

Päivi Majaranta · I. Scott MacKenzie
Anne Aula · Kari-Jouko Rähä

Effects of feedback and dwell time on eye typing speed and accuracy

Published online: 13 July 2006
© Springer-Verlag 2006

Abstract Eye typing provides a means of communication that is especially useful for people with disabilities. However, most related research addresses technical issues in eye typing systems, and largely ignores design issues. This paper reports experiments studying the impact of auditory and visual feedback on user performance and experience. Results show that feedback impacts typing speed, accuracy, gaze behavior, and subjective experience. Also, the feedback should be matched with the dwell time. Short dwell times require simplified feedback to support the typing rhythm, whereas long dwell times allow extra information on the eye typing process. Both short and long dwell times benefit from combined visual and auditory feedback. Six guidelines for designing feedback for gaze-based text entry are provided.

Keywords Eye typing · Text entry · Feedback modalities · People with disabilities

1 Introduction

Eye typing refers to the production of text using the focus of the eye (aka gaze) as a means of input. It is especially needed by people with severe disabilities, where controlling the eyes is sometimes the only means of interaction with the world. Their need for an effective communication system is acute.

Research on technical aspects of eye typing extends over 20 years. However, there is little research on design

issues [16]. The authors' work [17, 18, 19] is an attempt to partly fill this gap. Such work investigated how feedback can facilitate the tedious [5] eye typing task. This paper summarizes the results of three experiments studying various aspects of feedback during eye typing.

Because gaze has some unique features, the paper briefly introduces gaze as an input method. This is followed by a description of eye typing. Examples of the most relevant research on eye gaze feedback are reviewed. The methods and results of the experiments are then presented, followed by guidelines gleaned from the results.

1.1 Gaze input

Gaze is naturally used to obtain visual information. For example, gaze location shows the focus of attention [13]. As an input method, gaze has both advantages and disadvantages. It is a natural mode of input as it is easy to focus on items by looking at them [12, 25]. Another advantage is that target acquisition using gaze is very fast, provided the targets are sufficiently large [26].

However, gaze is not as accurate as the mouse. Inaccuracy originates partly from technological reasons and partly from features of the eye [12]. The size of the fovea and the inability of the remote camera to resolve the fovea position restricts the accuracy of the measured point of gaze to about 0.5° , equivalent to a region spanning approximately 15 pixels on a typical display (17" display with a resolution of $1,024 \times 768$ pixels viewed from a distance of 70 cm). One problem is drifting; even if a newly calibrated eye tracking device is accurate at first, with continued use the measured point of gaze drifts away from the actual point of gaze. This is partly due to the technology and partly due to the interaction between head movement and eye movement. Therefore, the practical accuracy with about $0.5\text{--}1^\circ$ of drifting corresponds to about 1–1.5 cm on the screen at a normal viewing distance.

P. Majaranta (✉) · A. Aula · K.-J. Rähä
Unit for Human–Computer Interaction (TAUCHI),
Department of Computer Sciences, University of Tampere,
Tampere 33014, Finland
E-mail: Paivi.Majaranta@cs.uta.fi
Tel.: +358-3-35518550

I. S. MacKenzie
Department of Computer Science and Engineering,
York University, Toronto, ON, Canada M3J 1P3

When we look at things, we fixate (focus) on them, with fixations typically lasting from 200 to 600 ms [12]. For a computer to distinguish whether the user is looking at an object to obtain information or to select it, an interval longer than the typical fixation interval is needed. Stampe and Reingold [25] used a dwell time (an extended look at the object) of 750 ms in their eye typing study. A thousand milliseconds is usually long enough to prevent false selections. For simple tasks, 700 ms or less is enough. Requiring the user to fixate for long intervals is good for preventing false selections (thus, preventing the so-called Midas touch problem [11]), but this is uncomfortable for most users [25].

It is important to note that the use of a dwell time criterion for key selection places an upper limit on eye typing speed. In other words, no amount of skill acquisition will allow a user to “eye press keys” at a rate faster than $1/t_d$, where t_d is the dwell time. If, for example, $t_d = 1,000 \text{ ms} = 1 \text{ s}$, the upper limit for typing speed is $(60/1)/5 = 12$ words per minute (wpm) (following the accepted method in computing typing speed of 1 word = 5 inputted characters).

1.2 Typing by gaze

A typical setup has an eye tracker and an on-screen keyboard (see Fig. 1). The tracking device follows the user’s point of gaze and software records and analyses the gaze behavior. Based on the analysis, the system decides which letter the user wants to type.



Fig. 1 On-screen keyboard and eye tracking device

During eye typing, the user first locates the letter by moving his or her gaze (focus) to it. When using dwell time for selection, the user continues to look at the letter for a pre-defined time interval. Feedback is typically shown on both focus and selection (for a review, see [16]).

2 Previous research

It is known that interaction in conventional graphical user interfaces is enhanced by adding sound [6]; e.g., the beep used with Windows’ warning dialogs. According to Brewster and Crease [4], the usability of standard graphical menus is improved by adding sound. In particular, combining visual and auditory feedback is claimed to improve performance and reduce subjective workload, compared to visual feedback alone.

