Seamless Interaction with Scrolling Contents on Eyewear Computers Using Optokinetic Nystagmus Eye Movements

Shahram Jalaliniya*
Diako Mardanbegi†

Abstract

In this paper we investigate the utility of an eye-based interaction technique (EyeGrip) for seamless interaction with scrolling contents on eyewear computers. EyeGrip uses Optokinetic Nystagmus (OKN) eye movements to detect object of interest among a set of scrolling contents and automatically stops scrolling for the user. We empirically evaluated the usability of EyeGrip in two different applications for eyewear computers: 1) a menu scroll viewer and 2) a Facebook newsfeed reader. The results of our study showed that the EyeGrip technique performs as good as keyboard which has long been a well-known input device. Moreover, the accuracy of the EyeGrip method for menu item selection was higher while in the Facebook study participants found keyboard more accurate.

Keywords: Eye tracking, Optokinetic Nystagmus eye movements (OKN), eyewear computers, scrolling, implicit input

Concepts: •**Human-centered computing → Interaction techniques;**

Figure 1: *A right-to-left fast scrolling menu on a head-mounted display where it stops on the content that has attracted the user's visual attention*

1 Introduction

Scrolling for navigation on small-screen devices (e.g. smartphones) has its own usability and inefficiency problems [Harms et al. 2015] which can be even more challenging on eyewear computers such as Google Glass. The only mechanism for scrolling the main menu in Google Glass UI, is to perform touch gestures on the touch-sensitive surface on the right side of the device. However, this manual mechanism is not always the best modality where the users' hands are busy with other tasks. Moreover, on a small screen, finding the desired content that has gone out of the screen requires a lot of touch gestures which is not always easy and sets a limitation to how fast the scrolling can be done. An alternative approach is to use eye-based techniques for hands-free interaction with scrolling contents in eyewear computers. But despite the great potentials of using our eyes for interaction, eye-based interaction techniques are still not widely used. Several challenges need to be tackled to make gaze interaction more pervasive. First, existing eye trackers need to be calibrated for each user due the differences between individual eye geometries. Furthermore, other factors such as relative movements of the eye and the eye tracker, ambient light conditions, and calibration quality affect the accuracy of gaze tracking. Finally, eye is a perceptual organ and is not suitable to use as an explicit input [Jacob 1990].

*e-mail: jsha@itu.dk
†e-mail:dima@itu.dk

ETRA 2016, March 14 - 17, 2016, Charleston, SC, USA
ISBN: 978-1-4503-4125-7/16/03
DOI: http://dx.doi.org/10.1145/2857491.2857539

Variety of gaze interaction techniques have been proposed in the recent years to overcome the above-mentioned challenges of eye-based interaction. Overall, we see trends towards more implicit way of using eye input in non-command interfaces [Nielsen 1993; Mardanbegi et al.]. The *EyeGrip* method is a novel implicit eye-based method for interaction with scrolling contents that addresses all of the above-mentioned challenges [Jalaliniya and Mardanbegi]. EyeGrip is a calibration-free method that uses natural reflexive Optokinetic Nystagmus eye movements for hands-free interaction with dynamic user interfaces and helps the user intuitively stop a sequence of moving (scrolling) visual contents displayed on the computer screen. In this paper, we investigate the utility of the EyeGrip method for eyewear computers that are becoming increasingly popular, to allow the user to seamlessly select an item among a series a scrolling contents in the near-eye display (Figure 1).

