Influences of dwell time and cursor control on the performance in gaze driven typing

Jens R. Helmert
Applied Cognitive Research / Institute of Psychology III, Technische Universitaet Dresden

Sebastian Pannasch
Applied Cognitive Research / Institute of Psychology III, Technische Universitaet Dresden

Boris M. Velichkovsky
Applied Cognitive Research / Institute of Psychology III, Technische Universitaet Dresden

In gaze controlled computer interfaces the dwell time is often used as selection criterion. But this solution comes along with several problems, especially in the temporal domain: Eye movement studies on scene perception could demonstrate that fixations of different durations serve different purposes and should therefore be differentiated. The use of dwell time for selection implies the need to distinguish intentional selections from merely perceptual processes, described as the Midas touch problem. Moreover, the feedback of the actual own eye position has not yet been addressed to systematic studies in the context of usability in gaze based computer interaction. We present research on the usability of a simple eye typing set up. Different dwell time and eye position feedback configurations were tested. Our results indicate that smoothing raw eye position and temporal delays in visual feedback enhance the system's functionality and usability. Best overall performance was obtained with a dwell time of 500 ms.

Keywords: Gaze based computer interaction, Dwell time, Cursor control, Eye tracking, Usability

Introduction

A gaze mouse – in its simplest version a gaze cursor – can be seen as a means to control a computer’s cursor with the eyes. Wherever a user looks, the cursor is there as well. Unfortunately, this has de facto no use in operating a computer. What is needed to add real benefit to such a gaze mouse is to enable a user to interact with the computer.

In this paper we will show how the use of dwell time as a means to initiate an action is related to a wide range of topics from the field of visual perception research. Especially, the nature and purpose of fixations have to be taken into consideration in seek for optimal dwell times in different contexts. We will reflect the functions of fixations on the base of the two visual systems approach (e.g. Ingle, 1967; Ungerleider & Mishkin, 1982; Velichkovsky, Joos, Helmert, & Pannasch, 2005). Additionally, we will show that knowledge derived from the Levels-of-Processing (e.g. Challis, Velichkovsky, & Craik, 1996) approach can be helpful in designing a gaze based interaction tool. These approaches provide a good base to discuss and investigate the pitfall of the Midas touch problem (Jacob, 1991). In seek for an optimised way of interacting with the computer in gaze based manner we will take into consideration how an actual eye position is linked to the subject’s conscious experience depending on the actually ongoing task (Helmert, Pannasch, & Velichkovsky, in press.; Velichkovsky, 1995).

On the basis of these considerations a study is presented using a simple virtual keyboard setup. Different designs to control the computer’s cursor through eye movements and to initiate “clicks” are tested and discussed.

Fixations, durations and functions

In the late 60s of the 20th century a distinction of two visual processing routes was promoted in a special issue
of Psychologische Forschung in 1967 (Ingle, Schneider, Trevarthen, & Held, 1967). In the following, different dichotomies were proposed, e.g. ambient-focal (Trevarthen, 1968), noticing-examining (Weiskrantz, 1972), or foveal vs. ambient (Stone, Dreher, & Leventhal, 1979). Later, attempts were made investigating the balance of these two systems in eye movement behaviour. Velichkovsky et al. (2005) could show that ambient processing is related to shorter fixations within saccades of large amplitudes, whereas focal processing is dominant in sequences of long fixations embedded in short saccadic amplitudes. The thresholds used for discrimination were at 180 ms for durations of fixations and 5° of visual angle for the amplitude of saccades. The same pattern was elicited in a virtual driving environment, though fixation duration thresholds were somewhat higher (around 300 ms) compared to the static setting (Velichkovsky, Rothert, Kopf, Dornhoefer, & Joos, 2002). In case of focal processing being required, we can conclude from these findings that the dwell time criterion should be well above the threshold for ambient fixations, namely above 200 ms.