Non-speech sound also supports “scanning” as an input method (often used by people with severe disabilities). Scanning means that on-screen objects are highlighted one at a time. To select an object, the user presses a switch. Brewster et al. [3] showed that auditory feedback supports the scanning rhythm, helping users anticipate the correct time to press a switch for selection.

Animation is another way to enhance visual feedback, with progress bars as a typical example. Animation can, for example, help to clarify the meaning and purpose of an icon [1]. Furthermore, a shift between two conditions is easier to understand if the change is animated. For example, in Cone Trees by Robertson et al. [22] changes in 3D trees are animated (e.g., rotation, zooming). Animation allows the perceptual system to track transformations in the perspective.

User interface guidelines (e.g., [20]) suggest that continuous feedback should be used for continuous input (e.g., moving a cursor), and discrete feedback for discrete input (e.g., highlighting a selected object). In eye typing, the user controls a visible or invisible cursor by moving her gaze. This is an example of continuous input. Using dwell time as the activation command, the user fixates at the desired target and waits for the action to happen. This is an example of discrete input. Since eye typing requires both continuous and discrete actions, choosing the proper feedback is a relevant design issue.

Seifert [23] studied feedback in gaze interaction by comparing (1) continuous feedback using a gaze cursor, (2) discrete feedback by highlighting the target under focus, and (3) no feedback for the gaze position. Seifert found no differences in performance between the gaze cursor and the highlight conditions. However, the condition with no visible feedback caused significantly shorter reaction times, fewer false alarms and fewer misses. In Seifert’s study, there were only three (large) letters displayed at a time. In eye typing, using the QWERTY keyboard layout with considerably smaller on-screen targets, having no feedback is not a realistic option, since it requires a very accurate eye tracker.

There are eye typing systems that display only a few large keys and do not require such accuracy (see [8] for an example). They often use intelligent word prediction methods. In this study, the QWERTY keyboard layout and no word prediction have been used in order to keep the experimental setup as simple as possible.

That Seifert found no performance differences between the (continuous) cursor and the (discrete) highlight condition is surprising, since previously it was assumed that the constant movement of a gaze cursor distracts the user [12]. The distraction is compounded by problems with calibration, which cause the cursor to gradually drift away from the focus of attention [8, 12]. Therefore, it was decided not to show the cursor in the current study. The question of whether showing the cursor is beneficial remains a subject for further study.

Animation is also exploited in gaze-aware systems. For example, EyeCons developed at Risø National Research Laboratory [7] shows an animation of a closing eye to indicate dwell time progress. However, EyeCons are not convenient in eye typing, since they require the user to concentrate on the EyeCon icon instead of the target letter.

The ERICA system [10] uses animation in eye typing by showing a shrinking rectangle to indicate the progress of selection [14]. First, a key is highlighted by drawing a rectangle around the symbol on the key. After indicating focus, the rectangle starts to shrink. The key is selected at the end of the shrinking process. The research presented here includes a modification of this approach. Instead of a rectangle, a shrinking letter is used—the letter on the key. By shrinking the symbol itself, the feedback is further simplified. Since motion is an effective pre-attentive feature of vision to guide attention [9], it can be hypothesized that shrinking letter draws the attention toward the center of the key.

3 Methods and metrics

Three experiments were conducted to study the effects of feedback on eye typing speed, accuracy, gaze behavior, and user experience. Feedback was varied in each of the experiments. This section starts with a brief description of the setup and procedure, since they were basically the same for all experiments, followed by an introduction to the metrics used in the experiments. The experiments and the results are then presented in more detail.

3.1 Setup and procedure

Two computers were used along with an iView X RED-III eye tracking device by SensoMotoric Instruments (Berlin, Germany). The eye tracker samples at 50 Hz with 1° gaze position accuracy. The data collected were fixation data, raw eye data, and event data, as described below.

The experimental software had an on-screen keyboard, a “Ready” key, a “Del” key, and two text fields (one each for the source and typed text, see Fig. 2). The Finnish speech synthesizer Mikropuhe by Timehouse Oy was used for spoken feedback (with default parameters).

For all experiments, the task was to type short phrases of text. Participants were instructed to first read and memorize the source phrase and then to eye type it as quickly and accurately as possible.

The participant sat in front of the monitor, with a distance of 70–80 cm between eyes and the tracker. The participants were instructed to sit still. However, their (head) movements were not restricted in any way. The eye tracker was then calibrated (and re-calibrated before phrases were shown). Some practice phrases were then entered.

During the experiment, each participant was presented with short, simple phrases of text, one at a time. All phrases were in Finnish as that was the native language of the participants in all experiments. After typing the given phrase, the participant looked at the Ready key to load the next phrase.

Participants could correct errors—delete the last letter typed—by looking at the Del key. They were told to correct errors if noticed immediately, but not to correct errors in the middle of a phrase after an entire phrase was typed. In the analyses, both corrected errors and errors left in the final text are analyzed.

Figure 2 illustrates an example gaze path for a participant eye typing one phrase. The gaze progression is roughly as follows: the participant read the source text, then typed the phrase by fixating on the letters on the virtual keyboard (correcting errors as they occur), and finally looked at the Ready key in the lower right corner.

3.2 Metrics

The performed analysis included typing speed, accuracy, gaze behavior, and responses from the interviews.

Typing speed is measured in wpm, where a word is any sequence of five characters, including letters, spaces, punctuation, etc. [15].

