# Improving Eye Cursor's Stability for Eye Pointing Tasks

**Xinyong Zhang**†, **Xiangshi Ren**†*, **Hongbin Zha**‡
† Dept. of Information Systems Engineering, Kochi University of Technology, Kochi, Japan
* DUB Group, University of Washington, Seattle, USA
‡ State Key Lab of Machine Perception, Peking University, Beijing, China
zxybit@163.com, ren.xiangshi@kochi-tech.ac.jp, zha@cis.pku.edu.cn

## ABSTRACT

In order to improve the stability of eye cursor, we introduce three methods, force field (FF), speed reduction (SR), and warping to target center (TC) to modulate eye cursor trajectories by counteracting eye jitter, which is the main cause of destabilizing the eye cursor. We evaluate these methods using two controlled experiments. One is an attention task experiment, which indicates that both FF and SR significantly alleviate the instability of eye cursor, but TC is not as we anticipated. The other is a 2D pointing task experiment, which shows that FF and SR as well as the improved implementation of SR (iSR) indeed improve human performance in dominant dwell-based eye pointing tasks of eye-based interactions. The method iSR is especially effective to accelerate eye pointing (10.5% and 8.5%) and reduce error rate (6.1% and 2.7%) when target diameter $D$ = 45 and 60 pixels.

## ACM Classification Keywords

H5.2. User Interfaces: Input devices and strategies.

## Author Keywords

Eye pointing, eye cursor, eye jitter, stability.

## INTRODUCTION

Since Bolt [5] introduced the possibility of using the eyes in multimodal interfaces more than 25 years ago, lots of researchers have explored the feasibility of eye-based interactions in the field of human-computer interaction (HCI) [14, 6]. Using an eye tracking system (eye tracker), the position that the user is looking at on the screen can be detected and used as a native "pointer". Thus, eye tracker can be used as a kind of augmented input device or a kind of direct pointing device. Eye-based interaction systems are especially useful for quadriplegics because it provides them with a unique opportunity to communicate with the rest of the world.

However, eye-based interactions are still problematic [28, 9, 10, 4, 3], because of issues like the "Midas-Touch" problem [9], lack of analogous functions for single-clicking and double-clicking etc., limitations in tracking accuracy, calibration errors and drift. Another crucial problem is the inherent jittery motions of the eyes. When we humans intentionally fixate on a static object, our fixations actually consist of three types of involuntary movement (tremors, drifts and microsaccades) [1]. Therefore, eye cursor controlled by noisy gaze input signals cannot be accurately fixed on a single point like a mouse cursor. This stability problem probably causes eye cursor to momentarily leave the desired target, thus interrupting dwell time continuity, which is the most common protocol used for command activation [28, 7].

In this paper, therefore, we mainly concentrate on the stability issue of eye cursor. We introduce three methods, force field (FF), speed reduction (SR), and warping to target center (TC), to redress eye cursor trajectories. We evaluate these methods using an attention task experiment, which shows us that both FF and SR significantly counteract eye jitter, enhancing the stability of eye cursor, but TC is not as we anticipated. The details of this experiment also provide insights into the optimization on FF and SR. Another 2D pointing task experiment further revealed that FF and SR as well as the improved implementation of SR (iSR) indeed improve human performance in eye pointing tasks.

## RELATED WORK

To improve the feasibility of eye-based interactions, there have been a number of methods, which can be divided into four categories as follows.

### Integrating with Other Input Modalities

Researchers in the HCI field have proposed different multimodal systems integrated with eye gaze input to overcome problems such as the "Midas-Touch" problem. One kind of multimodal system is the combination of gaze with a hardware button [28, 29, 11]. For example, Zhai et al. [29] proposed the MAGIC technique for pointing tasks. This MAGIC pointing technique firstly warps the cursor to the vicinity of the target when the user stares at and wants to select it. Then the user can use a manual pointing device to finally confirm the selection. Kumar et al. [11] recently developed a prototype system, $EyePoint$, which incorporated gaze input with keyboard input using a fluid look-press-look-release action. The user can look at the desired target in the screen and press a pre-defined hotkey for the desired action, then in a magnified squared view of the region related to the user's gaze, the user can stare at the target again and release the hotkey to finally perform the desired action.

*CHI 2008*, April 5-10, 2008, Florence, Italy.

Another kind of multimodal system is the fusion of gaze and speech [30, 17]. To remove the ambiguities of speech recognition, for example, Zhang et al. [30] defined a screen region that depended on the user's gaze position to refine the $n$-best list to limit the vocabulary of speech recognition, decreasing the probability of ambiguities. If more than one verbal command, associated with a specific screen target, has been recognized as one of the $n$-best candidates, only the one whose associated target is nearest to the gaze position will be chosen as the most desired. Miniotas et al. [17] proposed a similar method for the selection of small closely spaced targets. In this method, a number of neighboring targets will be highlighted in different colors. The user can then select the desired one by verbally announcing its color.

Furthermore, Surakka et al. [26] proposed a multimodal system which integrated gaze with facial muscle activation. Their method employed a voluntary frowning action of the facial muscle as the counterpart for the mouse click to finally activate target selection when the user is looking at it.

### Adding Intelligence to Eye-Controlled Interfaces

The gaze position reported by eye tracker probably does not exactly correspond to where the user is looking due to eye tracker's accuracy limitations and calibration errors. One solution is to infer the user's actual intention in a given context even if the user's gaze points as reported by eye tracker are not accurately mapped on the desired target. Salvucci [20] proposed a fixation tracking algorithm which employed hidden Markov models to interpret the user's eye movements as the most intentional fixation sequence. This method did not necessitate a dwell time to confirm each target (letter) or a looser spacing between targets to accommodate the relatively lower precision of eye trackers. It was suitable for eye typing as the user could quickly move his/her eyes to sequentially glance at some letters in the on-screen keyboard to output the most likely word that the user desired. Cooperating with Anderson, Salvucci [21] further applied a Bayesian probabilistic model to determine the target with the highest probability as the desired target when a gaze point has been observed. This approach was based on a two-dimensional Gaussian distribution, which expresses the probability that the eye gaze can be observed at a certain point when the user is staring at a specific target, and a prior selection probability distribution of different targets.

Ward et al. [27] integrated a language model into an eye typing system, called $Dasher$, to dynamically predict which letter is the next possible one after several letters in a word have been "input" by gaze direction. After the language model worked well, the user of $Dasher$ could fast input words by naturally searching and navigating in a shelf-space on the screen where the possible letters were arranged according to the prediction results.

### Designing Novel Eye-controlled Interfaces

Enlarging targets to make them big enough is another solution that addresses the accuracy issue of eye gaze interaction in the acquisition of small targets. When the user looks at the screen, the area immediately around the gaze point can be linearly magnified and redisplayed in a zoom window [4, 13, 11]. The purpose of employing a magnified view is to accommodate the accuracy problem of the eye tracker so as to exactly map gaze points to a desired target. As a consequence, the desired actions can be easily and correctly performed in the zoom window [13].

