

# Eye Gaze Interaction with Expanding Targets

Darius Miniotas<sup>1</sup>, Oleg Špakov<sup>1</sup>, I. Scott MacKenzie<sup>2</sup>

<sup>1</sup>Unit for Computer-Human Interaction  
University of Tampere  
FIN-33014 Tampere, Finland  
{dm, oleg}@cs.uta.fi

<sup>2</sup>Department of Computer Science  
York University  
Toronto, Ontario, Canada M3J 1P3  
smackenzie@acm.org

## ABSTRACT

Recent evidence on the performance benefits of expanding targets during manual pointing raises a provocative question: Can a similar effect be expected for eye gaze interaction? We present two experiments to examine the benefits of target expansion during an eye-controlled selection task. The second experiment also tested the efficiency of a “grab-and-hold algorithm” to counteract inherent eye jitter. Results confirm the benefits of target expansion both in pointing speed and accuracy. Additionally, the grab-and-hold algorithm affords a dramatic 57% reduction in error rates overall. The reduction is as much as 68% for targets subtending 0.35 degrees of visual angle. However, there is a cost which surfaces as a slight increase in movement time (10%). These findings indicate that target expansion coupled with additional measures to accommodate eye jitter has the potential to make eye gaze a more suitable input modality.

## Keywords

Eye movements, eye tracking, pointing, human performance

## ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User interfaces – input devices and strategies, interaction styles

## INTRODUCTION

The complexity of modern software with ever-more UI widgets on a display requires a careful approach to screen space management. Recent research has focused on dynamically expanding targets to facilitate pointing within the user’s focus of attention [3, 7]. In other words, iconic targets are expanded to a “pointing-friendly” size when the user needs to interact with them; otherwise they appear in a reduced size. Consequently, this is one solution to the problem of limited screen space.

Empirical evidence that dynamic target expansion speeds up performance in a point-select task was first reported in [3]. Furthermore, there is an improvement even when

expansion occurs only after 90% of the movement distance is covered. A subsequent study [7] validated these findings even where target expansion was randomized along with shrinking and static targets.

The problem of limited screen space is even more severe in eye gaze interaction. In the applied eye tracking literature, it is accepted that eye gaze accuracy is limited to one degree of visual angle [2]. This limitation is dictated by the size of the fovea – the portion of the retina providing high acuity vision of the object of current interest. As a result, targets must subtend at least one degree of visual angle for sufficiently reliable pointing with an eye tracker.

The findings on expanding targets acquired by conventional pointing devices encourage investigation on whether target expansion is also beneficial with other pointing devices. Caution is warranted, however, when applying the idea to eye gaze interaction due to fundamental differences.

First, for eye gaze, the benefits of *dynamic* expansion are arguable due to the jumpy nature of eye movements. An object of interest is only viewed during a short period of relative stability called a *fixation*. Fixations are connected by *saccades* – sudden motions of the eye – that allow navigation between objects of interest. These motions are ballistic, meaning the destination is known before the onset of movement (see [2] for more details). Since visual feedback is not processed during a saccade, little may be gained from dynamically expanding targets.

Second, during a fixation the eye does not stay still. Instead, it makes micro movements to allow visual perception of the scene. Given this jittery behavior, target expansion is potentially a distraction. Finally, in a gaze-controlled interface, the cursor is redundant, since the gaze point acts as the pointer.

The implications of the above are that static target expansion seems more reasonable for eye gaze interfaces. By “static”, we mean the region of expansion is determined a priori and that the expansion is not visually presented to the user. In other words, the interface responds to gaze point within the boundaries of the expanded target area, even though the target’s appearance does not change.

To test the efficiency of static target expansion, we conducted two experiments involving a simple point-select task with movement time and error rate as performance

indicators. In the first experiment, an algorithm developed to accommodate inherent eye jitter was tested. The second experiment aimed to reveal the effect of abandoning this aid.

## METHOD

### Participants

The same twelve un-paid volunteers (8 male, 4 female) participated in both experiments. All were students at a local university and had normal or corrected vision. Four of the participants had prior experience with eye tracking technology.

### Apparatus

Both experiments were conducted on a Pentium III 500 MHz PC with a 17-inch monitor with a resolution of 1024 x 768. A head-mounted eye tracking system *EyeLink*<sup>TM</sup> from SensoMotoric Instruments served as the input device. The participant PC was connected to another PC (Celeron 466 MHz) for analysis of the captured eye images.

### Procedure

#### Task

Participants were seated at a viewing distance of  $\approx 70$  cm. Both experiments used a simple point-select task (Figure 1). At the onset of each trial, a home box appeared on the screen. It was visible to participants as a 20-by-20-pixel square (solid outline in Figure 1). The actual size of the home box, however, was 120 x 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 the center of the box.