When measuring *accuracy*, both corrected errors and errors left in the final text are taken into account. The metrics used were error rate and keystrokes per character (KSPC).

Error rate is calculated by comparing the transcribed text (text written by the participant) with the presented text, using the minimum string distance method described by Soukoreff and MacKenzie [24]. The method does not take into account corrected errors.

Keystrokes per character [24] measures the average number of keystrokes used to enter each character of text. Ideally, KSPC = 1.00 indicating that each key press produces a character. If participants correct mistakes during entry, KSPC is greater than 1. For example, if “hello” is entered as h e l x [del] l o, the final

result is correct (0% error rate), but KSPC is $7/5 = 1.4$ (seven keystrokes for entering five characters). KSPC is an accuracy measure reflecting the overhead incurred in correcting mistakes.

In addition to typing speed and accuracy, aspects of gaze behavior were studied.

Read text events (RTE) refers to a participant switching point of gaze from the virtual keyboard to the typed text field to review the text written so far (see Fig. 2 for one such example). Instead of reporting raw counts, RTE is normalized and reported on a “per character” basis. The ideal value is 0, implying that participants were confident enough to proceed expeditiously without verifying the transcribed text. Typically, however, participants occasionally review their work. This is known to occur more frequently for inexperienced participants [2]; however, the type of feedback may also have an effect, as seen in the results of the experiments, discussed below.

Re-focus events (RFE) is a measure of the average number of times a participant re-focuses on a key to select it. As with RTE, RFE is normalized and reported on a “per character” basis. RFE is ideally 0, implying that the participant focused on each key just once. If the participant’s point of gaze leaves a key before it is selected, and then re-focuses on it without selecting anything else in between, RFE is greater than 0.

Participants’ subjective experience was also of interest. Participants’ opinions were collected via interviews and questionnaires.

The statistical analyses were done using repeated measures ANOVA and Bonferroni corrected *t*-tests. Data collection for a phrase started on the press of the first character and ended on the press of the Ready key. (“Press” in this context means a successful selection of the key by gaze.) Each experiment and the results are reported in detail in the subsequent sections.

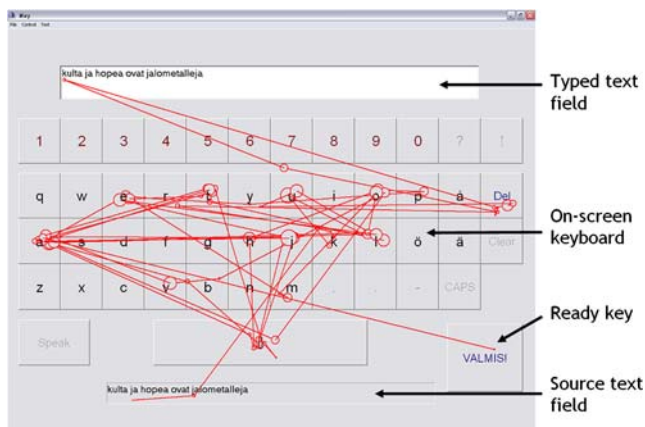


Fig. 2 A gaze path of a participant eye typing a phrase

4 Experiment 1: Effects of auditory and visual feedback on eye typing with a long dwell time

The first experiment used a relatively long dwell time while studying the effect of auditory feedback on user performance. Speech and non-speech auditory feedback, as well as no auditory feedback, were tested. The initial hypothesis was that added auditory feedback would improve performance.

4.1 Participants and design

Sixteen participants volunteered for the experiment. Data from three participants were discarded due to technical problems. In the end, there were five females and eight males (mean age 23 years). All were able-bodied university students with normal or corrected-to-normal vision. None had previous experience with eye tracking or eye typing, but all were familiar with desktop computers and a QWERTY keyboard layout.

Four feedback modes were tested (see Table 1):

Visual Only: The key is highlighted on focus and its symbol shrinks as dwell time progresses. On selection, the letter turns red and the key goes down.

Click + Visual: The same as the Visual Only mode, with the addition of a short audio “click” on selection.



Speech + Visual: The same as the Visual Only mode, with the addition of synthetic speech feedback. The letter on the key is spoken on selection.

Speech Only: The Speech Only mode does not use visual feedback. The symbol on the key is spoken on selection.

The dwell time for selection was the same, 900 ms, for all modes. This consisted of a delay before the onset of shrinking (400 ms) and the shrinking itself (500 ms).

The experiment was a 4×4 repeated measures design with four feedback modes and four blocks. Participants were tested in four sessions containing four blocks each. Each block had a different feedback mode (in randomized order). A block involved the entry of five short phrases of text. There was a pause after each block. In the last session, the participants were interviewed and filled-in a questionnaire. The results are based on a total

Table 1 Feedback modes in the first experiment

Feedback mode	While focused	When selected
Visual Only	highlight, shrinking 	red letter, key down 
Click+Visual	highlight, shrinking	red letter, key down, “click”
Speech+Visual	highlight, shrinking	red letter, key down, letter spoken
Speech Only	none	letter spoken

of 1,040 phrases (13 participants \times 4 feedback modes \times 4 blocks \times 5 phrases).