The area immediately around the gaze point can also be nonlinearly distorted as if it were covered by a fisheye lens. The work of Ashmore et al. [3] indicated that distorting the gaze area around the desired target with a fisheye lens could benefit the performance of eye pointing provided the fisheye lens is not always slaved to the user's gaze.

Another way to magnify a target is to temporarily expand the target itself rather than zooming the area around it. When the user's gaze falls on the vicinity of the desired target, the target's size could increase enough to involve the gaze point into the enlarged target. Miniotas et al. [16] verified target expansion's performance for eye gaze interaction. Špakov and Miniotas [23] subsequently applied this approach to vertical menu selections. Their technique dynamically expanded the gaze-dwelled items as the candidates for selection to facilitate eye pointing in menu manipulation.

### Adding Special Algorithm to Counteract Eye Jitter

The accuracy problem of eye gaze interaction mainly resulted from the limitation of visual angle, calibration errors, drift, and inherent eye jitter. The approaches reviewed above did not concentrate on eliminating any concrete source of the accuracy problem but treated eye gaze interaction as a whole. With respect to eye jitter, however, some researchers specifically pointed out the necessity of diminishing the effects of eye jitter to improve the stability of eye cursor by a smoothing algorithm [13, 3, 8, 11, 12]. For example, $EyeDraw$ [8] used the average location of every six successive gaze points as eye cursor's position. Kumar et al. [12] proposed a one-sided triangular filter to calculate the fixation point as a weighted mean of the point set in the current fixation window. They also applied two Kalman filters to process the gaze data, one for the entire raw gaze data, and the other only for the data within fixation windows.

However, few researchers except for Miniotas et al. [16] systematically presented their algorithm and evaluated its effectiveness for eye pointing tasks. The algorithm proposed by Miniotas et al. was called the grab-and-hold algorithm (GHA). GHA had the same effect as if the gaze was held on the desired target during periods of fixation, thus effectively reducing the probability of restarting the selection timer before the end of the dwell time. Therefore, dwell-based eye pointing could benefit from this algorithm. This is quite similar to the goal of our work. We will describe our methods in the following section.

## EYE CURSOR REDRESSMENT METHODS

We initially considered three methods, including force field, speed reduction and warping to target center, to modulate eye cursor's trajectories. Here, we describe them in detail.

### Force Field

Originally, Ahlström [2] implemented the idea of force field in cascading pull-down menus to improve the steering of mouse cursor. Within a parent item, for example, the user generally steers the cursor toward the little arrow in the right side to navigate in the submenu and select the desired item. Using a virtual force field on the parent item with a force point at the position of the arrow, the cursor was accelerated in the rightward direction but damped in the leftward and vertical directions. Thus, the force point behaved like a magnetic field as it attracted the mouse cursor.

Ideally, jittery eye movements should not cause the eye cursor to escape from effective contact with the target that has been gazed at by the user i.e. jittery eyes should not interrupt dwell time. We hypothesized that the force field method will be suitable for this purpose. When the force point is located on the target center, the centripetal movements of the eye cursor can be reinforced, whereas the centrifugal movements can be weakened. The warping algorithm of the eye cursor, being similar to that of the mouse cursor for menu selections, can be expressed as follows:

$$\vec{n} = \vec{a} + s \cdot \|\vec{a} - \vec{p}\| \cdot \frac{\vec{f} - \vec{a}}{\|\vec{f} - \vec{a}\|} \qquad (1)$$

where, the four symbols, $\vec{n}$, $\vec{a}$, $\vec{p}$ and $\vec{f}$, denote the vectors pointing to the new position of the eye cursor after being modulated, the current gaze location on the screen, the previous position of the eye cursor and the position of the target center (force point), respectively. The symbol $s$ is a coefficient reflecting the strength of the force field.

Figure 1a illustrates the relationship between those four vectors. It is clear that the new position of the eye cursor is closer to the target center than the current gaze location.

### Speed Reduction

The cursor speed reduction method significantly decreases the likelihood that eye jitter will cause the cursor to temporarily leave the target, interrupting dwell time continuity. This is similar to reducing control-display (C-D) gain in manually controlled interfaces. Thus, we also hypothesized that reducing the speed of the eye cursor can be helpful in keeping the eye cursor in the desired target.

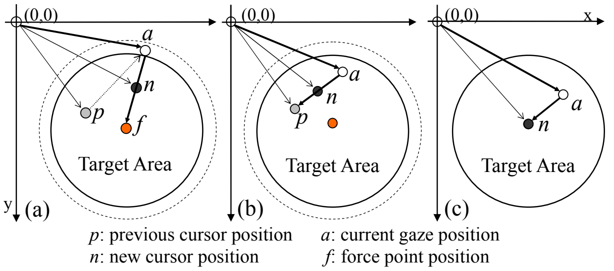

**Figure 1. Cursor redressment principles. (a) Force field, (b) Speed reduction, and (c) Warping to target center. For the first two methods, it looks as if there is a virtual force drawing the eye cursor back toward the force point (FF) or the previous position (SR) even if the gaze point is slightly outside of the target.**

To reduce the eye cursor's speed, the cursor should not be directly warped to the current gaze location but drawn back a certain distance from the current gaze location toward the previous position of the eye cursor. This process can be expressed by a vector formula as follows:

$$\vec{n} = (1 - r) \cdot \vec{a} + r \cdot \vec{p} \qquad (2)$$

where, the three symbols, $\vec{n}$, $\vec{a}$ and $\vec{p}$, have the same meanings as those in Equation 1. The symbol $r$ denotes the ratio of speed reduction. Figure 1b depicts the spatial relationship between those three vectors.

### Warping to Target Center

This method directly warps the eye cursor and holds it on the target center once an observed eye gaze enters the target that the user is dwelling on. This will probably make the user feel that the eye cursor is exactly controlled on the desired target. We suppose a fixed cursor within the target probably can improve the behavior of the eyes, progressively resulting in less deviation of the eye cursor from the desired target.

The condition for the first two algorithms to run is that the eye cursor (i.e. point $n$) is inside the target, while the condition for the the third algorithm is that the current eye gaze (i.e. point $a$) falls on the target. Sometimes, when point $a$ is outside of the target, but point $n$ is still inside the target (see Figure 1a). All these three algorithms are executed at the frequency of 50 $Hz$, which is fast enough to move the cursor. As eye movements are ballistic and very fast [28], these algorithms are easily suspended and resumed when the subject moves the eyesight from one target to another.

