% of the participants preferred the 1-Level Visual feedback, 33% liked the 2-Level Visual feedback best, and 20% preferred Speech. We asked the participants to give reasons for their preference. The participants appreciated the simplicity of the 1-Level Visual feedback. Some also felt that the 1-Level Visual feedback was the most fluent and pleasant mode to type with.

Participants who preferred the 2-Level Visual feedback appreciated the extra confidence given by the highlighting: "I instantly see what letter is going to be selected and can quickly adjust my gaze if necessary". They also found the short time between the focus and selection (450 – 150 = 300 ms) long enough to react, and to adjust the point of gaze.

Participants who liked the Speech feedback best wanted to hear what they write. They said it helped them to follow the typing and also helped in correcting errors. Many of the participants who otherwise preferred visual feedback commented that the spoken feedback helped in correcting errors. Many commented that with a little longer dwell time they might like the speech *combined* with visual feedback.

About half of the participants found the highlighting for focus in 2-Level Visual mode distracting and disturbing. They felt that it caused extra visual noise that made it hard to concentrate on the typing. The focus that “flashed” around the screen also made participants concerned that something might be selected accidentally.

One third of the participants thought that the extra highlight in 2-Level Visual did not give them any advantage over the 1-Level Visual feedback. For them, the 300 ms between focus and selection was not long enough to adjust the focus of gaze.

12 participants (80%) found the typing speed “just right”. No-one felt that the speed was too slow, but for 3 participants the speed was too fast. This only applies to the preferred choice of feedback. There were differences in how participants perceived the typing speed between the feedback modes. 5 participants commented that the typing felt too fast in the 2-Level Visual mode, much faster than in the 1-Level Visual mode (as mentioned earlier, the dwell time for selection was the same for all modes). The highlight probably caused extra stress, as one participant noted. Similarly, many commented that the speech felt too fast, and for that reason caused a lot of errors. (Eight participants preferred the mode that was actually their fastest mode.)

In addition to our questions, participants also gave other opinions. The location of the Del key, and the red color used in the visual feedback modes were noted by several participants. The location of the Del key was not optimal. In the current setup, the Del key was located in the top right corner of the virtual keyboard (similar to the location of the Backspace key in standard keyboards). Participants, however, commented that the Del key should be located as near as possible to the typed text field, since they often need to check the text during or after deleting characters. Our observations during the experiment also point out the need to somehow help the users in correcting text: Typically, the participants checked the typed text field after every correction made. This is a subject for further studies.

The red color indicating the selection annoyed some participants. They commented that red color is too strong compared to the light gray background. One participant also commented that red means denial or warning, thus a more neutral color would be better. The possibility to change the color is not only important because of the user’s preference but it may actually affect performance [Wolfson and Case 2000].

5 Discussion

This study showed that in typing speed, the speech feedback was significantly slower than either of the two visual feedback modes. The analysis of the error rates and types of errors showed that the main problem with speech feedback was that it caused participants to do unintended double entries. The 2-Level Visual feedback was found confusing, even though the measured performance was reasonably good. With a short dwell time the simple 1-Level Visual feedback yielded best results.

Compared to the earlier experiment [Majaranta et al. 2003], in the current study the typing speed was faster in all modes (obviously, since shorter dwell time was used), and the overall error rates were higher. The decrease in accuracy is no surprise since there always is a trade off between speed and accuracy in text entry tasks. However, with the spoken feedback the accuracy decreased considerably more than with the visual modes.

The duration of the spoken feedback is a problem when short dwell times are used in eye typing. However, the problems with the spoken feedback in this study do not mean that it could not work under different conditions. The length of the spoken feedback did not cause any problems with longer dwell time durations [Majaranta et al. 2003]. On the contrary, the Speech Only (with no visual feedback) or combined Visual + Speech produced fewer errors than visual feedback alone. Therefore, we assume that by adjusting the properties of the speech synthesis, better results could be achieved. The reason for not doing that in the current experiment is simply because we realized how many problems it caused only after the data from this experiment was analyzed. One possibility for adjusting the speech would be to make the speech synthesizer speak faster and sharper with no soft fading.

In addition to the length of the spoken feedback, a problem with speech is that it cannot achieve the sharpness and clearness of the very short red flash used in the visual feedback. Thus, it may not be clear for the users, if the selection is made as soon as the speech synthesizer activates, or only after the letter is spoken.