4.2 Results and discussion

4.2.1 Typing speed

The grand mean for typing speed was 6.97 wpm. This is quite typical for eye typing [5, 16], but is too slow for fluent text entry. However, the experiment showed that participants improved significantly with practice over the four blocks of input ($F_{3,36} = 10.92, p < 0.0001$, see Fig. 3).

Feedback mode had a significant effect on text entry speed ($F_{3,36} = 8.77, p < 0.0005$). The combined use of click + visual feedback yielded the fastest entry rate, with participants achieving a mean of 7.55 wpm in the last (fourth) block. The other fourth-block means were 7.14 wpm (speech + visual), 7.12 wpm (visual only), and 7.00 wpm (speech only).

4.2.2 Accuracy

The mean character level error rate was quite low (0.54%) and the participants' accuracy also improved significantly with practice ($F_{3,36} = 0.09, p = 0.005$). A significant main effect of feedback mode on error rate was found ($F_{3,36} = 5.01, p = 0.005$). Surprisingly, eye typing with speech only feedback was the most accurate technique throughout the experiment, with error rates under 0.8% on all four blocks. Visual only had the highest mean error rate (0.95%).

While the very low error rates overall sound encouraging, accuracy is also reflected in KSPC. Before presenting the results for KSPC, an additional comment on eye typing interaction is warranted. In eye typing, users frequently commit errors (especially with short dwell times) and immediately correct them. Thus, measuring accuracy only as errors in the final text is insufficient. On the other hand, a KSPC figure of, for example, 1.12 reflects about a 12% keystroke overhead due to the errors committed and corrected. Of this, 6% is for the initial error and 6% is for activating the Del

key. Thus, KSPC = 1.12 is roughly equivalent to a 6% error rate.

The grand mean for KSPC was 1.09, meaning there was about a 9% keystroke overhead in correcting errors (roughly corresponding to a corrected error rate of 4.5%). The KSPC for the Visual Only feedback mode was highest in all four blocks, ranging from 1.15 on the first block to 1.10 on the fourth block (the effect on feedback mode was significant, $F_{3,36} = 3.60, p < 0.05$).

4.2.3 Gaze behavior

No significant differences were found in the total number of fixation events between the feedback modes. However, there were significant differences in the participants' gaze path behavior.

In the first experiment, the grand mean was 0.064 RTE per character. By feedback mode, the RTE means were 0.047 (Speech Only), 0.051 (Click + Visual), 0.049 (Speech + Visual), and 0.110 (Visual Only). The differences were statistically significant ($F_{3,36} = 30.06, p < 0.0001$). In particular, RTE for Visual Only feedback was over 100% higher than for any other mode (see Fig. 4). Participants moved their point of gaze to the typed text field approximately once every ten characters entered for the Visual Only feedback mode, but only about once every 20 characters for the other modes. This may be due to the fact that auditory feedback (used with all except the Visual Only mode) significantly reduces the need to review and verify the typed text, and brings in a sense of finality which simply does not surface, at least to the same degree, through visual feedback alone.

4.2.4 Subjective satisfaction

The Click + Visual feedback mode was preferred by 62% of participants in the first experiment (15% Speech + Visual, 15% Speech Only, 8% Visual Only). Participants felt that spoken feedback or the "click" sound suitably supported visual feedback. The synthesized voice annoyed some participants, though.

By the end of the experiment (after four sessions of eye typing), all participants agreed that the dwell time

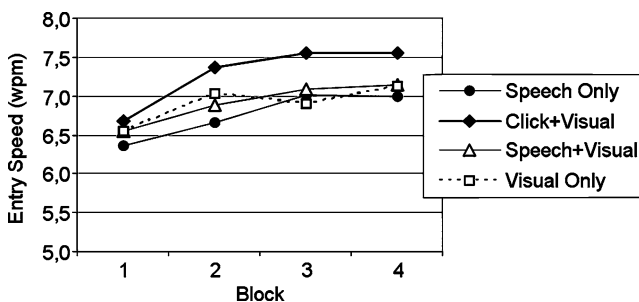


Fig. 3 Entry speed (wpm) by feedback mode and block

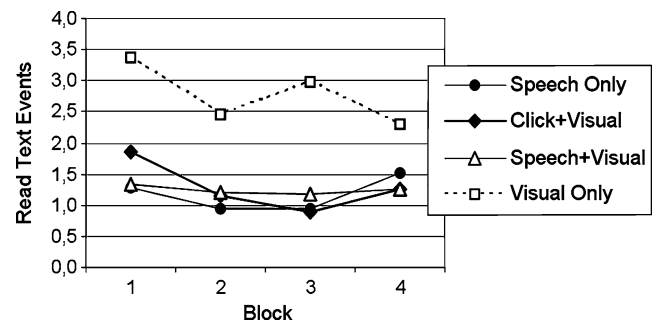


Fig. 4 Mean RTE by feedback mode and block

(900 ms) was too long, even if it was appropriate at the beginning of the experiment. Participants reported that the long dwell was tiring to the eyes and made it hard to concentrate.

5 Experiment 2: Effects of animated feedback with a long dwell time

The second experiment was similar to the first, except for a more close investigation of the “shrinking letter” condition. It was felt that shrinking not only serves as a good indicator of dwell time progress, but also draws the user’s attention in, thus helping the user to concentrate on the center of the key.