## EXPERIMENT 1: EVALUATION WITH ATTENTION TASKS

We conducted an attention experiment to verify if these three algorithms really improve eye cursor's stability.

### Apparatus

The experiment was conducted on a 2.7 $GHz$ Pentium$^{TM}$ IV PC running Windows $^{TM}$ XP with a 19-inch CRT display at 1024×768 resolution. A head mounted eye tracking system, $EyeLink$ II [24], served as the eye gaze input device. $EyeLink$ II was installed in a 700 $Hz$ Pentium$^{TM}$ III PC, which processed the captured eye movement videos and transferred the calculated gaze data to the experiment computer almost in real time by an Ethernet link. $EyeLink$ II worked in pupil only mode at the sampling rate of 250 $Hz$.

### Participants

Eighteen subjects (4 female, 14 male), temporarily recruited from a local university campus, successfully performed the experiment. Subjects ranged in age from 19 to 33 with an average age of 23. Among them, 12 subjects had normal vision without any correction, one used contact lenses and the rest wore glasses. None of the subjects was colorblind. All were experienced computer users, but had no prior experience using an eye tracker.

### Experimental Task and Procedure

Subjects were seated about 70 cm from the display without a head-rest. They were told to adjust the height or position of

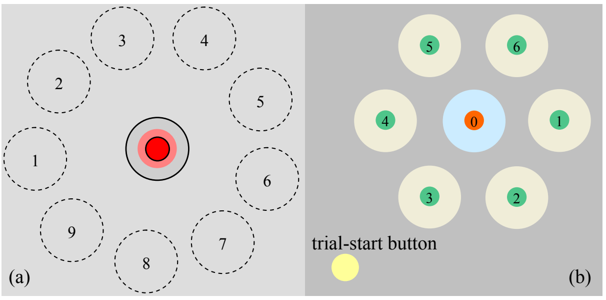

**Figure 2. (a) Experimental interface for attention tasks. Note that the numbered dashed circles illustrate the order of positions for the target. (b) Experimental interface for pointing tasks.**

the chair by themselves if they could not look horizontally at the center of the vertical display. The experimenter briefly introduced the principles of eye tracking, making them feel comfortable as some of them worried about possible harm to the eyes. Then the experimenter demonstrated the experimental interface and explained the task to familiarize them with the experiment. After correctly setting up the eye tracking device and performing a 9-point calibration (sometimes repeatedly to obtain a lower acceptable calibration error) for each subject, the experimenter opened the experimental interface to formally begin the experiment.

As Figure 2a shows, at the beginning for each target size, a round target appeared at the center of the screen. Then, the experimenter sequentially moved the target to 9 positions averagely arranged on a layout circle of 134.5 mm diameter. When the target appeared in each position from 1 to 9, subjects were explicitly instructed to focus on the target center, a small red solid circle of 5.3 mm diameter for a short duration of 7 seconds as a trial, but not to chase the cursor. At the same time, two feedback modes, a semi-transparent red bubble expanding from the target center to the edge, or shrinking contrarily, were alternatively presented at different positions to indicate the elapsing of the short period. At the end of the current trial, the target center changed to gray with the word “Next” on it. Before moving the target to the next position, the experimenter pronounced “Next” to make subjects aware that the next trial would begin. For each trial, “entering target event” ($ETE$) of the eye cursor and the corresponding times when it entered and left the target were recorded.

After being moved around from positions 1 to 9, the target would be relocated at the center of the screen with an increased diameter. This position was not used for trials but for drift correction, where the experimenter generally suspended the experiment to perform a 1-point recalibration to compensate for possible headband slippage, muscle tremor, or environmental vibration [24]. At other positions, if the calibration obviously deteriorated due to subject’s occasional head or body movements, the experimenter would also perform a 1- or 9-point recalibration and then repeat the corresponding trial. We allowed subjects to blink their eyes naturally but not continually. Subjects could have a rest including taking off the helmet whenever they wanted to. With rest times, the experiment lasted approximately one and a half hours.

### Design

The experiment was a repeated measures within-subject design. The independent factors and levels were as follows:

- Four kinds of eye cursor redressment methods ($CRM$): force field (FF), speed reduction (SR), warping to target center (TC), and no method (NM). The $4^{th}$ method means that the cursor was directly placed at the eye gaze location on the screen without any other modulation. This was used as the control condition to which the other three methods were compared.
- Five target diameters ($D$): 35, 45, 60, 80, and 100 pixels. In the experimental setup, these diameter sizes were approximately equal to 11.5, 14.8, 19.8, 26.4, and 33.0 mm, respectively, or corresponded to about 0.94, 1.21, 1.62, 2.16, and 2.70 degrees of visual angel, respectively.
- Three feedback modes: expanding bubble, shrinking bubble and no feedback.

A fully crossed design resulted in 60 combinations. Subjects performed three trials for each combination. The experiment was divided into 4 sections by cursor redressment method. For each method, the trials with the same target diameter were grouped together, consisting of 9 trials (3 feedback modes $\times$ 3 repeated trials) respectively performed at 9 positions as Figure 2a shows. That is to say, three feedback styles rotationally appeared in the order of no feedback, expanding bubble feedback and shrinking bubble feedback for 3 times among those 9 positions. In the methods of FF and SR, three different parameters (0.6, 0.8, and 0.9 for the coefficient $s$ of Equation 1, 0.2, 0.6, and 0.8 for the coefficient $r$ of Equation 2) were set respectively in the 3 repeated trials for each combination. The trial groups appeared in ascending order of $D$ from 35 to 100 pixels.

### Results

#### *Eye cursor’s entering target event*

When the subject was staring at the target, the eye cursor probably entered and left the target frequently as the user was unable to hold it steadily in the desired position. Thus, the number of times $ETE$ happened was the main dependent measure, reflecting eye cursor’s stability. Using repeated measures ANOVA, we found that there was a significant main effect for $CRM$ ($F_{3,51} = 21.90, p < .001$), target diameter ($F_{4,68} = 33.39, p < .001$), but no significant main effect for feedback mode ($F_{2,34} = .01, p = .992$) on $ETE$. We did not find any other interaction effects on $ETE$.