2 Related Work

The idea of using eye input in a scrolling task was originally suggested by Jacob et al. [Jacob 1990] and further studied (e.g. by [Kumar and Winograd]) but these work were limited to only enhancing the task of reading digital documents by automatically scrolling the page based on where the user is looking at. [Vidal et al. 2013] proposed the Pursuits method for interaction with dynamic user interfaces. Pursuits allows users to select an object among several moving objects on a display by following the object with their eyes which leads to a smooth pursuit eye movement. Since in the Pursuits method the trajectory of the moving objects should not to be identical, it is not possible to use Pursuits for interaction with scrolling contents with an identical trajectory. On the contrary, the EyeGrip method [Jalaliniya and Mardanbegi] enables us to detect an object of interest among a set of moving objects that all move in the same direction at the same speed. The earlier study on EyeGrip [Jalaliniya and Mardanbegi] was done on a desktop computer and a machine learning approach was used to detect the object of interest while in this paper, we implemented the EyeGrip method on a mobile setup with a head-mounted display (HMD) and a different classifier that analyses users' eye-strokes.

3 EyeGrip Method

Optokinetic nystagmus (OKN) has a sawtooth-like pattern that consists of alternating pursuits movements made in the direction of stimulus (slow phase) followed by saccacdes (fast phases). When a user is doing a visual search by looking at a scrolling sequence of contents on a computer screen we can see that the eye follows some objects longer than the others. The longer the user is following an object of interest, the higher the peak of the slow phase appears in the eye movement signal. In Figure 2 the blue signal (E) represents the horizontal eye movements (over time) recorded by an eye tracker in a visual search task where a user looks at horizontally (right to left) scrolling images and searches for a particular image (target) among other pictures. The upper/lower bounds of the signal (marked on the vertical axis of the graph) correspond to the left and right sides of the screen. The slow and fast phases of the Look OKN are visible in the figure as well as the longer slow phases that has happened when the target images have drawn user's attention two times (marked by green circles). These longer slow phases in the signal are denoted by *target-peak* in this paper. Implementing the EyeGrip method requires a tight communication and synchronization between the eye tracker and the UI. The eye tracker tracks the user's eye movements and generates a set of feature vectors (e.g. pupil center or even gaze coordinates). For the sake of synchronization later these vectors could be time-stamped (e.g. $\left[t_i, E_i\right]$ where t_i is time and E_i is the eye feature vector). The scrolling engine that updates the UI, controls what part of the moving sequence is within the display at each time instant (e.g. $\left[t_i, S_i\right]$ where S_i is the position an the state of the scroller). These two sets of vectors are pushed to a classifier that detects the target-peaks in the OKN signal. Once the classification is done and a target-peak is detected, the UI needs to be updated accordingly e.g. by stop scrolling and bringing the area or content of interest back to the screen.

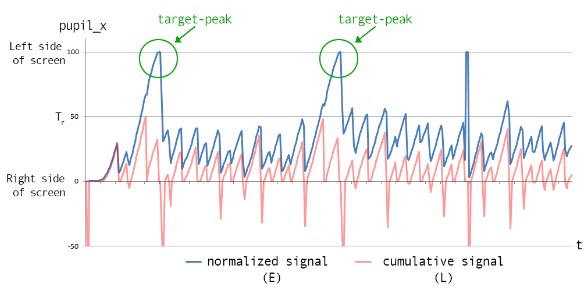

Figure 2: *The blue color signal (E) is generated from horizontal eye movements in a visual search task among images scrolling from right to the left side of the screen. The red color signal (L) is generated in the classification process. The target-peak is detected when L exceeds the T_r limit.*