5.1 Participants and design

Twenty university students (9 females, 11 males, mean age 27) volunteered for the experiment. All were able-bodied with normal or corrected-to-normal vision. None had previous experience with eye typing.

The following feedback modes were tested (Table 2):

Shrinking: This is the same as “Click + Visual” in experiment 1. Here, it is called “Shrinking” because the experiment is constrained to study only the effect of the shrinking letter.

No Shrink: The same as Shrinking, but the symbol does not shrink.





The experiment was a repeated measures design with two feedback modes. The order of feedback modes was counter-balanced. The results are based on the total of 200 phrases (20 participants \times 2 feedback modes \times 5 phrases).

5.2 Results and discussion

5.2.1 Typing speed

Grand mean for typing speed was 6.83 wpm. The feedback mode had a significant effect on text entry speed ($t = 2.94$, $df = 17$, $p < 0.01$); a significantly faster text entry rate was observed in the Shrinking mode with the mean of 7.02 wpm, as compared to the No Shrink mode

Table 2 Feedback modes in the second experiment

Feedback mode	While focused	When selected
Shrinking	highlight, shrinking 	red letter, key down, “click” 
No Shrink	highlight 	red letter, key down, “click” 

with 6.65 wpm. An explanation for the lower speed in No Shrink mode was found when studying gaze behavior (discussed below).

5.2.2 Accuracy

In this experiment, the feedback mode did not have a significant effect on error rates or KSPC. The character level error rates were quite low (0.43%) the grand mean for KSPC was 1.09.

5.2.3 Gaze behavior

There were no significant effects of feedback mode on RTE, which measures a participant’s gaze behavior within the typed text field. However, there were significant effects on the gaze behavior within a key (on the virtual keyboard).

Re-focus events was only studied in the second experiment, in order to understand the effects of the shrinking letter on participants’ gaze behavior within a key. Indeed, as shown in Fig. 5, RFE was about 59% higher for the No Shrink condition (0.297) than for the Shrinking condition (0.187) ($t = 4.56$, $df = 19$, $p < 0.001$). Higher RFE for No Shrink indicates that participants gazed away from a key too early, before it was selected, necessitating re-focus. Therefore, the shrinking helped participants maintain their focus on the key. The higher RFE probably also explains the decrease in typing speed (reported above) as re-focusing a key takes time.

5.2.4 Subjective satisfaction

Fifty percent of the participants preferred the Shrinking mode, with 65% finding shrinking helpful in concen-

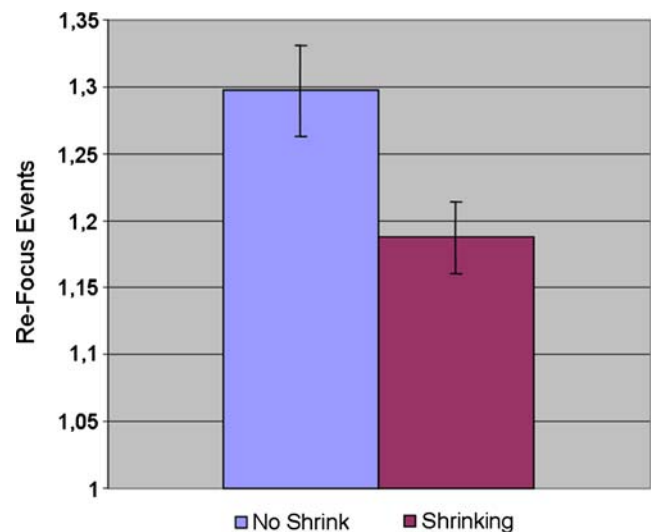


Fig. 5 Mean RFE per key (and the standard error of the mean)

trating on the key. Participants agreed that shrinking helped them understand the progression of dwell time. Some participants emphasized that “shrinking supports the typing rhythm”.

Two participants considered shrinking disturbing. Interestingly, some participants did not notice the difference between the modes, so the shrinking obviously did not disturb them. However, most participants agreed that shrinking might be disturbing and tiring in the long run, even though it helps novices in learning eye typing.

6 Experiment 3: Effects of feedback on eye typing with a short dwell time

In this experiment, the effects of feedback when a short dwell time is used were studied. It was felt that the results from the first experiment may not apply with short dwell times. For example, with longer dwell times, two-level feedback (focus + selection) is beneficial because the user has a comfortable opportunity to cancel before selection. With shorter dwell times, this may not be possible or may be more error prone.

6.1 Participants and design

Eighteen students volunteered for the experiment. Due to technical problems, data from three participants were discarded. In the end, there were ten males, and five females (mean age 25 years). All had normal or corrected-to-normal vision. All had participated in either experiment 1 or 2. The experiment involved experienced participants because a shorter dwell time was used, and it was important to compare the results with those from earlier experiments.







The visual feedback was simplified based on pilot tests and experiences from the previous experiments. The following feedback modes were tested (Table 3):

Speech Only: The symbol on the key is spoken on selection.

One-Level Visual: The key background turns red on selection.

Two-Level Visual: The key is highlighted on focus. On selection the key background turns red.

Table 3 Feedback modes in the third experiment

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

The dwell time for selection was 450 ms, i.e., half that (900 ms) for experiments 1 and 2. For the Two-Level Visual mode, the delay before highlight was 150 ms. Thus, the highlighting (before selection) lasted 300 ms. Based on the experience from pilot studies, the dwell time to re-select the current letter was increased by 120 ms to avoid erroneous double entries (e.g., ‘aa’). Thus, the dwell time for the second of the two consecutive letters was $450 + 120 = 570$ ms.