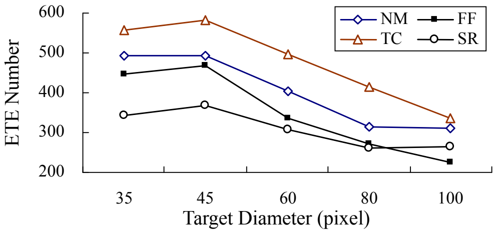

**Figure 3. Sum of $ETE$ numbers by $D$ for different methods.**

The overall numbers of $ETE$ were 1745, 1544, 2385, and 2018 for the methods FF, SR, TC, and NM, respectively. Post hoc pair-wise comparison tests indicated that the mean number of $ETE$ under the condition of SR was significantly smaller, all at the $p < .005$ level, while that under the condition of TC was significantly bigger than the others, all at the $p < .05$ level. Moreover, FF also led to a significantly smaller $ETE$ number than NM ($p = .001$). With respect to $D$, the overall average $ETE$ number was progressively decreased from the level of 35 to the level of 100 pixels, except that there was no significant difference between the levels 35 and 45 pixels ($p = .456$). The difference between the levels 80 and 100 pixels was statistically significant but not very strongly ($p = .031$). We performed further pair-wise comparison tests on the average $ETE$ numbers for $CRM$ according to different $D$ levels. As seen in Figure 3, SR produces significantly smaller $ETE$ numbers than NM at all levels. FF also does this but it is significant only at the level of 100 pixels. Unfortunately, the method TC led noteworthily to sometimes significantly bigger $ETE$ numbers than NM at all levels. This was not as we had anticipated.

### *Frequency Distribution of Eye Cursor's Dwell Time*

Besides the $ETE$ numbers, the number of times the eye cursor could dwell on the target for a long or short time after entering the target also reflected its stability. For each factor combination, therefore, we counted the frequency distributions of dwell time in the given bin ranges (see Figure 4) and then normalized them to percentage distributions as a kind of multivariate repeated measure (9 dimensions) for the stability evaluation. An ANOVA for multivariate repeated measures revealed a strong main effect for $D$ ($F_{36,252} = 2.82, p < .001$), $CRM$ ($F_{27,135} = 3.30, p < .001$), and a significant interaction effect between them ($F_{108,1836} = 1.37, p < .01$) on the multivariate measure, frequency distribution of dwell-time. No other main effect or interaction effect was found.

Univariate tests for each dimension, i.e. for the percentage of dwell time's frequency in each range, further showed a significant main effect for $D$ in the whole 9 ranges, all at the level of $p < .001$ except for the $7^{th}$ at $p < .01$ and the $8^{th}$ at $p < .05$, while for $CRM$ in the first four ranges, the $7^{th}$ and the last range, all at the level of $p < .05$ but the first and the last at the level of $p < .001$. The corresponding interaction effect existed in the $5^{th}$ and the last three ranges ($p < .05$). There was no a significant main effect for feedback mode or any other interaction effect.

Post hoc pair-wise comparison tests indicated that the mean dwell time frequency in the first five ranges was generally decreased with the increase of $D$ and the decrease was significant in most cases. On the contrary, the mean dwell time frequency was significantly ascending in the last range. With respect to $CRM$, both FF and SR, compared with NM, strongly decreased the mean dwell time frequency in the first range ($p < .001$), while significantly increased it in the last range ($p < .005$). However, TC was not significantly different to NM in all ranges but the last one with a significant lower frequency ($p < .05$). As Figure 4 illustrates, these results imply that both FF and SR can effectively reduce the possibility that the eye cursor would leave very quickly from the desired target because of the eyes' jittery movements, while increase the possibility of keeping the eye cursor in targets for longer, even until the trail end.

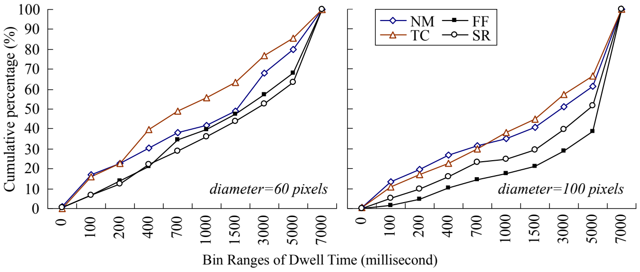

**Figure 4. Cumulative percentage of dwell time frequency in the given bin ranges for different methods when $D$ = 60 and 100 pixels. The situations under other $D$ conditions were not presented due to the page limitation, but they are similar to the situation presented in this figure.**

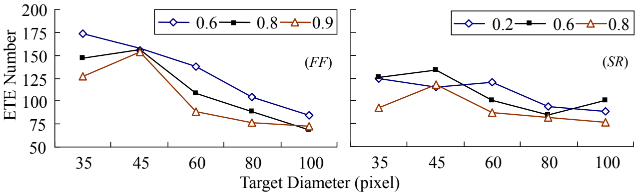

**Figure 5. Sum of $ETE$ numbers by $D$ for different parameters of the method FF and SR.**

### *Effects of Parameter Factor in Both Methods FF and SR*

With respect to the coefficient $s$ or $r$ in the algorithm of FF or SR, three different parameters had been applied in experimental trials to verify whether there was room to optimize both methods. Repeated measures ANOVA, based on the separate data of FF and SR, revealed that the parameter factor had a significant effect on the $ETE$ number in the condition of FF ($F_{2,34} = 7.90, p < .005$) as well as SR ($F_{2,34} = 6.21, p < .05$). The overall mean $ETE$ numbers were 2.43, 2.11, and 1.92 when the strength coefficient $s$ was equal to .6, .8, and .9, respectively, and 2.01, 2.02, and 1.69 when the speed reduction ratio $r$ was equal to .2, .6, and .8, respectively. Post hoc pair-wise comparison tests indicated that increasing $s$ from .6 to .8 or .9 could significantly decrease the $ETE$ number ($p < .05$ or .005), and increasing $r$ from .2 or .6 to .8 could lead to similar effects ($p < .01$ or .001). However the $s$ values .8 and .9, and the $r$ values .2 and .6 did not generate significant differences on $ETE$. As Figure 5 shows, therefore, we chose .9 for $s$ and .8 for $r$ as the optimal parameters in both methods FF and SR.

Moreover, for the FF method, the parameter factor significantly impacted the dwell time frequency only in the last bin range ($F_{2,34} = 4.22, p < .05$), while for SR method in the second bin range from 100 ms to 199 ms ($F_{2,34} = 4.60, p < .05$). Post hoc tests revealed that the FF method with the $s$ coefficient values of .8 and .9 resulted in significantly higher frequencies in the last range than with the value of .6 ($p < .05$), and the SR method with the $r$ coefficient value of .8 caused significantly lower frequency in the second range than with the value of .6 ($p < .005$).

These results mean that FF with a higher $s$ value was more effective at keeping the eye cursor within the target until the end of the trials. On the other hand, SR with a higher $r$ value was more effective at reducing the probability of ineffectively short dwell times. With the significance of these different aspects, choosing a higher parameter value in both methods could result in the same level of effect in decreasing the $ETE$ number.

### EXPERIMENT 2: EVALUATION WITH POINTING TASKS

In Experiment 1, we determined that both methods FF and SR can improve the stability of eye cursor, effectively counteracting the effects of eye jitter, and that the SR method is generally more effective than the FF method. The attention task used in that experiment was well suited to the verification of our hypotheses, but it could not answer the question of whether or not they can improve human performance when dwell-based activation is used for eye pointing tasks. In this experiment, therefore, we evaluate the human performance of both methods in 2D pointing tasks.