Upon fixating on the home box for one second, a rectangular target appeared in the peripheral field of view. Participants were instructed to look at the target as quickly as possible (timing started), and fixate upon it until selection (timing ended). A window of three seconds was given to complete a trial. If no selection occurred within three seconds, a TNC-type (trial not completed) error was recorded. Then, the next trial followed.

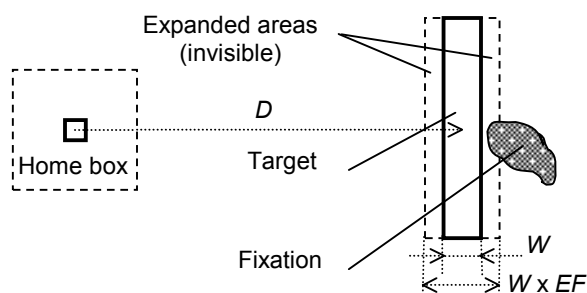


Figure 1. Experimental task

### Experiment 1: Grab-and-Hold Algorithm Introduced

As revealed in our pilot tests, selecting targets as narrow as 20 pixels was not always straightforward. Quite often, only a small fraction of the gaze points belonging to the same fixation entered the target (see Figure 1 for illustration). As mentioned before, this effect is due to inherent eye jitter. To handle this, we introduced a simple grab-and-hold algorithm (GHA). The algorithm works as follows.

Upon appearance of the target, there is a settle-down period of 200 ms during which the gaze is expected to land in the target area and stay there. Then, the algorithm filters the gaze points until the first sample inside the expanded target area is logged. When this occurs, the target is highlighted and the selection timer triggered. The selection timer counts down a specified dwell time ( $DT$ ) interval.

The target is selected irrespective of the actual location of the gaze point at the moment of the  $DT$  expiry, provided no interruptions (i.e., interspersing saccades) occurred in the fixation throughout the  $DT$  interval. Thus, the gaze is virtually held on the target once it is “grabbed”. This way some intelligence is added to the interpretation of the eye tracker data: the gaze point is allowed to deviate from its intended destination as long as this deviation does not extend beyond the boundaries of the current fixation.

If the eye makes a saccade before the end of the  $DT$  countdown, however, the target is de-highlighted resetting the selection timer. Then the algorithm starts hunting for the next gaze point in the expanded target area, and the process is repeated.

### Experiment 2: Grab-and-Hold Algorithm Disengaged

In the absence of the algorithm, the target is highlighted whenever the gaze is over it. A highlight starts the selection timer. Selection occurs only if the gaze does not leave the expanded target area for the duration of the  $DT$  interval. If an exit occurs during this interval, the target is de-highlighted, and the selection timer resets, starting the countdown for a new  $DT$ . The process is repeated until either the gaze meets the stringent no-quit criterion for target selection, or the three-second time limit expires.

## Design

### Experiment 1

The experiment was a 3 x 4 x 3 x 3 x 3 repeated measures factorial design. The factors and levels were as follows:

Dwell time	750, 1000, 1250 ms
Direction	left, right, up, down
Distance ( $D$ )	128, 256, 512 pixels
Width ( $W$ )	12, 24, 36 pixels
Expansion Factor ( $EF$ )	1, 2, 3
Trial	1, 2, 3

Note that  $EF = 1$  serves as a baseline condition as it represents “no expansion”. Although no learning effects were expected due to the highly intuitive nature of eye-gaze

based pointing, participants were still randomly assigned to one of three groups. Each group received the dwell time conditions in a different order using a Latin square.

For each *DT* condition, participants performed 12 blocks of trials (3 blocks per movement direction) in one session. The three sessions were run over consecutive days with one session lasting approximately half an hour. Each block consisted of the 27 *D-W-EF* conditions presented in random order. For each *D-W-EF* condition, 3 trials were performed. The trial for any condition, however, was not repeated within the same block, but was administered in a separate block to allow resting the eyes. Thus, a block consisted of 27 trials. The conditions above combined with 12 subjects resulted in 11,664 total trials in the experiment.

The 27 *D-W-EF* conditions were chosen to cover a range of task difficulties spanning 1.13 to 5.45 bits, according to Fitts' index of difficulty:

$$ID = \log_2(D/W+1)$$

The dependent measures were movement time (*MT*) and error rate (*ER*).