Another line of research leading to a functional distinction of fixations comes from the levels of processing approach (e.g. Craik & Lockhart, 1972; Velichkovsky, 1999). Separable fixation duration segments could be attributed to levels of different depths of processing, with deeper processing being associated with longer fixations (Velichkovsky, Sprenger, & Unema, 1997). On the superficial level of figurative processing fixation durations are within the range of 120 to 250 ms, whereas the level of semantic processing spans a range from 250 to 450 ms. Provided that selecting a letter from a keyboard in the course of writing involves semantic processing the dwell time criterion should be above 400 ms.

Using gaze as an input device for the computer has advantages and disadvantages. On the one hand, it is intuitive and easy to learn, and as the direction of gaze and the focus of attention usually match (e.g. Just & Carpenter, 1976), it is also a faster pointing device than e.g. a mouse. On the other hand, problems arise from the fact that in gaze controlled environments the shifting gaze does not solely serve information uptake but also as control instance (Majaranta & Räihä, 2002). The Midas touch problem was described as the difficulty to separate explorative fixations serving visual processing from those, that are concerned with the intentional initiation of an action in the context of gaze based computer interaction (Jacob, 1991). One way to circumvent this interference of fixational functionality could be the use of dwell-free approaches, as e.g. selection by anti-saccades (for an overview see Huckauf & Urbina, this issue) or by combining gaze (pointing) and facial muscle activity (initiation of action) (Surakka, Illi, & Isokoski, 2004). However, in contexts where users have to rely solely on their gaze behaviour as input device, the combination of the latter example will fail, whereas selection by antisaccades is heavily dependent from a system with high spatial resolution, and a user whose eye movements can be tracked nearly perfectly.

**Visual feedback and usability**

In every day life, eye movements serve the demands of ongoing activities, like e.g. orientation in the environment and identification of task related objects. However, little attempt has been made to consider such natural activities in research on the relationship of visual attention and eye movements (e.g. Tatler, 2001; Velichkovsky, 1995). In the study by Velichkovsky (1995) experts and novices cooperated on a computer based puzzle task. They were free to communicate verbally, and in one third of the trials experts’ eye fixations were projected on both screens. In this condition the expert could guide the novice which piece was to be taken next not only verbally but also by showing him with his eyes. In retrospect experts reported difficulties to “catch their eye positions” because whenever the decision of making an eye movement to an intended location became conscious the eye marker was already in place.

Paradigms investigating the relationship of attention and eye movements in laboratory settings – i.e. ‘fixate & jump’ paradigms – were criticized for the unnaturalness of single saccade programming motivated by arbitrary instructions and the attention biasing effects of cues in such experiments (Fischer, 1999). Recent experiments (Helmert et al., in press.), aimed at closing the gap between natural and abstract settings, could show that the relationship between gaze position and subjectively perceived focus of visual work is heavily dependent from the ongoing task. Subjects were instructed to inspect a circular array of six pictograms in order to be able to tell which item they had looked at, at the moment the inspection screen was turned off. Two tasks were used, localisation and identification: In the former condition, participants were instructed to indicate their subjective impression where they had been looking when the screen was...
turned off. In the latter, the task was to indicate the content participants dealt with when the screen was turned off. In the localisation task, i.e. when the actual position of the eye should be reported, subjects indicated that their subjective focus of visual work was ahead of the actual eye movement, as could be expected from classical literature on visual attention (e.g. Posner, 1980). However, in the identification and categorisation task this could not be replicated. Instead, participants reported the actual eye position also in cases with very long fixation durations. The same task with short fixation duration indicated, that subjects’ perceived focus even lagged behind the actual eye position.

Obviously, these results should have implications on the way how the visual feedback of the actual eye position in a gaze controlled environment should be designed. The dissociation of perceived and actual eye position will be considered in the present study: A delay of cursor position feedback and/or smoothed cursor position will be contrasted with the direct feedback without any delays or smoothing.

The quality of different setups of an interactive environment can be measured on the base of processing times for a given task. More important for this study is to evaluate the usability of the different setups. According to the widely accepted definition of usability in the context of computer applications usability is the ‘Degree to which specified users can achieve specified goals in a particular environment with effectiveness, efficiency and satisfaction’ (ISO DIS 9241-11). ‘Effectiveness’ here refers to the completeness and accuracy with which goals can be achieved. Objective ‘efficiency’ refers to the effort or resources necessary to achieve the goals and ‘satisfaction’ refers to the subjective assessment by the user of factors such as workload, comfort, and ease of use whilst achieving these goals. These aspects of usability were considered in the present study and recorded using questionnaires.

