SPOCK: A Smooth Pursuit Oculomotor Control Kit

Simon Schenk
Technische Universität
München
Munich, Germany
simon.schenk@tum.de

Philipp Tiefenbacher
Technische Universität
München
Munich, Germany
philipp.tiefenbacher@tum.de

Gerhard Rigoll
Technische Universität
München
Munich, Germany
rigoll@tum.de

Michael Dorr
Technische Universität
München
Munich, Germany
michael.dorr@tum.de

CHI'16 Extended Abstracts, May 07–12, 2016, San Jose, CA, USA
ACM 978-1-4503-4082-3/16/05.
http://dx.doi.org/10.1145/2851581.2892291

Abstract

Gaze holds great potential for fast and intuitive hands-free user interaction. However, existing methods typically suffer from the Midas touch problem, i.e. the difficult distinction between gaze for perception and for user action; proposed solutions have required custom-tailored, application-specific user interfaces. Here, we present SPOCK, a novel gaze interaction method based on smooth pursuit eye movements requiring only minimal extensions to button-based interfaces. Upon looking at a UI element, two overlaid dynamic stimuli appear and tracking one of them triggers activation. In contrast to fixations and saccades, smooth pursuits are not only easily performed, but also easily suppressed, thus greatly reducing the Midas touch problem. We evaluated SPOCK against dwell time, the state-of-the-art gaze interaction method, in a simple target selection and a more challenging multiple-choice scenario. At higher task difficulty, unintentional target activations were reduced almost 15-fold by SPOCK, making this a promising method for gaze interaction.

Author Keywords

Gaze-based interaction; smooth pursuit; gaze gestures; gaze tracking

ACM Classification Keywords

H.5.2. [User interfaces]: Input devices and strategies.

Introduction

Gaze-based interfaces have the potential to provide an intuitive and fast input modality, for example when hands-free operation is desired. For severely motor-impaired users, gaze-based interfaces may even be the only means for computer-human interaction. While various solutions for gaze typing already exist [1, 17], and gaze pointing has been shown to outperform the mouse in some specific applications [3, 11, 15], a more general approach that could fully replace traditional input devices such as the mouse is still missing. Such an approach would open up the whole range of existing applications and their user interfaces to gaze-based interaction.

One fundamental problem is posed by the need to perform two different actions, namely to select a target (point) and to activate it (click), using gaze only. Because the eye primarily serves as a sensor and not as an actuator, volitional oculomotor control is limited and seemingly trivial activation methods such as blinking (and worse, not blinking) quickly become cumbersome. Moreover, eye movements are constantly used to sample the visual input and are needed by the user to identify both targets and non-targets in the first place, so that a major challenge for gaze-based interfaces is to distinguish between gaze for sensing and gaze for acting: Referring to the mythical king who turned everything he touched into gold, this is called the Midas touch problem [7]. Here, we propose a novel gaze interaction approach based on smooth pursuit eye movements that is largely immune to this problem.

Related Work

The eyes typically alternate several times per second between relatively stationary phases (fixations) and rapid movements (saccades). Only in the presence of slowly-moving targets, tracking eye movements (smooth pursuits) may also occur. All these gaze characteristics have been used to detect activation interactions.

For example, in dwell time approaches, fixating the target for a certain duration activates it [7]. However, the optimal duration threshold is difficult to determine: Short dwell times ($<$300 ms) are hardly distinguishable from natural fixations and long dwell times ($>$1000 ms) are very tiring and hard to perform; optimal dwell time also depends on the information to be processed before activation [14]. Furthermore, the optimal dwell time changes through strong learning effects and may need to be adjusted constantly [10].

Saccade-based selection methods were first introduced as single-, two-, or multi-stroke gestures. Here, the eye follows an imaginary pattern performing a sequence of saccades [8, 12], and distinct patterns correspond to distinct actions. These patterns combine selection and activation into a single gesture, which makes them incompatible with existing mouse-based user interfaces; they also require expert knowledge, i.e. memorization of all possible gestures. Another way to utilize saccades while retaining this separation of selection and activation is the use of antisaccades: A stimulus appears on one side of the selection target after a short fixation duration, but the target is activated by making a saccade to the opposite side, thus alleviating attentional capture [6]. These methods do not suffer from the Midas touch problem because they are based on unnatural gaze behaviour, resulting in an unergonomic user experience.