### Experiment 2

This experiment used only one *DT* condition (1250 ms) and did not use the grab-and-hold algorithm. Otherwise, the design was identical to Experiment 1. Thus, 3888 trials were completed in Experiment 2. This experiment primarily served to gage the effect of the grab-and-hold algorithm. Since the algorithm was absent in the second experiment, any learning effect from Experiment 1 to 2 tends to bias results *against* the algorithm. Thus, actual performance benefits in using the algorithm should be as good as, or better than, those observed.

## RESULTS

### Experiment 1

The grand means on the two dependent measures were 1672 ms for *MT* and 9.9% for *ER*. The main effects and interactions on each dependent measure are presented below.

#### Speed

As expected, the 750-ms *DT* condition was the fastest with a mean *MT* of 1444 ms. The 1000-ms *DT* condition was slower by 17% (1683 ms), and the 1250-ms *DT* condition by 31% (1887 ms). The differences were statistically significant ( $F_{2,22} = 191.0, p < .0001$ ). The main effect for *EF* was also significant ( $F_{2,22} = 73.4, p < .0001$ ), as was the *DT* × *EF* interaction ( $F_{4,44} = 4.2, p < .05$ ). The main effects and interaction are illustrated in Figure 2.

#### Accuracy

The lowest *ER* was in the 1000-ms *DT* condition (9%). It was followed by the 750-ms condition at 10% errors, and the 1250-ms condition at 10.9%. The differences were not significant ( $F_{2,22} = 1.8, ns$ ). The main effect of *EF* on *ER*, however, was significant ( $F_{2,22} = 101.0, p < .0001$ ).

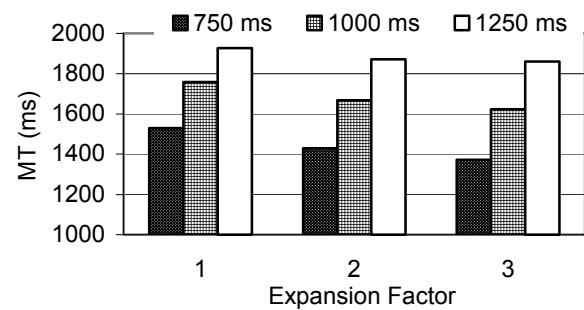


Figure 2. *MT* vs. *EF* for the three *DT* conditions

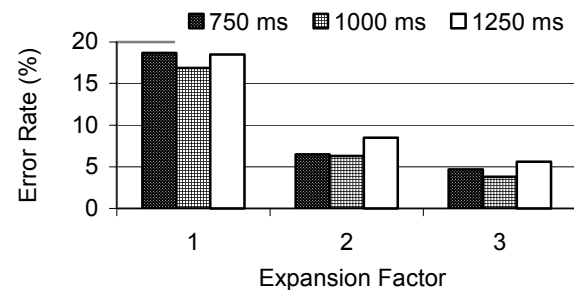


Figure 3. *ER* vs. *EF* for the three *DT* conditions

The *DT* × *EF* interaction was not significant ( $F_{4,44} = 1.3, ns$ ). The main effects and interaction are illustrated in Figure 3.

#### Speed-Accuracy Tradeoff

Regression analyses showed that the data for any of the three *DT* conditions did not fit the Fitts' law equation [1] very well, with  $r^2$  values falling just under 0.7. This is in contrast to the finding in [4], where the equation accounted for more than 98% of the variability in the data. In that study, however, the mouse cursor was visible as a feedback of the current gaze point location, whereas no mouse cursor was present in this study. The presence of the cursor might have influenced the strategy of performing pointing movements.

Despite the relatively low correlation, however, the slope of the linear regression line (99 ms/bit for the 750-ms *DT* condition) was consistent with the finding of [7] obtained for pointing at expanding targets with a puck on a tablet. This is a clear demonstration that a speed-accuracy tradeoff still takes place during eye gaze interaction, even in the absence of the pointer.

### Experiment 2

As seen in Figure 4, the *MT*s were slightly lower in Exp 2. Thus, the GHA bears a slight cost in terms of speed. Overall, *MT* was 10% higher in Exp 1 (1887 ms vs. 1722 ms), which included the algorithm. However, this is substantially offset by a dramatic reduction in errors (Figure 5). The overall error rate in Exp 2 (25.6%) is contrasted with the much lower error rate in Exp 1 (10.9%).