Methods

Subjects

The participants included 9 male (n=3) and female (n=6) undergraduate students ranging in age from 22 to 28 years, with a mean age of 24.7 years. They participated individually in a single session lasting approximately one hour. The participants received course credit in exchange for their time. All subjects reported normal or corrected-to-normal vision and normal hearing. Informed consent was obtained.

Apparatus

Eye movements were recorded at 1000 Hz, using the EyeLink1000 remote eye-tracking system (SR Research Ltd., Ontario, Canada) with on-line detection of saccades and fixations. Fixation onset was detected and transmitted to the presentation system with a delay of approximately 12 ms. Images were displayed using a GeForce2 MX card and a CRT display (19-inch Iiyama Vision Master 451) at 1152 by 864 pixels with a frame rate of 100 Hz. Viewed from a distance of 90 cm, the screen subtended a visual angle of 23° horizontally and 17.2° vertically.

The interaction with the virtual keyboard was based on eye movements. Three different algorithms to control the cursor via eye movements were applied. The raw data condition overlaid the exact actual eye position onto the cursor. In the smoothed data version, data collected from the eyetracker were stored, and a gliding average over 30 samples (equalling 30 ms) was used to determine the actual cursor position. In the smoothed and delayed data (referred to as delayed in the following) the gliding average corresponded to 20 samples with a delay of data storing of another 10 samples (cf. Figure 1).

![Figure 1: Designs of interaction between eye movement data and cursor position.](image)
keyboard. In contrast to normal keyboards no space bar was provided as the task only covered writing of single words. Additionally, for the course of trials, two extra keys were provided. Driftkorrektur was designed as an “emergency button” allowing for a calibration or drift correction within a trial. The “DONE” was pressed by subjects to indicate that they had finished the task.

Figure 2: Virtual keyboard set up used in the experiment.

For the experiment a word data base of all together 114 words was used. They were chosen from a list of the 10,000 most common words (cf. http://wortschatz.uni-leipzig.de/Papers/top10000de.txt). The words’ lengths were 3, 5, 7, 9, 12 and 15, respectively, 19 of each length. Dwell times manipulated in the experiment corresponded to 350, 500 and 700ms.

Procedure

The experiment was conducted in single sessions in a dimly lit laboratory. After an instruction, informed consent of participants was obtained. The allocation of subjects to the different experimental condition sequences, i.e. dwell time and cursor control, was realised applying a latin square. All subjects participated in all experimental conditions leading to a completely repeated measures design. Prior to the experiment subjects practised with six words (equals 51 letters). The settings for these practise trials were the same as in the following first experimental block. This design was chosen due to the assumption that the learning curve would be steepest within the first attempts of writing with the eyes.

The experimental procedure consisted of three blocks referring to the three dwell times, 700, 500 and 350 ms. Every subject started with the longest dwell times, then the medium and the shortest in the last block. Within each block, sequences applying the different cursor control conditions – raw data, smoothed data and smoothed & delayed data – were presented. The order of cursor control conditions was counterbalanced with a latin square that ensured that the last session in one block was always different to the first sequence in the next block. Within each sequence, participants wrote a total of twelve words equaling 102 letters. After each sequence, subjects had to judge the gaze-based interaction with the virtual keyboard on two dimensions: appropriateness of dwell time and cursor control. The scale of both ratings subtended a range of -100 to +100. The end points of the selection time scale were labeled too fast (-100) and too slow (+100), i.e. zero value expressed best comfort with the sequence. For key control the extremes of the bar stated poor (-100) versus good (+100). A value of +100 therefore means best handling of the sequence. The whole experiment took about one hour.