The experiment was a repeated measures design using a counterbalanced order of presentation. The results are based on the total of 450 phrases (15 participants \times 3 feedback modes \times 10 phrases).

6.2 Results and discussion

6.2.1 Typing speed

The grand mean for typing speed was 9.89 wpm. A faster entry speed was expected, since the dwell time was less and the participants were experienced. The feedback mode had a significant effect on text entry speed ($F_{2,28} = 6.54, p < 0.01$). The Speech Only mode was significantly slower than either of the two visual feedback modes with a mean of 9.22 wpm. The means for the other modes were 10.17 wpm (One-Level Visual) and 10.27 wpm (Two-Level Visual).

6.2.2 Accuracy

The overall error rates were higher in the third experiment, but still quite low, with a grand mean of 1.20%. The increased error rate is no surprise, since there always is a trade off between speed and accuracy in text entry tasks. In other words, reducing the dwell time tends to push entry speed up while reducing accuracy.

The effects of feedback mode on error rate were not significant. However, the differences in KSPC across feedback modes were significant ($F_{2,28} = 9.83, p < 0.005$). KSPC for Speech Only (mean 1.28) was significantly higher compared to the One-Level (mean 1.17) and Two-Level Visual (mean 1.19) modes (in both, $p < 0.05$). So, despite the relatively low error rates (about 1.20%), quite a few errors were committed and corrected, particularly with the Speech Only feedback mode (with 28% keystroke overhead, corresponding roughly to a 14% corrected error rate).

As with entry speed, the higher KSPC for Speech Only in the third study was likely due to participants’ tendency to pause and listen to the speech synthesizer. If the participant spent more time on the key than the defined dwell time, this caused an unintended double entry, as confirmed by a closer examination of the error types (see Fig. 6). There were significantly more (corrected) double entry errors in the Speech Only mode than in the other two modes ($F_{2,28} = 19.12, p < 0.001$). Even though the double entry problem had been

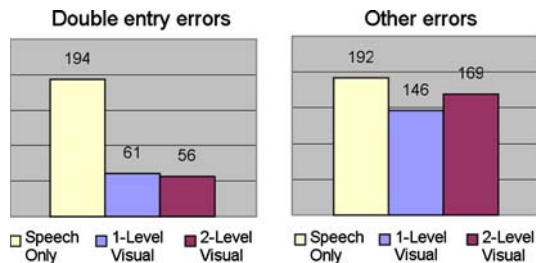


Fig. 6 Double entry errors (left), and other errors

anticipated based on the pilot tests, the obtained results indicate that the increase in dwell time ($450 + 120 = 570$ ms) is insufficient to avoid this for many participants, especially with the Speech Only mode.

6.2.3 Gaze behavior

The feedback mode had a significant effect ($F_{2,28} = 4.50, p < 0.05$) on RTE, with means of 0.139 (Speech Only), 0.087 (One-Level Visual), and 0.140 (Two-Level Visual). The One-Level Visual mode had significantly lower RTE than the Speech Only or Two-Level Visual mode.

The increased need to review the typed text in the Two-Level Visual mode can be attributed to a degree of confusion in separating focus and selection. Participants were not sure whether the key was already selected or not. This was confirmed by the comments received (discussed below). With the Speech Only mode, the higher RTE followed from an increased need to verify corrections (participants typically reviewed the typed text every time they corrected an error).

6.2.4 Subjective satisfaction

When a short dwell time was used, the One-Level Visual feedback was preferred by 47% of the participants (33% liked Two-Level Visual, and 20% Speech Only). The participants appreciated the simplicity of the One-Level Visual feedback. About half found the highlighting for focus in the Two-Level Visual mode disturbing, as it caused extra visual noise. 33% said they did not get any benefit from the shown focus, and the short time between the focus and selection ($450 - 150 = 300$ ms) was not long enough to react to. However, another 33% appreciated the extra confidence provided by this feature: they found the 300 ms long enough to adjust the point of gaze (all were experienced participants familiar with the two-level feedback in a previous experiment).

Many participants expressed a desire to have a “click” sound to support the visual feedback and typing rhythm. In addition, some participants said that they would like speech more if it were combined with visual feedback. Most participants (80%) found the typing speed with a 450 ms dwell time “just right”. No one

considered the dwell time too long, but for three (experienced) participants the dwell time was too short. A short dwell time may be problematic, especially for novices, as was discovered by Hansen et al. [8] who used a dwell time of 500 ms in their eye typing study.

7 Summary

The results of the experiments indicate that the type of feedback significantly impacts typing speed, accuracy, gaze behavior, as well as users’ subjective experience. Furthermore, dwell time duration impacts the suitability of certain types of feedback. Even though the differences were not especially large in some cases (a slight increase in accuracy or a few extra characters written due to increased speed), they are important in a repetitive task where the effect accumulates. Users may adapt to the shortcomings of the feedback up to a point. However, as seen in the first experiment, the effects (on performance and accuracy) were still significant after four sessions. Nevertheless, since the participants of the experiments were either first-time users (experiments 1 and 2) or only had little experience (experiment 3), the results best apply to novices.

