ISSN 1392 – 124X INFORMATION TECHNOLOGY AND CONTROL, 2005, Vol. 34, No. 2

DYNAMIC TARGET EXPANSION TO FACILITATE EYE-BASED POINTING AT MENUS

Oleg Špakov, Darius Miniotas

Unit for Human-Computer Interaction, University of Tampere
FIN-33014 Tampere, Finland

Abstract. With recent advances in eye tracking technology, eye gaze gradually gains acceptance as a pointing modality. Its relatively low accuracy, however, determines the need to use enlarged controls in eye-based interfaces. This renders the overall design quite distant from "natural". Another factor impairing pointing performance is deficient robustness of an eye tracker's calibration. To facilitate pointing at standard-size menus, we developed a technique that uses dynamic target expansion for on-line correction of the eye tracker's calibration. Correction is based on the relative change in the gaze point location upon the expansion. A user study suggests that the technique affords selection accuracy of 91%. User performance is thus shown to approach the limit of practical pointing. Effectively, developing a user interface that supports navigation through standard menus by eye gaze alone is feasible.

Keywords: eye tracking, eye-based interaction, pointing, menus, target expansion, human performance.

1. Indroduction

A growing number of uses for eye-based interactive systems manifest the ability of the eye to function as a pointing device (see, e.g., [1] for a survey). The most prominent examples are applications involving eye typing, eye drawing, and other eye-controlled tasks designed primarily for people with disabilities.

Nevertheless, the design of those user interfaces renders them quite distant from what is perceived as "natural" (i.e., today's standard GUIs with their widgets). One of the major differences is the size of on-screen objects.

Most standard GUI widgets (e.g., icons in a toolbar, checkboxes, etc.) span less than one degree of visual angle. For instance, a toolbar's icon in a standard MS Windows™ application (e.g., MS Word™) is 24 by 24 pixels in size. This translates into approximately 0.7 degrees for a 17-inch monitor with a resolution of 1024 x 768 and a viewing distance of 70 centimetres. Meanwhile, the size of a button in a window's title bar is even smaller (only 16 by 16 pixels, or 0.46 degrees). Moreover, icons in a toolbar are usually aligned side by side: there is no space between.

In traditional applied eye tracking research, however, targets below the one-degree limit are considered too small for facile eye gaze interaction [2, 5]. Consequently, gaze-operated objects are made substantially bigger to ensure facile interaction (i.e., to bring gaze pointing to the level of practical accuracy). This measure accommodates calibration errors of the eye tracker as well as inherent limitations in the accuracy of eye gaze.

For the same reason, objects are also spaced on the screen at relatively large distances from one another. In turn, this poses problems in managing the real estate of the screen. Therefore, it is not surprising that, apart from applications for people with disabilities, current gaze-based interfaces are still rare in solutions for the general population of computer users.

One solution to the problem of limited screen space is dynamic target expansion. Using this approach, iconic targets are expanded to a "pointing-friendly" size when the user needs to interact with them; otherwise they appear in a reduced size. Dynamic target expansion was first successfully applied in target acquisition tasks with conventional (i.e., manually operated) pointing devices [3, 6].

For eye-based pointing, target expansion was also shown to facilitate performance [4]. To accommodate the peculiarities of eye gaze input, we modified the approach by substituting static expansion for dynamic one. That is, the region of expansion was determined a priori, and the expansion was not visually presented to the user. In other words, the interface responded to gaze point within the boundaries of the expanded target area, even though the target's appearance did not change.

In [4] we argued that static target expansion was more reasonable for eye-based interfaces due to the jumpy nature of eye movements and inherent eye jitter. However, we believe that the dynamic approach could also be useful in certain situations. This is particularly true of interaction with GUI controls comprised of multiple items aligned in one dimension.