Results

Subjective Ratings

In a first step, the appropriateness of dwell time of all dwell time x cursor control conditions was analysed. Mean ratings for short, medium and long dwell times were at -15.7, -0.22 and 15.7, respectively. For the cursor control conditions, the mean values were -5.0 (raw), 3.7 (smoothed) and 1.7 (delayed). A 3 (Dwell Time) x 3 (Cursor Control) ANOVA for repeated measures was conducted. Significant differences both for dwell time, F(2,16) = 21.214, p < .001, and cursor control, F(2,16) = 3.684, p = .048, were obtained, whereas the interaction of both factors was not significant, F < 1.

Secondly, the appropriateness of cursor control was analysed. Mean ratings for short, medium, and long dwell times were at 42.7, 43.6, and 37.2, respectively. The mean values for the cursor control conditions were 42.7 (raw), 54.7 (smoothed) and 46.2 (delayed). A 3 (Dwell Time) x 3 (Cursor Control) ANOVA for repeated measures yielded significance for cursor control, F(2,16) = 7.638, p = .005, but not for dwell time, F < 1. The interaction between the factors was insignificant, too (cf. Figure 3).
Performance

One crucial parameter of quality in typing is the orthographical correctness of the finally written text. Here, two measures were applied to check for performance in the study: Firstly, we were interested in how many words containing errors were produced at all. And secondly, key strokes per character (KSPC, cf. Soukoreff & MacKenzie, 2003) were analysed.

The former was calculated as the frequency of correctly typed words within all produced words. For short, medium and long dwell times the percentages were at 97%, 99%, and 98.5%, respectively. For raw, smoothed and delayed cursor interaction 98.2%, 99.1%, and 97.3%, respectively, were correctly typed. An ANOVA for repeated measures with dwell time and cursor as independent variables was conducted. Neither dwell time, $F(2,16) = 3.484, p = .055$, nor cursor control, $F(2,16) = 2.415, p = .121$, had a significant influence on the rate of erroneous words. Also the interaction of both factors remained insignificant, $F < 1$.

The latter measure, KSPC, was determined in the following way: The length of given word was used as the expected value of key strokes. Every correction, like deleting a wrong letter with backspace and typing the right letter, produced additional key strokes leading to higher KSPC for this word. For every typed word, the KSPC index was calculated and its average was analysed in an ANOVA with the same factors as described above. KSPC values for short, medium and long dwell times were 1.18, 1.03, and 1.00, respectively. For cursor control conditions, KSPC values were 1.05 (raw), 1.06 (smoothed), and 1.05 (delayed). No significant influence was found for cursor control, $F < 1$. Dwell time conditions differed significantly, $F(2,16) = 74.706, p < .001$ (cf. Figure 4).

Discussion

Dwell time

Looking at typing speed results alone suggests that the shortest dwell time is most appropriate. In this study,
typing speed with 350 ms dwell time was about 1.5 times faster than with 700 ms. However, taking into consideration all other measures available the picture changes. Not only there is an increase in error rate, as indicated by significantly higher KSPC, but also the subjective ratings reveal worse usability compared to medium and long dwell times. Medium dwell time, i.e. 500 ms, can be seen as the best compromise between usability and typing speed based on the data of this study. Selection time is rated optimal, and the overall key control judgement comparable to all other conditions. Moreover, with intermediate results in typing speed, and good performance as measured in KSPC, the medium dwell time condition is preferable to the other two tested conditions.

This result corroborates the findings by Velichkovsky et al. (1997), that semantic processing is characterized by fixation duration around 400 to 500 ms. In gaze driven typing, there is indeed a necessity to semantically process the letters before the decision is made to actually select this very letter, i.e. to “press” the key. Additionally, there was no evidence for a benefit in terms of subjective ratings for the longer dwell time. In the opposite, the long dwell time conditions were rated as being too slow.