Participants’ preferences naturally varied in all experiments, but some consistent opinions were found. For example, the use of an audible “click” was generally liked. Participants also appreciated feedback that clearly marks selection, as well as feedback in support of their typing rhythm.

Typing rhythm is considered important because dwell time, as an activation command, imparts a sense of rhythm to the task. When typing “as fast as possible”, the participant no longer waits for the feedback, but learns to take advantage of the rhythm inherent in the dwell time duration. In other words, the interaction (type a letter, proceed to the next letter) is no longer based on reaction time, but follows from the rhythm given by the dwell (and search) time. Rhythm-based eye typing may also exacerbate the problem with erroneous double entries, since the typing rhythm is interrupted due to the reduced search time.

To fully appreciate the results, it is important to understand the basic difference between using dwell time as an activation command compared to, for example, a button click. With a button click, the user makes the selection and defines the exact moment when the selection is made. Using dwell time, the user only initiates the action; the system makes the actual selection after a predefined interval. Therefore, general guidelines on feedback on graphical user interfaces may not be suitable as such. The feedback should be matched with the dwell time.

8 Guidelines

Based on the results of the experiments, as well as knowledge from previous research, the following six guidelines on feedback in eye typing were formulated.

8.1 Use a short non-speech sound to confirm selection

Visual feedback combined with a short audible “click” produced the best results in the first experiment, and was also preferred by the participants. Even though “click” was not tested with a short dwell time, many participants in the third experiment commented that the visual feedback would benefit from an additional “click”. This is consistent with previous research. A non-speech sound not only confirms selection, but also supports the typing rhythm better than visual feedback alone [3, 4].

8.2 Combine speech with visual feedback

Even though the Speech Only mode produced good results in the first experiment (with a long dwell time), it was not liked by participants. Some of the participants found the spoken feedback on every keystroke quite disturbing. Furthermore, some letters are hard to distinguish by speech alone (e.g., ‘n’ and ‘m’ sound similar and are located next to each other in the QWERTY keyboard). Spoken feedback is especially problematic with short dwell times, since speaking a letter takes time. As demonstrated in the third experiment, people paused to listen to the speech. This, in turn, not only decreased typing speed, but also decreased accuracy. Since the time to speak a letter varies (e.g., ‘e’ vs. ‘m’ vs. ‘w’), spoken feedback does not support the typing rhythm (especially with short dwell times). However, if speech is combined with visual feedback, it can improve performance, as seen in the first experiment, and may be helpful especially for novices.

8.3 Use simple, one-level feedback with short dwell times

Short dwell times require sharp, clear feedback. For example, spoken feedback may be problematic, because it may not be clear to the user whether the selection occurs at the beginning or end of the synthetically spoken letter. When a short dwell time is used, there is no time to give extra feedback to the user. As seen in the third experiment, the two-level feedback was confusing and distracting, even though the measured performance was reasonably good. With very short dwell times, the feedback for focus and selection are not distinguishable anymore. The feedback should have a distinct point for selection; there should be no uncertainty of the exact moment when the selection is done.

General usability guidelines (e.g., [21]) indicate that feedback on actions should be provided within reasonable time. This also applies for gaze input. As the results of the conducted experiments show, people may gaze away from a key too early if feedback is delayed. Thus, a simple one-level feedback may not work well with long dwell times. A possible solution is to give feedback for focus before selection occurs. The separated two-level

feedback (focus + selection) should, however, be designed carefully to avoid confusion.

8.4 Make sure focus and selection are distinguishable in two-level feedback

As seen in the experiments, two-level feedback may cause confusion. A simple audible click helps to make the moment of the selection distinct and clear. As seen in the second experiment that used a click in both conditions, the shift between focus and selection is also strengthened by animation.

8.5 Use animation to support focus with long dwell times

Since animation takes time, it is difficult to use with short dwell times, or users may (at least first) find it confusing [8]. However, with long dwell times, animation provides extra information on the dwell time progress. It is not natural to fixate for a long time at a static target [25], so animation helps in maintaining focus on the target letter for long dwell times. In the second experiment, the shrinking letter improved performance by helping users focus on the center of the key. Animation gives continuous feedback for continuous waiting (for the dwell time to end). The animation should be designed carefully; it should be subtle and not distract the user from the task at hand. To the extent possible, animation should show in a direct, continuous fashion the time remaining to selection.

8.6 Provide a capability to adjust feedback parameters

The dwell time should, of course, be adjustable. The same 500 ms may be “short” for one user and “long” for another. The needs and the preferences of the users vary a lot; this is especially true with people with disabilities [10]. Therefore, the final guideline is to support user control of feedback parameters and attributes.

9 Conclusions

This paper has showed that the tedious task of eye typing is facilitated using proper feedback. The type of feedback impacts both user performance and experience. By adding a short audible “click”, both the typing speed and accuracy is improved. The duration of the dwell time affects the suitability of different types of feedback: Short dwell times require short and clear feedback while long dwell times enable showing extra information.

Acknowledgments The work presented in this paper was partly supported by the Graduate School in User-Centered Information Technology (UCIT) and by the COGAIN Network of Excellence

funded by the European Commission. The authors would like to thank all volunteers who participated in the experiments, and Mika Käki, Aulikki Hyrskykari, Salla Ovaska, and Harri Siirtola from the TAUCHI unit, and Nancy and Dixon Cleveland from LC Technologies, Inc., for consultation and inspiring discussions.