A menu serves as a good example for this. Menu items have captions usually spanning more than 60 pixels (1.7 degrees). Since this is markedly above the critical one-degree limit, no expansion is needed in the horizontal dimension. On the other hand, the height of a menu's item is only 20 pixels (0.6 degrees). Therefore, vertical expansion is required.

Another problem peculiar to eye tracking systems is deficient robustness of their calibration. As calibration typically deteriorates over time, the system returns false estimates for the gaze point location. In turn, this has an adverse impact on pointing performance.

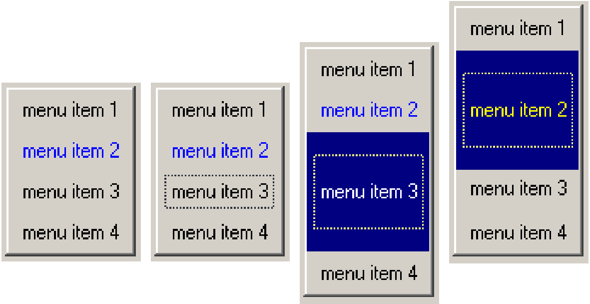
To address the issues above, we developed a technique that uses dynamic expansion of standard-size menu items to facilitate their selection in the presence of inaccuracies in the eye tracker's calibration.

The structure of this paper is as follows. After the technique is introduced, we present a pilot study to empirically determine the values for certain performance-critical parameters. Then, we describe results from a user study conducted to evaluate the technique.

2. Technique

First, we will explain the concept using a specific example. Then, we will present the general algorithm after defining the variables and parameters involved.

When the user does not look at the menu, it has a regular appearance (Figure 1a). Suppose that the user's task is selecting "menu item 2" from the menu. Hence, "menu item 2" is the target in this example. As requested, the user gazes at the target. Due to a drift in calibration, however, the eye tracker reports a gaze point location within the area of the item immediately below the target (i.e., "menu item 3"). "Menu item 3" is highlighted by a dashed outline to indicate the estimated target (Figure 1b). We will refer to it as the candidate.



a b c d

Figure 1. Dynamic target expansion and calibration correction

After a dwell time *DT*, the candidate expands in the vertical direction. As the expansion occurs, the other menu items shift vertically: those located above the candidate move up, whereas those below it move down (Figure 1c). Note that the candidate's caption stays in the same location. As additional feedback, the expansion is marked by a change in the colours of the candidate's background and caption.

In our example, the candidate's expansion results in the actual target ("menu item 2") shifting up (Figure 1c). Let us denote the magnitude of the shift as ΔH (see Figure 2). In response to this shift, the gaze point moves up by a distance ΔY approximately equal to ΔH. Then, the eye tracker reports a new location for the gaze point, which is now within the area of "menu item 2". As the new candidate, "menu item 2" expands while "menu item 3" shrinks (Figure 1d). Since the new candidate coincides with the target, the gaze makes no further jumps. This completes selection of the required item.

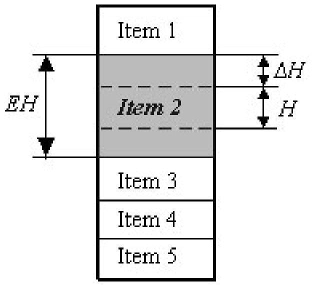


Figure 2. Expansion parameters. H is the height of a menu item ($H = 20$ pixels). EH is the height of the expanded item (candidate target): $EH = H \times EF$, where EF is expansion factor

In our technique, the target's selection process is governed by four parameters: *DT*, *P*, *EF*, and *TH*.

DT is the time the gaze point must dwell on an item to have it expanded. This verifies the user's willingness to initiate selection, as opposed to regular scanning of the screen.

P is the transition period following an expansion during which the gaze is expected to settle down in a new location.

EF is the expansion factor that determines ΔH:

$$\Delta H = \frac{EH - H}{2} = \frac{H \times EF - H}{2} = \frac{(EF - 1)H}{2} \quad (1)$$

Because of inherent eye jitter, the gaze point is never still. Therefore, the stimulus for the eye ΔH must be large enough to allow reliable discrimination between its response ΔY and the jitter. For this purpose, we introduce a threshold *TH*. It is used as a reference to evaluate ΔY before deciding whether to take any corrective action, or not.