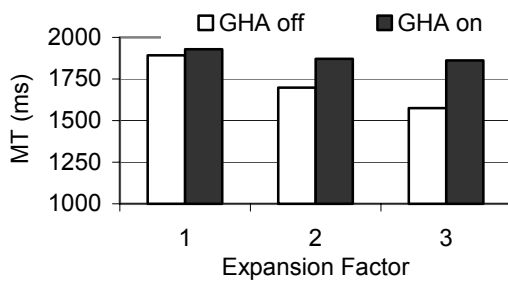


Figure 4. MT vs. EF for the two conditions

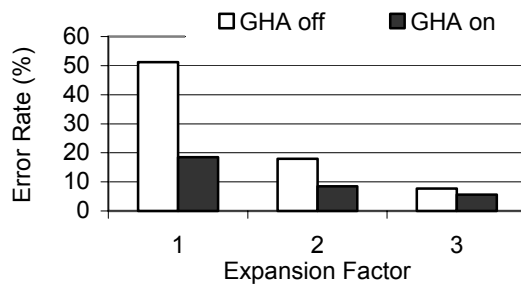


Figure 5. ER vs. EF for the two conditions

This translates into an error rate reduction of 57% with the addition of the grab-and-hold algorithm.

The impact of the algorithm on accuracy is particularly apparent when error rates are plotted against the effective target width, i.e.,  $W \times EF$  (Figure 6). For the smallest target width without expansion (i.e., 12 pixels, corresponding to 0.35 degrees of visual angle), there was a 68% reduction in ER when the algorithm was turned on. Facilitation was also observed for effective target widths of 24 and 36 pixels, the Student-Newman-Keuls pair-wise differences being quite reliable ( $p$ 's < .001). Meanwhile, for effective  $W \geq 48$  pixels, the algorithm's effect was not significant.

At the effective width of 48 pixels (1.4 degrees), ER did not exceed 8% in both the conditions. This finding is consistent with that of [6]. Employment of the grab-and-hold algorithm yields error rates under 10% even for a target as narrow as 12 pixels with a threefold expansion.

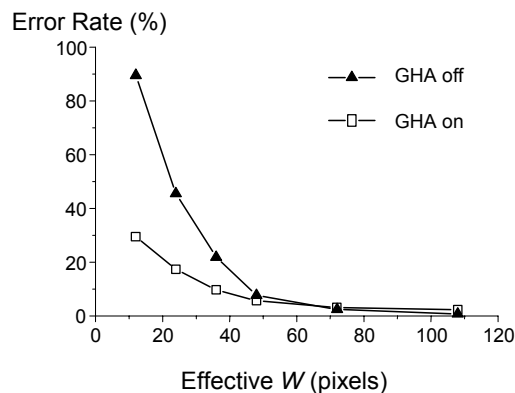


Figure 6. ER vs. effective  $W$  for the two conditions

## CONCLUSION

Our results indicate that target expansion in motor space (i.e., invisible to the user) facilitates pointing both in terms of speed and accuracy. Even though the spatial cost is permanent, the space occupied by the expanded areas is still available for non-interactive objects.

Moreover, the limited accuracy of eye gaze as an input technique is amenable to techniques that increase tolerance to the inherent eye jitter. As our evidence suggests, adding some intelligence to dwell-time based selection brings eye gaze input one step closer to supporting interactions with standard GUI widgets, such as scrollbars and pull-down menus. Previous findings in eye gaze control suggest that such targets are just too small for facile interaction. However, we believe that novel approaches, such as our grab-and-hold algorithm, might help in ultimately extending applications for eye-based systems.

More work is needed before eye gaze interaction enters more realistic settings involving numerous objects. A more sophisticated algorithm will be required for handling eye jitter under constraints imposed by multiple expanding targets that are close to one another. In the future, we also intend to supplement our grab-and-hold algorithm with an eye drift correction technique similar to that suggested in [5].

## ACKNOWLEDGEMENTS

This research was funded in part by grant #103174 to Miniotas from the Academy of Finland. Support from the Academy of Finland (project 53796) for MacKenzie is also greatly appreciated.

## REFERENCES

1. Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *J Exp Psych*, 47 (1954), 381-391.
2. Jacob, R. J. K. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Trans Info Systems* 9, 3 (1991), 152-169.
3. McGuffin, M. and Balakrishnan, R. Acquisition of expanding targets. *Proc CHI 2002*, ACM, 57-64.
4. Miniotas, D. Application of Fitts' law to eye gaze interaction. *Ext Abstracts CHI 2000*, ACM, 339-340.
5. Stampe, D. M. and Reingold, E. M. Selection by looking: A novel computer interface and its application to psychological research. In J. M. Findlay et al. (Eds.), *Eye Movement Research*, Elsevier, 1995, 467-478.
6. Ware, C. and Mikaelian, H. H. An evaluation of an eye tracker as a device for computer input. *Proc CHI+ GI 1987*, ACM, 183-188.
7. Zhai, S., Conversy, S., Beaudouin-Lafon, M., and Guiard, Y. Human on-line response to target expansion. *Proc CHI 2003*, ACM, 177-184.