In addition, we noted that the effectiveness of the SR method for the eye cursor's movements was identical in different directions, unlike the FF method, which enhances the movements toward the target center but dampens those toward perimeter. That is to say, the SR method probably can restrict the eye cursor's movement toward the target center. We indeed observed this phenomena during the first experiment. Although damping centrifugal movements is the crucial aspect for solving the stability issue of eye cursor, we believe that this would potentially weaken the effectiveness of the SR method. Thus, we refined the implementation of this method as expressed in the following equation.

$$\vec{n} = \begin{cases} (1-r)\cdot\vec{a} + r\cdot\vec{p} & , if\ \|\vec{c}-\vec{a}\| \geq \|\vec{c}-\vec{p}\| \\ \vec{a} & , otherwise \end{cases} \qquad (3)$$

where, the symbol $\vec{c}$ denotes the target center and the others are the same as those in Equation 2. This improved SR (iSR) method can differentiate between centrifugal and centripetal cursor movements, making it easier for the eye cursor to stay inside the target but also moving it toward the center as fast as the eyes could move. We compared the human performance of this method and the two initial methods with the human performance under the control condition, where there is no process to compensate for eye jitter.

#### Apparatus

The same apparatus was used as in the first experiment.

#### Participants

Sixteen able-bodied subjects (4 female and 12 male with average age of 23 in the range from 20 to 33 years) successfully completed this experiment. Among them, 7 subjects had taken part in experiment 1, the others had never been in an environment with eye tracker. 12 subjects had normal vision without any correction, one used contact lenses and the rest wore glasses. None of them was colorblind.

#### Experimental Task and Procedure

The experimental setup was similar to that in experiment 1. For those who did not participate in the first experiment, a brief introduction about the eye tracker was also given to eliminate their worry about the safety of their eyes. After explaining the experimental task, requirements and process with a corresponding demonstration on the screen, the experimenter helped the subject put the eye tracker on and then calibrated it for the succeeding trials. We did not give subjects a practice block in general as the experimental task which was to select a specified target by intuitive look, does not necessitate a training process [25], but we did offer a practice block when someone requested it.

As Figure 2b shows, at the onset of each trial, a circular trial-start button randomly appeared at one of the predefined positions on the screen. Instead of the actual diameter of 120 pixels, this button was rendered as a 32-pixel-diameter dull yellow solid circle to facilitate its acquisition [16]. After the subject briefly focused on the trial-start button, making the eye cursor stable in the target for 450 ms to "press" it, a target group, arranged as a hexagon, immediately appeared in the diagonal direction and the trial-start button disappeared. The subject was instructed to stare at the desired target as soon as possible. The desired target was located at the center of the hexagon, being marked with a small red solid circle.

When the eye cursor entered a target, the dwell timer was started and a semi-transparent bubble begun to expand from the target center to the target edge as an indicator of dwell time lapsed. When the cursor left the target, the dwell timer and the feedback bubble were reset. If the cursor continuously stayed in the desired target for 1000 ms, a correct selection was recorded. However, if the cursor stayed in any other distracter, a wrong-selection event was recorded. 5 seconds after the trial begun, if no target was selected, a no-selection event was recorded. The current trial was repeated 5 times at most if the desired target had not been successfully selected. After the target group disappeared at the end of each trial, the trial-start button was shown at a different position. During the process, subjects could have a rest if they wanted between blocks. For each trial, the start and end time, the time when the cursor initially entered the target, and the number of $ETE$ were all recorded.

#### Design

A repeated measures within-subject design was applied. The independent factors were cursor redressment method $CRM$ (NM, SR, FF, and iSR), target diameter $D$ (45, 60, and 80 pixels), amplitude $A$ (200, 380, and 850 pixels), and target gap $TG$ (20, 40, and 60 pixels). The experiment was broken up by $CRM$, with 4 blocks for each method. In each block, there were 27 combinations of $D$, $A$, and $TG$ with 4 trials respectively in 4 different diagonal directions for each combination, resulting in a total of 108 trials. The trials were presented in random order and the situation where the trial-start button would appear at the location where the last target group was displayed was avoided. The experiment was divided into two sessions held over consecutive days with each lasting approximately one hour.

## Results

### Eye Movement Time and Eye Pointing Time

Eye movement time ($EMT$) is defined as the time taken to move eyesight from the trial-start button to the desired target, approximately measured from the trial beginning when the target group appeared to the time when the eye cursor entered the desired target the first time. Including $EMT$ as the first stage, eye pointing time ($EPT$) is the duration until the desired target is successfully selected after a continuous dwell time threshold has been reached. We did not find a learning effect between different blocks. Therefore, the same ANOVA as in experiment 1 was used for correct trial records to reveal the effects of different factors as follows.

The main effect of $CRM$ on $EMT$ was statistically significant ($F_{3,45} = 3.06, p = .037$), but very strongly significant for $EPT$ ($F_{3,45} = 12.90, p < .001$). $D$ had a significant main effect on both $EMT$ ($F_{2,30} = 159.57, p < .0001$) and $EPT$ ($F_{2,30} = 263.77, p < .0001$) as well as $A$ on both $EMT$ ($F_{2,30} = 137.19, p < .0001$) and $EPT$ ($F_{2,30} = 60.9, p < .0001$). There were interaction effects $CRM \times D$ ($F_{6,90} = 5.84, p < .001$) on $EPT$, $CRM \times A$ ($F_{6,90} = 2.86, p < .05$) on $EMT$, and $D \times A$ on both $EMT$ ($F_{4,60} = 17.73, p < .001$) and $EPT$ ($F_{4,60} = 7.61, p < .001$). $TG$ had no significant main effect on $EMT$ and neither did $EPT$, but there was a statistical interaction effect $TG \times D$ ($F_{4,60} = 2.83, p = .032$) on $EPT$.

The overall means for $EPT$ were 1766, 1688, 1659, and 1626 ms for the NM, FF, SR, and iSR methods, respectively. Post hoc pair-wise comparison tests revealed that the later three methods significantly decreased $EPT$ ($p < .05, .005$, and .001, respectively). The difference between FF and SR and the difference between SR and iSR was not significant, but the difference between FF and iSR was significant ($p < .005$). That is to say, the improved implementation of SR further decreased $EPT$, making iSR significantly faster for eye pointing tasks than FF. Furthermore, pair-wise comparison tests according to different diameter levels showed that iSR still significantly decreased $EPT$ ($p < .05$) even at the diameter level 80 pixels where neither FF nor SR had significant differences when compared with NM. With respect to $EMT$, there were no significant differences between all pairs of methods except for the pair FF and iSR ($p = .005$). As Figure 6 illustrates, $EPT$ as well as $EMT$ will increase in general if $A$ increases and/or $W$ decreases under each $CRM$ condition. However, neither of them can be modeled using Fitts' law.