The technique can now be generalized as follows. If $\Delta Y < TH$, the candidate is recognized as the target, and selection follows right away. Otherwise, the target is assumed to be located above or below (depending on the sign of ΔY) the candidate. Then, the latter shrinks, while the adjacent item above/below expands. The eye tracker's calibration is adjusted by mapping the estimated gaze point to the current candidate's caption. The process is repeated until the gaze is directed towards the target to select it.

3. Pilot Study

To empirically determine the values for the parameters *P*, *EF*, and *TH*, we conducted a pilot user study. One second was adopted a priori for *DT*.

3.1. Method

3.1.1. Participants

Eight unpaid volunteers (3 male, 5 female) from a local university participated in the study. All participants had prior experience with eye tracking. Five participants wore glasses; three required no correction of vision.

3.1.2. Apparatus

The experiment was conducted on an AMD Athlon 1.3 GHz PC with a 17-inch LCD monitor with a resolution of 1024 x 768. A remote eye tracking system *iViewX*™ from SensoMotoric Instruments was used for collecting gaze data. Eye gaze input and associated events were recorded using experimental software developed in our laboratory.

3.1.3. Procedure

Participants were seated at a viewing distance of approximately 70 cm. Their task was to look at a target (10-by-10-pixel square) presented on the screen. The initial location of the target was chosen randomly within the 512-by-512-pixel central screen area. After a two-second delay, the target jumped upwards a specified distance (*D*). In this new location, it was displayed for another two seconds. The time limit was considered sufficient for the gaze to settle down.

Two measures were recorded for the vertical coordinate of the gaze point: 100 ms before the target's jump Y_B and one second after the jump Y_A. Both measures were not instantaneous values, but averages of five data samples collected over an interval of 100 ms (as determined by the eye tracker's temporal resolution).

3.1.4. Design

The experiment was a 6 x 10 repeated measures factorial design. The factors and levels were as follows:

Distance	0, 10, 20, 30, 40, 50 pixels
Trial	1, 2…10

Note that $D = 0$ serves as a baseline condition as it represents "no jump". This condition was used to obtain an estimate for *TH*.

All conditions were administered as a single block. A block consisted of the 6 *D* conditions presented in random order. For each *D* condition, ten trials were performed. Thus, a block consisted of 60 trials. The conditions above combined with 8 participants resulted in 480 total trials in the experiment.

Two dependent measures were used. The first was the absolute difference between the gaze point vertical coordinates before and after the target's jump ($\Delta Y = | Y_B - Y_A |$). The second dependent measure was reaction time to the target's jump.

3.2. Results

In the baseline condition ($D = 0$), the maximum vertical drift of the gaze point was 20 pixels. In 97% of the trials, however, ΔY did not exceed 15 pixels. Therefore, we adopted this value for the threshold *TH*. If $\Delta Y < 15$ pixels, the gaze drift was treated as a consequence of eye jitter.

For the other *D* conditions, we counted the percentage of the gaze shifts in excess of 15 pixels (Figure 3). Based on these empirical data, ΔY is reliably ($p \geq .99$) identified as a gaze shift for target jumps exceeding 35 pixels. With reference Equation 1, this corresponds to $EF = 4.5$ for the present target height ($H = 20$ pixels).

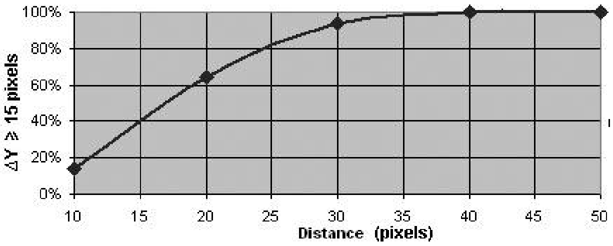


Figure 3. Percentage of cases where $\Delta Y \geq 15$ pixels for each nonzero *D* condition

As for reaction time, it averaged 305 ms (the range was from 220 ms to 400 ms). To have a safety margin, we adopted 500 ms for *P*.