Methods that work well with a conventional mouse may require extra careful design with an “eye mouse”. The problem with the exact point of selection also arose from the programming point of view when the feedback of a standard click event was used as feedback (during practice). A click consists of two events: key down and key up. The click event does not happen if the focus is moved away while the key is held down. That causes annoying errors if the user has “clicked” the key by her gaze but the selection is cancelled because the gaze moves away too fast (during the very short time to show the key visually going down). This error type was avoided by using the key down event as a trigger. Anyhow, this emphasizes the need to define a distinct point where the selection happens, and making sure the user behavior is in accordance with it.

Typing rhythm is another issue worth considering. Typing is a series of actions, including the search for, and the selection of the key. It takes time both to search for the next key, and also for the dwell time to elapse. When the same key is double-clicked the typing rhythm is broken because the search time is not included. This was noted by a couple of participants. They would have wanted an even longer dwell time for the key repeat than what was used in the experiment (the normal 450 ms plus the added 120 ms). Furthermore, a short “click” sound would have better supported the typing rhythm than visual feedback alone. Non-speech auditory feedback have been found to support temporal tasks (with rhythm) better than visual feedback alone [Brewster et al. 1996].

The feedback itself can also affect the typing rhythm. In Speech mode, the duration of the spoken feedback varied from about 200 ms to 350 ms, depending on the spoken letter (*e.g.*, ‘a’ takes considerably less time to speak than ‘m’). As discussed above, the point of selection should be clear and distinct. The selection should happen immediately after the defined dwell time has elapsed, allowing the user to instantly proceed. In other words, the user should not be forced to wait for the feedback to finish.

Whether the duration of the dwell time should be adaptive (not constant on every character) is yet another interesting question. We added 120 ms for the dwell time if the user continued focusing on the selected key to prevent false double entries. The space key might be another special case to consider. In our experiments, we have seen that especially novices tend to either forget the space altogether or only briefly glance the space and proceed to the next word. One explanation could be that people think in words and type words—space is something extra. Perhaps the dwell time for the space key should be shorter. However, if an adaptive (automatically adjusted) dwell time is used, the adaptation should not interfere with the typing rhythm. Simpson and Koester [1999] studied adaptive scanning in an alternative communication system. They discovered that the automatic adaptation increased errors, because the users had developed a scanning rhythm, which the automatic adaptation interfered with.

Most of the participants would have liked to hear a simple "click" sound. They commented that the very short red flash did not seem to be enough. The "click" was left out of this study to simplify the experimental setup. However, we assume that a "click" sound would have helped to improve typing performance. As discussed earlier in this paper, research has shown that adding non-speech auditory feedback is an effective way of improving the visual feedback and we found no reason not to believe this. Thus, we still recommend adding a "click" that confirms the selection together with visual feedback.

Many participants felt that visual feedback is very important and that spoken feedback alone is not sufficient. As a couple of participants commented, it was sometimes hard to discriminate some letters by the spoken feedback alone. For example, 'n' and 'm' sound quite similar and they are also located next to each other in the QWERTY keyboard layout. Added visual feedback would have confirmed the selection.

We were a bit surprised that the 2-Level Visual feedback did not cause more problems and increase in error rates. As demonstrated by this experiment, for many users, 300 ms is a long enough time to react (by gaze); participants actually corrected their point of gaze during the short time interval between the focus (at 150 ms) and the selection (at 450 ms). Thus, dwell times as short as 300 ms are possible. As commented by the participants (who had participated in an earlier experiment with a longer dwell time) "faster is better".

6 Conclusions

The results show that with short dwell times, it is essential that the feedback is sharp and clear. Thus, brief feedback should be used with short dwell times: the feedback should not consume the dwell time needed to select the next character. Furthermore, there should be a distinct point where the selection is made – both visual and auditory feedback take time. There should be no uncertainty of the exact moment when the selection is done. As demonstrated by the results of the experiment, that may be hard to achieve with spoken feedback alone, or with a feedback that has several phases (*e.g.*, separated focus and selection) that cause extra confusion.

Acknowledgements

We would like to thank all volunteers who participated in the experiment. We would also like to thank Scott MacKenzie, Mika Käki, Saila Ovaska, and Harri Siirtola for consultation and inspiring discussions. The work of the first two authors is funded by the Graduate School in User-Centered Information Technology (UCIT).

References

BATES, R. 2002. Have patience with your eye mouse! Eye-gaze interaction with computers can work. In *Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT)*, Trinity Hall, University of Cambridge, 2002. Available at http://www.cse.dmu.ac.uk/~rbates/research/cambseyemouse.htm (24.11.2003).