3.1 Our approach for detecting target-peaks

In this study, we used pupil center as the main feature obtained from the eye image and we were only interested in the horizontal OKN movements. The moving stimuli on the display were scrolling from right to left, and only the horizontal component of the pupil center is captured. We adopted a sliding window approach for detecting the target-peaks and once a peak in the signal is detected, the system reacts by stop scrolling the content. Although, the EyeGrip technique itself does not rely on the absolute position of the gaze point and it does not require any gaze estimation, our classification algorithm needs to be calibrated (not gaze calibration) because it relies on an absolute threshold operator in the decision process that needs to be adopted for each user. In order to calculate a default threshold value for different users, we have normalized the eye signal for all users and in our implementation, we asked each participant to look at two red circles to detect right and left borders of the screen. This is used to determine the lower/upper bounds of the OKN signal and to obtain the parameters of a linear mapping function (F_{norm}). This function is used to map the pupil coordinates to a normalized range of $[LB, UB]$ where LB and UB respectively correspond to the pupil position when a user looks at the right and left sides of the screen. For each incoming frame, we find the pupil center in the image denoted as E_{new}. Then we clean the data by removing zero values caused by pupil tracking failures. After applying a smoothing window and normalizing the new data ($F_{norm}(E_{new})$), the pre-processed data E_N is passed to the classifier. The classifier always keeps a window of the recent N observations $E_{1:N} = \left\{E_t | 1 \leqslant t \leqslant N\right\}$ (a set of observations of E_t in a sliding window of N frames within time span of $\left[t - N + 1, t\right]$). After a new observation the classifier updates the buffer. The classifier also buffers two other sets of features both generated from the main observation input of E_N. The first set is defined as $\Delta E_{1:N} = \left\{\Delta E_t = E_{t+1} - E_t | 2 \leqslant t \leqslant N\right\}$ which is basically a difference between adjacent items in the sliding window $E_{1:N}$. The other set is a cumulative sum of the items in the data set $\Delta E_{1:N}$ and is defined as $L_{1:N} = \left\{L_t | 2 \leqslant t \leqslant N\right\}$ where L_t is defined based on the following rules where $direction = \Delta E_t \times \Delta E_{t-1}$.

$$L_t = \begin{cases} \Delta E_t + L_{t-1} & \text{if } direction > 0 \\ 0 & \text{if } direction \leq 0 \end{cases} \tag{1}$$

Finally using a general threshold T_r the current frame E_N is classified into target-peak class (TP) for when $L_N > T_r$ and $E_N < 0.5 \times UB$ otherwise to no-interest-area class (NIA). Where the term $E_N < 0.5 \times UB$ is to ensure that when a target-peak is detected, the user has been looking at the left side of the screen. Whenever a target-peak is detected the two $\Delta E_{1:N}$ and $L_{1:N}$ get empty to ensure that the event is not detected twice. Based on a preliminary study, we have derived the optimized constant values used in the classification process as: $\left\{T_r = 50, LB = 0, UB = 100, N = 100\right\}$ The main reason for using the value L_N instead of E_N for the classification is to ensure that a fast movement from right to left side of the screen or noise in the data is not considered as a slow phase (see the last peak of the signal in Figure 2). Figure 2 (red line) shows the result of applying the classification process on an example signal.

3.2 Voluntary vs involuntary interactions

As humans, we can attend to objects one by one. Our visual attention can be attracted by salient stimuli that 'pop out' in our surroundings which is called bottom-up attentional mechanism. Attention can also be voluntarily directed to an object based on our longer-term cognitive strategies which is more like a top-down mechanism [Connor et al. 2004]. In EyeGrip, we detect users' object of interest in a scrolling UI where users' attention can be directed to a certain object either voluntarily due to a predefined task (top-down mechanism) or involuntarily to operate on raw sensory input such as an attractive colour or fast movements in the user interface (bottom-up mechanism). We exploit these two attentional mechanisms to implement EyeGrip interaction technique for two different types of applications. In the first application, users have a predefined plan to search for a particular menu item on the screen. In this case, EyeGrip supports the top-down attention mechanism by stopping the menu scroller when a menu item draws users' attention. We call this type of using EyeGrip, *voluntary interaction* since the user is voluntarily searching for a specific object. In the second application, there is no predefined goal in the visual search task. Users might look at the computer screen with no specific goal or task. In such cases they might be attracted to an image due to the novelty of the image or transients such as motion, and change [Pashler and Harris 2001]. In this type of *involuntary interactions*, the bottom-up attention mechanism directs users' attention.