4. Evaluation Study

4.1. Method

4.1.1. Participants and Apparatus

Participants and apparatus were the same as in the pilot study.

4.1.2. Procedure

Participants were seated at a viewing distance of 70 cm. At the onset of each trial, a home box appeared on the screen (left-hand side of Figure 4). It was visible to participants as a 20-by-20-pixel square (solid outline). The actual size of the home box, however, was 120-by-120 pixels (dashed outline). The expansion in motor space facilitated homing through increased tolerance to instabilities in calibration of the eye tracker. On the other hand, making only the central portion of the home box visible ensured bringing the gaze closer to its center.

Participants were first asked to gaze at the home box. After one second, it disappeared. Simultaneously, a menu containing five items appeared on the right (right-hand side of Figure 4). The home box and menu were aligned so that their centres were always on the same hypothetical horizontal axis in the middle of the screen.

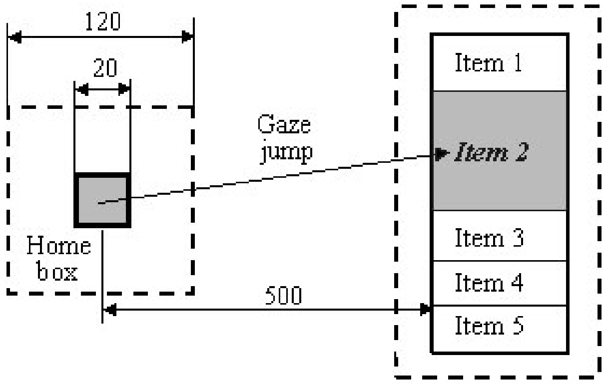


Figure 4. Experimental setup. Home box and menu were not displayed simultaneously; they are shown here together for schematic illustration only. Distances measure in pixels

As with the home box, the menu was also expanded in motor space. That is, the actual size of the menu (dashed outline) was bigger than that displayed (solid outline). The extra areas above and below the visible part (each 30 pixels high) were treated as extensions of the top and bottom menu items, respectively. This was a precaution against a potential loss of the menu when its outside item was to be selected in the presence of a calibration drift. Similarly, the extra areas on the sides of the visible menu were each 30 pixels wide.

When the menu appeared on the screen, participants were instructed to look at the target (the item with a caption in different colour) as soon as possible. Then, the target's selection process proceeded as described in Section 2.

A window of six seconds was given to complete a trial. If no selection occurred within six seconds, a TNC-type (trial not completed) error was recorded. Then, the next trial followed.

4.1.3. Design

Participants completed five blocks of trials (10 trials each). Within each block, the five menu items were presented in random order. All blocks were performed in one session. A short break was made between the blocks if participants needed to rest their eyes. For each participant, the eye tracker was calibrated only at the beginning of the session; no recalibrations were allowed. Total amount of trials was: 8 (participants) x 5 (blocks) x 10 (trials) = 400. No learning effects were expected due to the highly intuitive nature of eye gaze input.

The dependent measures were selection time and error rate.

4.2. Results

As seen in Figure 5, participants completed successfully 91% of the trials. The target was selected without calibration correction 56.4% of the time. Thus, the technique contributed additional 34.6% to the overall success rate.

Of the failed attempts, 2% were beyond the technique's control as they were due to a fault in the hardware (data samples lost by the eye tracker). This resulted in a timeout (TNC-type error).

In 7% of the cases a selection occurred, but the selected item was other than the target. There were two kinds of the technique's failures. 5% of the time it overreacted: the eye tracker's calibration was adjusted even though the estimated gaze point location was correct. There was an opposite situation 2% of the time: the technique was required to adjust the calibration, but it did not.