References

- Baecker R, Small I, Mander R (1991) Bringing icons to life. In: Proceedings of CHI 1991. ACM Press, Cambridge, pp 1–6. DOI 10.1145/108844.108845
- Bates R (2002) Have patience with your eye mouse! Eye-gaze interaction with computers can work. In: Proceedings of 1st Cambridge workshop on universal access and assistive technology, pp 33–38. <http://www.rehab-www.eng.cam.ac.uk/cwuaat/02/7.pdf>
- Brewster SA, Rätty V-P, Kortekangas A (1996) Enhancing scanning input with non-speech sounds. In: Proceedings of ACM ASSETS '96, pp 10–14. DOI 10.1145/228347.228350
- Brewster SA, Crease MG (1999) Correcting menu usability problems with sound. *Behav Inf Technol* 18(3):165–177. DOI 10.1080/014492999119066
- Frey LA, White KP Jr, Hutchinson TE (1990) Eye-gaze word processing. *IEEE Trans Syst Man Cybern* 20(4):944–950
- Gaver WW (1989) The SonicFinder: an interface that uses auditory icons. *Hum Comput Interact* 4(1):67–94
- Glenstrup AJ, Engell-Nielsen T (1995) Eye controlled media: present and future of state. Technical report, University of Copenhagen. <http://www.diku.dk/~panic/eyegaze/>
- Hansen JP, Johansen AS, Hansen DW, Itoh K, Mashino S (2003) Command without a click: dwell time typing by mouse and gaze selections. In: Proceedings of INTERACT '03. IOS Press, Amsterdam, pp 121–128
- Hillstrom AP, Yantis S (1994) Visual motion and attentional capture. *Percept Psychophys* 55(4):399–411
- Hutchinson TE, White KP, Martin WN, Reichert KC, Frey LA (1989) Human-computer interaction using eye-gaze input. *IEEE Trans Syst Man Cybern* 19(6):1527–1534
- Jacob RJK (1991) The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Trans Inf Syst* 9(2):152–169. DOI 10.1145/123078.128728
- Jacob RJK (1995) Eye tracking in advanced interface design. In: Barfield W, Furness TA (eds) *Virtual environments and advanced interface design*. Oxford University Press, New York, pp 258–288
- Just MA, Carpenter PA (1976) Eye fixations and cognitive processes. *Cogn Psychol* 8:441–480
- Lankford C (2000) Effective eye-gaze input into windows. In: Proceedings of ETRA '00. ACM Press, Cambridge, pp 23–27. DOI 10.1145/355017.355021
- MacKenzie IS (2003) Motor behaviour models for human-computer interaction. In: Carroll JM (ed) *Toward a multidisciplinary science of human-computer interaction*. Morgan Kaufmann, San Francisco, pp 27–54
- Majaranta P, Riih  K-J (2002) Twenty years of eye typing: systems and design issues. In: Proceedings of ETRA '02. ACM Press, Cambridge, pp 15–22. DOI 10.1145/507072.507076
- Majaranta P, MacKenzie IS, Aula A, Riih  K-J (2003a) Auditory and visual feedback during eye typing. In: Ext Abstracts CHI 2003. ACM Press, Cambridge, pp 766–767. DOI 10.1145/765891.765979
- Majaranta P, MacKenzie IS, Riih  K-J (2003b) Using motion to guide the focus of gaze during eye typing. In: Proceedings of 12th European conference on eye movements—ECEM 12, University of Dundee, O42
- Majaranta P, Aula A, Riih  K-J (2004) Effects of feedback on eye typing with a short dwell time. In: Proceedings of ETRA 2004. ACM Press, Cambridge, pp 139–146. DOI 10.1145/968363.968390
- Microsoft Windows User Experience (2002) Official guidelines for user interface developers and designers. Microsoft Corporation
- Nielsen J, Mack RL (1994) *Usability inspection methods*. Wiley, New York
- Robertson CG, Mackinlay JD, Card SK (1991) Cone trees: animated 3D visualizations of hierarchical information. In: Proceedings of CHI '91. ACM Press, Cambridge, pp 189–194
- Seifert K (2002) Evaluation multimodaler computer-systeme in fr hen entwicklungsphasen (in German). PhD Thesis, Technical University Berlin. Summation of the results involving gaze interaction (in English) available at <http://www.roetting.de/eyes-tea/history/021017/seifert.html>
- Soukoreff RW, MacKenzie IS (2001) Measuring errors in text entry tasks: an application of the Levenshtein string distance statistic. In: Ext Abstracts CHI '01. ACM Press, Cambridge, pp 319–320. DOI 10.1145/634067.634256
- Stampe DM, Reingold EM (1995) Selection by looking: a novel computer interface and its application to psychological research. In: Findlay JM, Walker R, Kentridge RW (eds) *Eye movement research: mechanisms, processes and applications*. Elsevier, Amsterdam, pp 467–478
- Ware C, Mikaelian HH (1987) An evaluation of an eye tracker as a device for computer input. In: Proceedings of CHI/GI '87. ACM Press, Cambridge, pp 183–188. DOI 10.1145/29933.275627