### Eye Cursor's Entering Target Event

We also collected $ETE$ for each trial as we did in experiment 1. The main effect of $CRM$ ($F_{3,45} = 36.56, p < .001$), $D$ ($F_{2,30} = 141.61, p < .0001$), and $A$ ($F_{2,30} = 9.59, p < .005$) on $ETE$ was significant. There was an interaction effect $CRM \times D$ ($F_{6,90} = 17.03, p < .001$) on $ETE$. The overall mean $ETE$ numbers were 1.40, 1.21, 1.18, and 1.16 for NM, FF, SR, and iSR method, respectively. Post hoc pair-wise comparison tests showed a similar priority relationship among these four methods based on $ETE$ to that based on $EPT$.

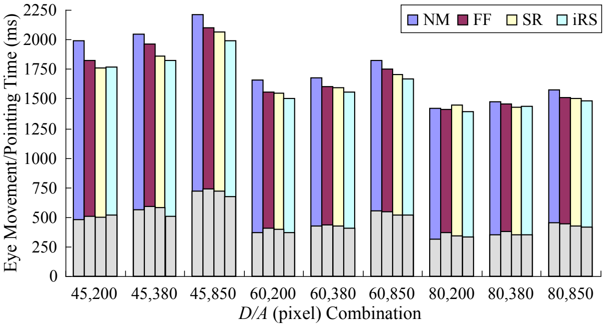

**Figure 6. Mean $EPT$ and $EMT$ by $D/A$ combination for different methods. The bottom section of each column represents $EMT$.**

### Spatial Distribution Features of Eye Cursor

When the eye cursor entered the desired target, its positions were sampled at the frequency of 10 $Hz$, generating a total of more than three hundred thousand points. For each trial, we calculated the average distance ($AD$) from the sampled points to the target center and the standard deviation of corresponding distances ($SDD$) as two measures to reflect the spatial stability of the eye cursor in targets.

The main effect of $CRM$ on both $AD$ ($F_{3,45} = 39.24, p < .001$) and $SDD$ ($F_{3,45} = 35.99, p < .001$), $D$ on both $AD$ ($F_{2,30} = 474.89, p < .0001$) and $SDD$ ($F_{2,30} = 706.1, p < .0001$) as well as $A$ on both $AD$ ($F_{2,30} = 21.00, p < .001$) and $SDD$ ($F_{2,30} = 31.59, p < .001$) was significant. There were also significant interaction effects $CRM \times D$ on both $AD$ ($F_{6,90} = 4.18, p < .005$) and $SDD$ ($F_{6,90} = 3.03, p < .05$), and $D \times A$ on both $AD$ ($F_{4,60} = 4.88, p < .005$) and $SDD$ ($F_{4,60} = 5.04, p < .005$).

The overall mean $AD$ values were 15.17, 13.51, 15.71, and 13.79 pixels and those of $SDD$ were 4.05, 4.46, 3.49, and 3.81 pixels for the methods NM, FF, SR, and iSR, respectively. Post hoc pair-wise comparison tests indicated that there were significant differences among these methods except for the pair FF and iSR for $AD$ and the pair NM and iSR for $SDD$. These results indicated that SR made the eye cursor more aggregative than any of the other methods by reducing its speed (the minimum $SDD$), but it indeed damped the eye cursor's movements toward the target center (the maximum $AD$). The iSR method was effective to overcome this drawback as we expected.

### Error Rate

In the experimental process, if a trial was not successful, the subject had to repeat it at most for 5 times before aborting the trial for a recalibration process to compensate for the deterioration in tracking accuracy. All trials generally had successful records. We categorized the records where the desired targets were not correctly selected at the first attempt as errors, seldom observing wrong-selection events (0.3%).

Therefore, the analysis only aimed at no-selection event data, indicating that there were main effects for $CRM$ ($F_{3,45} = 20.07, p < .001$), $D$ ($F_{2,30} = 152.53, p < .0001$), and $A$ ($F_{2,30} = 9.16, p < .005$) and two corresponding interaction

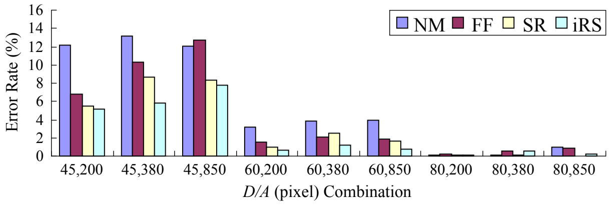

**Figure 7. Error rate by $D/A$ combination for different methods.**

effects $CRM \times D$ ($F_{6,90} = 8.70, p < .001$) and $D \times A$ ($F_{4,60} = 6.34, p < .001$) on error rate. Overall error rates were 5.43%, 3.91%, 2.91%, and 2.30% for NM, FF, SR, and iSR, respectively. As Figure 7 shows, these three methods are significantly effective for decreasing error rates and furthermore iSR is significantly more effective than FF (all at least at the level $p < .05$) when $D = 45$ *or* 60 pixels. Increasing $D$ also significantly decreased error rate ($p < .001$) until all error rates are close to zero when $D = 80$ pixels. When $D = 45$ pixels, decreasing $A$ from 380 or 850 to 200 pixels also significantly decreased the error rate ($p < .005$).

## DISCUSSION

In this study, we defined different dependent variables to measure the stability of eye cursor from different aspects. We have shown the significant effectiveness of the methods FF and SR for improving the stability of eye cursor in experiment 1, and the benefits of both methods as well as the improved SR (iSR) for eye pointing tasks in experiment 2.

Although some researchers stated that slaving a cursor to eye movements was a destructive form of feedback because of its potential to distract the user's gaze [10], we still let the eye cursor remain visible in both experiments. Because another supplementary experiment[1], which duplicated the task in experiment 1 with the additional factor of eye cursor's visibility in the experimental design, revealed that the visibility factor had no a main effect on $ETE$ ($F_{1,13} = .51, p = .487$), the overall effect of $CRM$ on $ETE$ was similar to that in experiment 1. Further analysis of that experiment confirmed that there was a significant difference between TC and NM ($p < .05$) when the eye cursor was visible, but not ($p = .149$) under the invisible condition. We can explain this as follows. According to the principle of TC, the method TC is essentially equivalent to NM when the eye cursor is invisible, while it often makes the cursor suddenly approach or fly away from the target center when the cursor is visible. This phenomenon probably means subjects cannot help chasing the cursor with their eyes as researchers stated, destabilizing the eye cursor. These results indicate that subjects were able to accommodate their eyes to focusing on the target center instead of the eye cursor when they were explicitly required to do so in our experiments unless the eye cursor suddenly and dramatically moved in an unnatural manner such as under the condition of TC. That means that displaying the eye cursor in both experiments did not bias the main results.