It remains however unclear, in how far additional training could change the obtained results. In the course of this study we made one experience that showed such an influence. In pretests the order of conditions was different to the one presented here: Again, subjects participated in all combinations of dwell times and cursor control variations. But the blocks were dedicated to the cursor controls, i.e. in the first block e.g. the raw data condition with firstly 750 ms, then 500 ms and lastly 350 ms. And in the next block, the same was done with another cursor conditions. In these pretests, subjects complained that it was hard to use long dwell time after they had experienced shorter dwell times already. Moreover, there are studies showing that subjects experienced in gaze controlled typing can reach dwell times of around 250 ms are capable. The irritations in the described pretests and the very low dwell times for some experienced “eye typists” could be explained by further dealing with functions of fixations of shorter durations. In the nomenclature of Velichkovsky et al. (1997) shorter fixations serve figurative processing and localisation. One could hypothesize that experienced typists do not need to semantically process every single letter but rather rely on the location information provided by the experience with a certain keyboard design. Further studies will deal with a possible transfer of expertise in traditional type writing to gaze based typing.

**Cursor Control**

The results of this study show that the smoothed conditions of cursor control were judged better in contrast to the raw data condition. However, no significant differences were found between smoothed only and additionally delayed condition. Concerning the appropriateness of cursor control no significant interaction between dwell times and cursor control conditions was obtained. However, looking at the actual data reveals an interesting trend (cf. Figure 6). With decreasing dwell time the subjective judgement for the raw data cursor control conditions increases. For the shortest dwell time of 350 ms the results are even slightly better than for the delayed condition. It has been described in the literature (Majaranta, Aula, & Räihä, 2004) that short dwell times require sharp and clear feedback in contrast to longer dwell times. Based on the results of the actual study it could be hypothesized that this could be true for the way the cursor is controlled by the eye, too.

![Figure 6: Interaction of dwell time and cursor control in cursor interaction ratings.](image)

Unstructured postexperimental interviews with the participants asked for their impressions on how they felt writing with the eyes. First of all, most subjects reported to be very fascinated by this new technique. None of them had ever participated in an experiment dealing with gaze controlled interfaces. Interestingly, some participants tried to explain how the single cursor controls differed; especially the difference between the raw data condition and the other two were mentioned. Taking their reports together, the overall message was that the delay
both in the delayed as well as in the smoothed conditions caused the impression of taking the cursor to the next point with the eye. In contrast, in the raw data condition the impression was that the cursor always had been in place whenever the eye arrived at a new point on the screen. This notion is congruent with propositions in Velichkovsky (1995).

The setup of this simple eye typing environment was based on a 1000 Hz eye tracking system. Usually, especially due to the high prizes of such systems, one would not expect a system of such high temporal resolution in the context of real applications, e.g. as a system for the daily use by disabled persons (cf. e.g. Gips, DiMattia, Curran, & Olivieri, 1996). However, our intention was to firstly see how smoothing and delaying algorithms in cursor feedback can have influence on the usability of eye typing environments as such. Some preliminary data show that the algorithms tested here are also suitable for temporal resolutions as low as 50 or 100 Hz. It even turned out that the effect of such algorithms is more dramatic, because the higher pauses between acquisition of position data leads to noisier signals. Smoothing data in such cases removes some portions of noise.

Conclusions

In summary, the study shows that dwell times and the way a cursor is controlled by the eyes are – apart from the design of the actual visible interfaces – the key factors for the usability of a gaze based typing environment. For novice users a dwell time of around 500 ms is an appropriate starting point, as could be derived from the literature on fixation durations in the context of the Levels-of-processing approach (Velichkovsky et al., 1997). Smoothing the eye data, and thereby delaying the eye-cursor interaction, additionally improved the usability of the interface.

Further studies will have to deal with the influence of experience, even in a twofold way: (1) It is worth to investigate how dwell time and cursor control setup change with increasing experience with a certain interface. (2) In our study we found some hints that there might be transfer effects of experience with handed typing to the eye typing task. In both cases, it becomes clear that factors like controlling the cursor with the eye as well as dwell times of the buttons have to be subject to individual adjustability. Only this will provide the key to include users’ changing abilities into the setup of an existing gaze based interaction system.

Acknowledgment

This research has been supported by two grants from EU Commission PERCEPT (043261) and COGAIN (511598). Thanks are due to Sven-Thomas Graupner for technical support and an anonymous reviewer for valuable comments.

References