BREWSTER, S.A. RATY, V.-P. AND KORTEKANGAS, A. 1996. Enhancing scanning input with non-speech sounds. In *Proceedings of ACM ASSETS'96* (Vancouver, Canada), ACM Press, 10-14. Available at http://www.dcs.gla.ac.uk/~stephen/papers/Assets96.PDF (24.11.3003).

BREWSTER, S.A. AND CREASE, M.G. 1999. Correcting Menu Usability Problems With Sound. In *Behaviour and Information Technology* 18(3), 165-177. Available at http://www.dcs.gla.ac.uk/~stephen/papers/BIT99.pdf (24.11.2003).

GAVER, W. W. 1989. The SonicFinder: An Interface that Uses Auditory Icons. In *Human-Computer Interaction* 4(1), 67-94.

HANSEN, J. P., JOHANSEN, A. S., HANSEN, D. W., ITOH, K., AND MASHINO, S. 2003. Command without a click: Dwell time typing by mouse and gaze selections. In *Proceedings of INTERACT 2003*, IOS Press, 121-128.

JACOB, R.J.K. 1993. Eye Movement-Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces. In H.R. Hartson and D. Hix (Eds.) *Advances in Human-Computer Interaction*, Vol. 4, Ablex Publishing Co., Norwood, N.J. 151-190. Available at http://www.cs.tufts.edu/~jacob/papers/hartson.pdf (24.11.2003).

JACOB, R. J. K. 1995. Eye Tracking in Advanced Interface Design. In W. Barfield and T.A. Furness (Eds.), *Virtual Environments and Advanced Interface Design*, Oxford University Press, New York, 258-288. Available at http://www.cs.tufts.edu/~jacob/papers/barfield.html (24.11.2003)

MACKENZIE, I. S. 2003. Motor behaviour models for human-computer interaction. In J. M. Carroll (Ed.) In *Toward a multidisciplinary science of human-computer interaction*, San Francisco: Morgan Kaufmann, 27-54. Available at http://www.yorku.ca/mack/carroll.html (24.112003).

MAJARANTA, P., AND RÄIHÄ, K.-J. 2002. Twenty years of eye typing: systems and design issues. In *Proceedings of ETRA 2002,* New Orleans, LA, ACM Press, 15-22. Available at http://www.cs.uta.fi/~curly/publications/ETRA2002-Majaranta.pdf (24.11.2003).

MAJARANTA, P., MACKENZIE, I. S., AULA, A., AND RÄIHÄ, K.-J. 2003. Auditory and visual feedback during eye typing. In *Proceedings of CHI 2003,* Extended Abstracts of the ACM Conference on Human Factors in Computing Systems, ACM Press, 766-767. Available at http://www.cs.uta.fi/~curly/publications/CHI2003-Majaranta.pdf (24.11.2003)

MICROSOFT WINDOWS USER EXPERIENCE 2002. Official Guidelines for User INTERFACE Developers and Designers. Microsoft Corporation. Available at http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnwue/html/welcome.asp (24.11.2003).

SEIFERT, K. 2002. *Evaluation Multimodaler Computer-Systeme in Frühen Entwicklungsphasen.* (in German) PhD thesis, Department of Human-Machine Systems, Technical University Berlin. Available at http://edocs.tu-berlin.de/diss/2002/seifert_katharina.pdf. Summation of the results involving gaze interaction (in English) available at http://www.roetting.de/eyes-tea/history/021017/seifert.html (24.11.2003).

SIMPSON, R. AND KOESTER, H. 1999. Adaptive One-Switch Row-Column Scanning. In *IEEE Transactions on Rehabilitation Engineering, 7(4)*, 464 - 473.

SOUKOREFF, R. W., AND MACKENZIE, I. S. 2001. Measuring errors in text entry tasks: An application of the Levenshtein string distance statistic. In *Proceedings of CHI 2001*, Extended Abstracts of the ACM Conference on Human Factors in Computing Systems, New York: ACM, 319-320. Available at http://www.yorku.ca/mack/CHI01a.htm (24.11.2003).

STAMPE, D. M. AND REINGOLD, E. M. 1995. Selection by looking: A novel computer interface and its application to psychological research. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Eye movement research: Mechanisms, processes and applications.* Amsterdam: Elsevier Science Publishers, 467-478. Available at http://www.psych.utoronto.ca/~reingold/publications/Stampe_&_Reingold_1995/ (24.11.2003).

WOLFSON S. AND CASE G. 2000. The effects of sound and colour on responses to a computer game. In *Interacting with Computers*, 13 (2), 183-192.