4 User Study

To characterise involuntary and voluntary types of interactions, we developed two different applications where users can select a particular item among scrolling visual contents using EyeGrip: 1) a menu scroll-viewer and 2) a Facebook newsfeed reader. In order to maximize the performance of EyeGrip the velocity and image width should be adjusted carefully [Jalaliniya and Mardanbegi]. Based on the earlier study on EyeGrip and our preliminary study we selected the optimized values for the size of the content ($W_{content}$), size of the constant offset between the contents (W_{offset}), and scrolling speed ($Speed$) as: $W_{content} = 0.6 \times W_{display}$, $W_{offset} = 0.1 \times W_{display}$, and $Speed = 35° /second$

4.1 Participants

We recruited 11 participants (mean age = 35, from 28 to 52 years old, and 3 females) among local university staff to try both systems. Three participants wear glasses and one uses contact lenses. Rest of the participants have perfect visual acuity. Also 9 of 11 participants are Facebook users. Each participant completed the tasks for both studies in a single session in approximately 30 minutes.

4.2 Apparatus

Due to the limitation in processing power of Google Glass, we simulated the Google Glass info cards in a desktop application with an HMD. The screen of the desktop application is mirrored on a binocular HMD (ICUITI DV920) with the resolution of 640×480. For tracking eye movements, we used a home-made wearable monocular gaze tracker and the open-source Haytham gaze tracking software[1]. The eye tracking camera is mounted on the HMD. The frequency of the sampling eye data is 20Hz. In the Facebook reader app the newsfeed is scrolled horizontally unlike the original Facebook newsfeed in mobile devices which scrolls vertically. Due to the dimensions of the display in state of the art eyewear computers it is easier to explore Facebook posts horizontally.

4.3 Study 1: gaze enabled menu scroll-viewer

The main concept in the Google Glass UI is a side-scrolling stream of info cards providing updates for different categories of interests, such as email, messages, weather, etc. The main challenge of using info cards is that if you want to find a particular card, you have to scroll through every existing card which requires a lot of touch gestures on the touchpad. EyeGrip can be a solution to this problem. The gaze-enabled menu scroll-viewer system helps users stop scrolling info cards whenever target card passes in front of the users' eyes. We investigate the utility of EyeGrip for such menu selection tasks. To have a baseline for comparison, we implemented a manual method (keyboard) for menu selection and the participants were asked to select info cards with both manual and EyeGrip methods.

4.4 Study 2: Facebook newsfeed reader

When people are browsing their Facebook page, particularly in smart phones, it's often the case that they quickly scan their newsfeed by scrolling down or up until they find something interesting to stop on. If Google Glass users want to explore their Facebook newsfeed, they would need to keep swiping back and forth on the touchpad of the Glass. Just like the menu scroll-viewer app, EyeGrip can provide a hands-free automatic mechanism to stop scrolling when an interesting Facebook post draws users' attention. To investigate the utility of EyeGrip in such applications, we developed a Facebook newsfeed reader app for eyewear computers. The participants, tried the Facebook app immediately after the menu scroll-viewer. To have a baseline for comparison, the participants were asked to use both manual (keyboard) and EyeGrip methods.

4.5 Procedure

The session started with a short introduction on purpose of the study. Since none of the participants were familiar with the concept of info cards in Google Glass, first we asked them to wear a Google Glass and scroll between different cards using the touchpad of the Glass. Next, we asked them to wear the HMD and eye tracker set.

Menu scroll viewer: In a menu selection task usually users are familiar with the menu items; therefore, we showed our sample 14 menu items (14 is an arbitrary big enough sample size) to the user before starting the task. Then the participant started the manual selection mode where the menu items (cards) start scrolling on the screen and the user is asked to press the space key to stop scrolling when the target item appears on the screen. The participants could correct their error using left and right arrow keys if they stop before or after the target item by mistake. The task is repeated five times for five different target items. In the next condition, the participants performed the same task using EyeGrip. In this step, the EyeGrip method automatically stopped the scrolling content as soon as the participant found the target item. Also in this condition, the task was repeated five times for five different target items, and we recorded the accuracy of the automatic menu selection and the number of corrections. In order to remove the order effect, the manual and EyeGrip conditions are counterbalanced and the cards were randomly positioned in the queue.