For correct trials, the grand mean for selection time was 2.65 s. When no calibration correction was involved, it took on average 2.34 s to select the target. Selection time was 3.08 s with a one-step correction, and 3.81 s with a two-step correction.

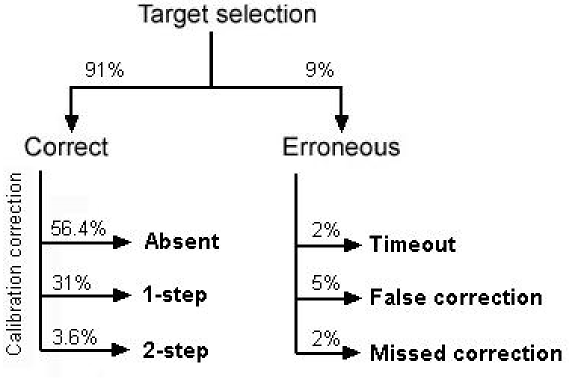


Figure 5. Selection accuracy of menu items

5. Conclusions

Our results indicate that the suggested technique for selection of standard-size menu items allows achieving 91% accuracy. This is a remarkable success rate for eye-based interaction unassisted by any other input modality. Moreover, the level of accuracy yielded by the technique is comparable to that reported for manual pointing with devices such as a laptop's isometric stick [7].

Furthermore, due to the iterative nature of the technique, menu items are selected even in the presence of substantial calibration errors. However, there is a cost which surfaces as a noticeable increase in selection time (24% for a two-step versus one-step correction).

We expect to improve both pointing speed and accuracy by optimising the values of parameters *EF*, *TH*, *P*, and *DT*. Increasing expansion factor *EF* is likely to enhance accuracy as the signal-to-noise ratio increases (i.e., there is a more reliable distinction between gaze response to target expansion and eye jitter). Similar considerations apply to fine-tuning threshold *TH*. For menus with many items, however, the available screen space may be limited. Naturally, a footprint-accuracy trade-off is inevitable then.

In turn, reducing dwell time *DT* and transition period *P* can make the technique faster. Since they both are constant components of total selection time, it is important to keep them as short as possible. Once again, however, a compromise is required: *DT* cannot be made too short to avoid the "Midas Touch" problem (i.e., when an object is selected unintentionally during regular scan of the screen contents). Similarly, reducing *P* should not aggravate selection process because of unfinished transients. Further work is needed to explore the merits of such modifications and the risks involved.

Acknowledgments

This research was supported by the Graduate School in User-Centered Information Technology and by grant #103174 from the Academy of Finland.

References

[1] **A.T. Duchowski.** A breadth-first survey of eye tracking applications. *Behavior Research Methods, Instruments & Computers* 34, 2002, 455-470.

[2] **R.J.K. Jacob.** The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Trans. Info Systems* 9, 1991, 152-169.

[3] **M. McGuffin, R. Balakrishnan.** Acquisition of expanding targets. *Proc. CHI* 2002, *ACM Press*, 2002, 57-64.

[4] **D. Miniotas, O. Špakov, I.S. MacKenzie.** Eye gaze interaction with expanding targets. *Ext. Abstracts CHI* 2004, *ACM Press*, 2004, 1255-1258.

[5] **C. Ware, H.H. Mikaelian.** An evaluation of an eye tracker as a device for computer input. *Proc. CHI+GI* 1987, *ACM Press*, 1987, 183-188.

[6] **S. Zhai, S. Conversy, M. Beaudouin-Lafon, Y. Guiard.** Human on-line response to target expansion. *Proc. CHI* 2003, *ACM Press*, 2003, 177-184.

[7] **S. Zhai, C. Morimoto, S. Ihde.** Manual and gaze input cascaded (MAGIC) pointing. *Proc. CHI* 1999, *ACM Press*, 1999, 246-253.