The duration that the subject took to stare at the target in experiment 1 as well as the supplementary experiment was 7 seconds, while the counterpart in experiment 2 on the average was less than 1.8 seconds. That means that the extent to which natural blinks or corrective saccades affected the data of these experiments probably varied. However, we observed a consistent priority relation between different methods. Thus, the data confirming the effectiveness of FF and SR for improvements in eye cursor's stability is convincing.

In experiment 1, we found that animated feedback had no main effect on the stability of eye cursor. That is not to say that feedback is not useful for eye-based interaction, but it suggests that providing animated feedback when the subject is staring at the target with a focus center would not significantly distract the subject's attention. No matter what kind of feedback mode (expanding or shrinking), the difference is not significant. The animated feedback can benefit eye-based interaction [13, 14, 15] without extra cost, such as weakening the eye cursor's stability. Therefore, choosing the expanding mode as the dwell time progress cue for eye pointing tasks would not bias the experimental results.

For the methods FF or SR in experiment 1, the three trials of each combination of target diameter and feedback mode were performed under three different parameter conditions, respectively. However, we did not differentiate trials with different parameters but mixed them together when analyzing the effect of $CRM$ on the eye cursor's stability. This was reasonable and acceptable because if we observed a significant difference between FF or SR and NM without considering the impact of parameters, a more significant difference would be observed when choosing an optimal parameter for FF or SR. According to the results, we believed that the priority relations between FF or SR and the other methods were correctly presented. Of course, the priority relation between FF and SR probably could change when two optimal parameters were applied to them. Therefore, both FF and SR were included in experiment 2 for further evaluation, although SR was considered better than FF in experiment 1.

We also did not differentiate trials performed at different screen positions (see Figure 2a). We tried to keep the calibration error at an acceptably low level during the experimental process. Thus, the position effect, i.e. the local calibration error [9], was expected to be insignificant. To ensure that the position effect did not bias the experimental results in any way, we picked out the data of NM, which did not include the impacts of different parameters at different positions as that of FF or SR, and analyzed it again, finding that there was indeed no significant main effect for the position factor on $ETE$ ($F_{8,136} = .88, p = .536$).

As Figure 4 illustrates, the span of bin ranges is not uniform, providing more information in low-value ranges because the dwell time criterion for eye-based interaction varied from 100 to 3000 ms in different studies [7]. For example, Sibert and Jacob [22] applied a very short dwell time of 150 ms in their work. In Lankford's work, the dwell time was customizable to fit different users, meaning an experienced

---

[1] We had another group of 14 subjects formally perform this experiment in two sessions, but we could not completely report it in this paper due to the page limitation.

user can effectively work at a dwell time as short as 500 ms [13]. Majaranta et al. [15] studied the effects of a long dwell time (900 ms) and a short dwell time (450 ms) on eye typing tasks. A short dwell time can maintain the speed advantage of eye movements, but can inevitably cause unintentional activation of commands due to the "Midas-Touch" problem. Stampe and Reingold [25] pointed out that a dwell time of 1000 ms was long enough to prevent the "Midas-Touch" problem without causing difficulty to users. Therefore, we chose 1000 ms for pointing tasks in experiment 2. Under this condition, we observed the effectiveness of the methods FF, SR, and iSR for improving human performance in eye pointing tasks. Although we did not investigate the situation when a shorter dwell time was applied, such as 700 ms, which can work well for simple tasks [25], we can infer that these methods are probably still effective because both FF and SR, especially SR, can significantly reduce the cumulative percentage of the frequencies of dwell time at different bin ranges (see Figure 4). That means at each corresponding dwell time criterion, the probability of resetting the dwell timer under the condition of FF or SR was lower than that under the condition of NM, resulting in faster selections.

With respect to the experimental designs, the method order presented in experiment 1 was different from that in experiment 2, and the order in both experiments was not counterbalanced across subjects in order to reduce learning or fatigue effects because eye movements can make eye pointing tasks near fatigue-free [19] and subjects can perform a wide variety of control tasks without the need for any training [25]. Furthermore, no subject reported that he/she felt any obvious difference between those methods in appearance except for TC in experiment 1. Therefore, we believed this to be an acceptable, although not perfect solution without serious impact on the results.

Hansen et al. [7] pointed out there were three types of dwell-based activation: continuous dwell activation, accumulated dwell activation, and adaptive dwell activation. Omitting the actual positions of the eye gaze, GHA accumulates the time when the eye gaze is inside or outside of the desired target provided that there is no a saccade happening to interrupt this accumulation [16]. By contrast, our methods modulate the gaze positions, making eye cursor stay in the desired target as long as possible. Therefore, our methods belong to the type of continuous dwell activation, while GHA belongs to the type of accumulated dwell activation. According to the empirical evaluations, GHA could decrease error rate, but increase selection time (overall about 10%), but fortunately, our methods decreased them at the same time.

According to the effect of $CRM$ on $EMT$ in experiment 2, there was no obvious delay in eye cursor's movement between different targets when our methods were applied (see Figure 6). Qvarfordt [18] had employed a similar formula to Equation 2 for displaying the indicator of tourists' eye gaze on a map in $iTourist$ but with a slight delay. The method used in $EyeDraw$ also resulted in a slight delay of 133 ms. One reason for the fact that the eye cursor had no delay was probably that our eye tracking device worked at a higher speed (250 $Hz$) than theirs. Another more significant reason was that our methods, without the need to detect saccades, only processed the gaze points related to a specific target. By contrast, other methods processed all the gaze points on the screen in order to smooth them. Jacob pointed out that smoothing gaze data could improve performance during a fixation, but it would probably dampen sudden saccades [9]. The report of Kumar et al. [12], which also depicted the results of a general Kalman filter, verified this point. Those special algorithms with the need to detect saccades based on a predefined time threshold, such as GHA, the one-sided triangular filter and the improved Kalman filter [12], also had to endure a slight time cost when determining whether or not a new fixation had happened. Kalman filter is likely very useful for pursuit eye movements [1], but it seems to be excessive for eye pointing tasks due to the fact that the user's eyes jump ballistically [9] almost without pursuit eye movements in eye pointing tasks. In our view, the key for dwell-based activation is not to filter the noises but to maintain the continuity of dwell time. The work presented in this paper emphasized this point.

## CONCLUSIONS

Dwell-based eye pointing is often frustrated due to the eye cursor's instability that mainly results from eye jitter and the limited accuracy of eye tracking. In this paper, therefore, we introduce several methods, such as force field and speed reduction, to counteract the jittery movements of the eyes so as to improve the stability of eye cursor.