Facebook reader: this part started immediately after the first part and participant went through a similar procedure (manual and EyeGrip conditions). The only difference was that instead of menu items, 50 Facebook posts (randomly selected from public Facebook pages such as CNN, National Geographic, etc.) was scrolling on the screen, and participants were not familiar with the posts. In contrast to the study 1, there was no plan for stop scrolling. The users were allowed to stop scrolling whenever they found something interesting among the scrolling content. After finishing each part, the users were asked to complete a questionnaire with 5-point likert scale questions polling their experience completing the task. After filling out the questionnaire, the participants were interviewed briefly.

4.6 Results

Study1: We defined the error as the total number of items between the selected item and the target item. This means if the target card is selected correctly the error is zero. The statistical paired t-test with Holm-Bonferroni corrections indicates a significant difference in the accuracy of menu selection for manual (mean = 2, $\sigma = 1.6$) and EyeGrip (mean = .90, $\sigma = .94$) conditions; $t(10) = 2.12$, $p = .02 < \alpha = .05$. The results of the usability questionnaire is illustrated in Figure 3-b. Pairwise comparisons showed that, participants found EyeGrip significantly more intuitive than the manual for selecting menu items; $t(10) = 2.6$, $p = .01 < \alpha = .05$. Moreover, the users evaluated EyeGrip as a more comfortable method for selecting items compared to the manual method; $t(10) = 5.16$, $p = .0002 < \alpha = .05$. Finally, EyeGrip is also recognized as a faster

[1] http://eyeinfo.itu.dk

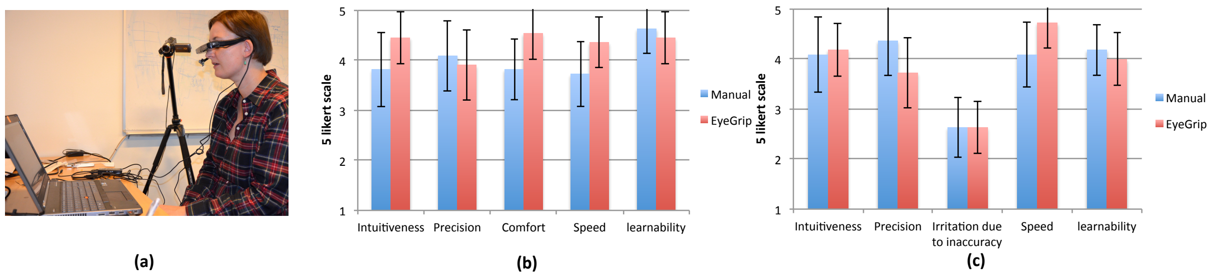

Figure 3: *a) A participant performing the task, b) Result of the questionnaire for study1, and c) Result of the questionnaire for study2*

method for selecting menu items compared to the manual method; t(10) = 2.05, p = .03 $< \alpha = .05$.

Study2: The results of the questionnaire is represented in Figure 3-c. We compared different aspects of the usability for each condition using statistical paired t-tests with Holm-Bonferroni corrections. The statistical analysis indicated that participants experienced the manual method significantly more precise than EyeGrip for the Facebook app, t(10) = 1.88, p = .04 $< \alpha = .05$. However, EyeGrip is evaluated significantly faster compared to manual method for reading the Facebook newsfeed, t(10) = 2.6, p = .01 $< \alpha = .05$. Other usability aspects indicated no significant difference.

In both studies, we asked participants how much temporal tension they felt during the task. They felt significantly more temporal tension in the Facebook study (mean = 3.09 $\sigma = .69$) compared to the scroll-viewer study; t(10) = 2.69, p = .01 $< \alpha = .05$.

5 Discussion & Conclusions

In the menu selection study, EyeGrip is more accurate than the manual method. It can be due the high speed of the scrolling menu items in the screen which requires a fast reaction to select the target item as soon as it appears on the screen. EyeGrip can be potentially a faster than the manual technique since in EyeGrip, as soon as the eyes react to an item on the screen the system stops scrolling without any additional motor task. But the manual method requires a very high coordination between eyes, brain and our motor control system which decreases the user performance. Due to the same reason, in the first study, participants found EyeGrip faster than the manual method. The participants also found EyeGrip more comfortable and intuitive method for selecting menu items compared to the manual method. The comfort and intuitiveness of the EyeGrip can be explained by being a hands-free interaction technique which is easier to use in eyewear devices. In the second study, the EyeGrip is again evaluated as a faster method for exploring Facebook compared to the manual method. However, the manual method is recognized significantly more accurate than EyeGrip. The reason can be the significant role of the bottom-up attention mechanism in the Facebook reader app which directs users' attention based on properties of the visual contents. Since this type of attention is involuntary, sometimes even the user does not know if s/he is paying attention to a particular object in the scene. This might be the reason why the manual approach is evaluated significantly more precise method in the Facebook app and not in the menu selection app. This reveals the limitations of using EyeGrip in involuntary applications. In both studies, we compared the usability of EyeGrip with keyboard which is a very old and well-known input device. Nonetheless, the EyeGrip technique is evaluated event better in some usability aspects such as speed and intuitiveness. The result of interviews also showed that 90% of the participants preferred the EyeGrip method for both applications. One of the points made by most of the participants was the need for a hands-free modality to start the scrolling movement. In a real application, the manual method that we used to start scrolling can be replaced with another hands-free technique such as dwell-time or even head gestures that are detected by eyewear computers.

Apart from functional utilities of EyeGrip, many participants particularly expressed that the EyeGrip interaction technique is *different* and *fun*: *"It can be relaxing if you just lay down at home, wear a HMD, and let your unconscious attention together with EyeGrip decide what you should see in Facebook newsfeed." (Participant 4)*

In the future, we shall investigate the use of EyeGrip with vertical scrolling stimuli as vertical OKN has different characteristics. The way of handling deviations in OKN movements, once they are detected by the system, could also be the subject of future studies.

References

CONNOR, C. E., EGETH, H. E., AND YANTIS, S. 2004. Visual attention: bottom-up versus top-down. *Current Biology 14*, 19, R850–R852.

HARMS, J., KRATKY, M., WIMMER, C., KAPPEL, K., AND GRECHENIG, T. 2015. Navigation in long forms on smartphones: Scrolling worse than tabs, menus, and collapsible fieldsets. In *INTERACT 2015*. Springer, 333–340.

JACOB, R. J. K. 1990. What you look at is what you get: Eye movement-based interaction techniques. In *Proc. of CHI '90*, ACM, New York, NY, USA, 11–18.

JALALINIYA, S., AND MARDANBEGI, D. Eyegrip: Detecting targets in a series of uni-directional moving objects using optokinetic nystagmus eye movements. In *Proc. of CHI '16*.

KUMAR, M., AND WINOGRAD, T. Gaze-enhanced scrolling techniques. In *Proc. of UIST '07*.

MARDANBEGI, D., HANSEN, D. W., AND PEDERSON, T. Eye-based head gestures. In *Proc. of ETRA '12*.

NIELSEN, J. 1993. Noncommand user interfaces. *Commun. ACM 36*, 4 (Apr.), 83–99.

PASHLER, H., AND HARRIS, C. R. 2001. Spontaneous allocation of visual attention: Dominant role of uniqueness. *Psychonomic Bulletin & Review 8*, 4, 747–752.

VIDAL, M., BULLING, A., AND GELLERSEN, H. 2013. Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proc. of UbiComp '13*, ACM, New York, NY, USA, 439–448.