We carried out an attention task experiment to thoroughly evaluate their effectiveness in improving eye cursor's stability. The experimental results show that both methods, force field and speed reduction, especially the later, can effectively enhance the stability of eye cursor. Using a pointing task experiment, we found that both the force field and speed reduction methods as well as the improved implementation of speed reduction could effectively improve the human performance in eye pointing tasks, especially when the target is relatively small. Furthermore the improved method of speed reduction was significantly better than the force field method. At the same time, the results in the first experiment were also confirmed in the second, enhancing their reliability. Overall, our work clarifies that it is feasible to overcome eye jitter and improve human performance by a simple algorithm.

## ACKNOWLEDGMENTS

This work was partially funded by CASIO Science Promotion Foundation, and Academic Frontiers Promotion Program (No.7560040411) in Japan. The authors are grateful to the members of the Ren Lab in Kochi University of Technology and the State Key Lab of Machine Perception in Peking University for their help. We also thank Shumin Zhai and the anonymous CHI reviewers for their valuable comments.

## REFERENCES

1. Abd-Almageed, W., Fadali, M. S., and Bebis, G. A non-intrusive Kalman filter-based tracker for pursuit eye movement. In *Proc. American Control Conference* (2002), 1443-1447.

2. Ahlström, D. Modeling and improving selection in cascading pull-down menus using Fitts' law, the steering law and force fields. In *Proc. CHI 2005*, ACM Press (2005), 61-70.
3. Ashmore, M., Duchowski, A. T., and Shoemaker, G. Efficient Eye Pointing with a FishEye Lens. In *Proc. Graphics Interface*, Canadian Human-Computer Communications Society (2005), 203-210.
4. Bates, R. and Istance, H. O. Zooming interfaces! Enhancing the performance of eye controlled pointing devices. In *Proc. the fifth international ACM SIGCAPH conference on Assistive Technologies (ASSETS)*, ACM Press (2002), 119-126.
5. Bolt, R. A. Eyes at the interface. In *Proc. CHI'82*, ACM Press (1982), 360-362.
6. Duchowski, A. T. A Breadth-First Survey of Eye Tracking Applications. In *Behaviour Research Methods, Instruments, and Computers (BRMIC)*, 34(4) (2002), 455-470.
7. Hansen, J. P., Johansen, A. S., Hansen, D. W., Itoh, K., and Mashino, S. Command without a click: dwell time typing by mouse and gaze selections. In *Proc. INTERACT'03*, IOS Press (2003), 121-128.
8. Hornof, A. J. and Cavender, A. EyeDraw: Enabling children with severe motor impairments to draw with their eyes. In *Proc. CHI 2005*, ACM Press (2005), 161-170.
9. Jacob, R. The use of eye movements in human-computer interaction techniques: what you look at is what you get. In *ACM Transactions on Information Systems*, Vol.9, No.3, ACM Press (1991), 152-169.
10. Jacob, R. J. K. Eye Movement-Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces. In *Advances in Human-Computer Interaction*, Vol.4, Ablex Publishing (1993), 151-190.
11. Kumar, M., Paepcke, A., and Winograd, T. EyePoint: Practical pointing and selection using gaze and keyboard. In *Proc. CHI 2007*, ACM Press (2007), 421-430.
12. Kumar, M., Klingner, J., Puranik, R., Winograd, T., and Paepcke, A. Improving the Accuracy of Gaze Input. Technique Report CSTR 2007-03, Stanford University HCI Group, March 2007.
13. Lankford, C. Effective eye-gaze input into windows. In *Proc. ETRA 2000*, ACM Press (2000), 23-27.
14. Majaranta, P. and Räihä, K.-J. Twenty Years of Eye Typing: Systems and Design Issues. In *Proc. ETRA 2002*, ACM Press (2002), 15-22.
15. Majaranta, P., MacKenzie, I. S., Aula, A., and Räihä, K.-J. Effects of feedback and dwell time on eye typing speed and accuracy. In *Journal of Universal Access in the Information Society*, Vol.5 No. 2 Springer-Verlag (2006), 199-208.
16. Miniotas, D., Špakov, O., and MacKenzie, I. S. Eye Gaze Interaction with Expanding Targets. In *Ext. Abstracts CHI 2004*, ACM Press (2004), 1255-1258.
17. Miniotas, D., Špakov, O. Tugoy, I., and MacKenzie, I. S. Speech-augmented eye gaze interaction with small closely spaced targets. In *Proc. ETRA 2006*, ACM Press (2006), 67-72.
18. Qvarfordt, P. Eyes on multimodal interaction. *Linköping Studies in Science and Technology*, Dissertation No.893, Ph. D. Thesis, Linköping University, Sweden (2004).
19. Saito, S. Does fatigue exist in a quantitative measurement of eye movements? *Ergonomics* 35(5-6) (1992), 607-615.
20. Salvucci, D. D. Inferring intent in eyemovement interfaces: Tracing user actions with process models. In *Proc. CHI'99*, ACM Press (1999), 254-261.
21. Salvucci, D. D. and Anderson, J. R. Intelligent gaze-added interfaces. In *Proc. CHI 2000*, ACM Press (2000), 273-280.
22. Sibert L. and Jacob, J. Evaluation of Eye Gaze Interaction. In *Proc. CHI 2000*, ACM Press (2000), 281-288.
23. Špakov, O. and Miniotas, D. Dynamic target expansion to facilitate eye-based pointing at menus. *Information Technology And Control*, Vol.34, No.2 (2005), 135-139.
24. SR Research Ltd. EyeLink II system. http://www.eyelinkinfo.com/index.php.
25. Stampe, D. M. and Reingold, E. M. Selection by looking: A novel computer interface and its application to psychological research. In *Eye Movement Research: Mechanisms, Processes, and Applications*, Elsevier Science Publishing (1995), 467-478.
26. Surakka, V., Illi, M., and Isokoski, P. Gazing and frowning as a new human-computer interaction technique. In *ACM Transactions on Applied Perceptions*, Vol.1, No.1 (2004), 40-56.
27. Ward, D. J. and MacKay, D. J. C. Fast hands-free writing by gaze direction. *Nature*, Vol.418, No.6900 (2002), 838.
28. Ware, C. and Mikaelian, H. H. An evaluation of an eye tracker as a device for computer input. In *Proc. CHI'87*, ACM Press (1987), 183-188.
29. Zhai, S., Morimoto, C., and Ihde, S. Manual and gaze input cascaded (MAGIC) pointing. In *Proc. CHI'99*, ACM Press (1999), 246-253.
30. Zhang, Q., Imamiya, A., Go, K., and Mao, X. Resolving ambiguities of a gaze and speech interface. In *Proc. ETRA 2004*, ACM Press (2004), 85-92.