Smooth pursuit (SP) may be considered as a fixation on a moving object and thus requires a moving stimulus. This simplifies the Midas touch prevention, since typical user interfaces are mostly static. Even though SP gestures tend to be hard to detect, they have been utilized in a number of special purpose applications. Most of these approaches are

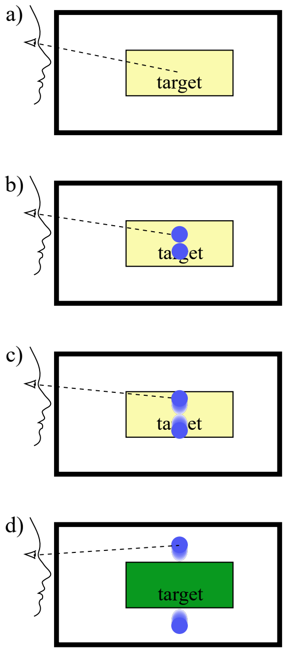


Figure 1: Schematic illustration of SPOCK: Looking at the target (a) selects it and two overlaid discs appear that move up- and downwards, respectively (b/c). Following one of these discs with gaze activates the target (c/d).

based on dynamic user interfaces [2, 4, 9, 16], where each possible selection target moves along a unique path on the screen. In order to select one of these targets, it has to be followed with the eyes, and it is activated once the trajectories of gaze and target have been similar enough for a certain duration. This requires customized human-computer interfaces specific to gaze interaction; they also still run the risk of visual inspection periods being misclassified as activations.

Vidal et al. [16] also outlined how their approach might be extended for a static desktop environment, but did not empirically evaluate this idea. In an extensive user study, Esteves et al. [4] showed how smooth pursuit based interaction can successfully be used on smartwatches. However, their interface is also not general purpose and cannot be easily transferred to a desktop environment.

The SPOCK Method

We here propose the Smooth Pursuit Oculomotor Control Kit (SPOCK), a method that is not based on a single gaze characteristic, but that combines different eye movement types for selection and activation and thus avoids the Midas touch problem. This method may be used with only marginal changes to existing user interface layouts and is illustrated schematically in Figure 1: Upon looking at the selection target, two small discs appear in the centre of the selection target and begin to slowly move up- and downwards, respectively. For target activation, the user then has to follow one of these discs with a smooth pursuit eye movement. If the user has not activated the target within a certain time window (i.e. once the discs have reached a certain eccentricity), the discs start again at the target centre. This cycle is repeated until the target is activated or the user looks away. The use of only one disc that suddenly appears would introduce involuntary eye movements due to attentional capture [18]. In our symmetric design, however, the attentional capture of one side cancels out the one of the other side and unintentional eye movements are minimized.

Because of the separation of selection and activation, the user is free to visually inspect potential targets without the Midas touch problem, which should lead to fewer unintentional activations. This is particularly important when the activation is only secondary to a primary decision-making process.

Experiment

We conducted a case study with 18 participants (13 male, 5 female; 23–34 years) to compare SPOCK with dwell time - the state of the art activation method - in regards of completion time and fail attempts. Two different scenarios were chosen to achieve a broad coverage of the design space: a simple selection task favouring dwell time and allowing for a fast completion time, and a more complex multiple choice task facilitating Midas touches.

Design and Procedure

The study followed a within-subject design with *activation method* (dwell time, SPOCK) as main factor. Each experiment comprised two blocks, one per activation method. In each block the selection scenario had to be performed first, followed by the question scenario. Half of the participants started with dwell time as activation method, the other half with SPOCK. The questions of the two blocks were different, but the overall order of questions was always the same. Before the selection scenario, each block contained a training session with nine selection targets.

Apparatus

Binocular gaze was recorded by an EyeLink 1000 Plus eye tracker in the tower configuration running at 1000 Hz. No

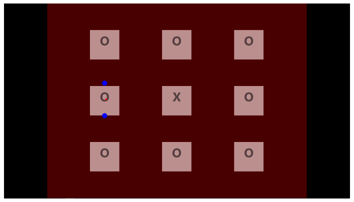
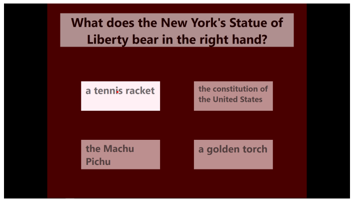


Figure 3: Screenshot showing the selection scenario with SPOCK.


What does the New York's Statue of Liberty bear in the right hand?

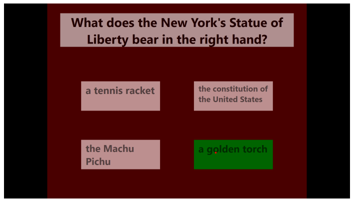
a tennis racket

the constitution of the United States

the Machu Pichu

a golden torch

Figure 4: Screenshot showing a fail attempt of dwell time in the multiple choice scenario.


What does the New York's Statue of Liberty bear in the right hand?

a tennis racket

the constitution of the United States

the Machu Pichu

a golden torch

Figure 5: Screenshot showing a correct selection of dwell time in the multiple choice scenario.

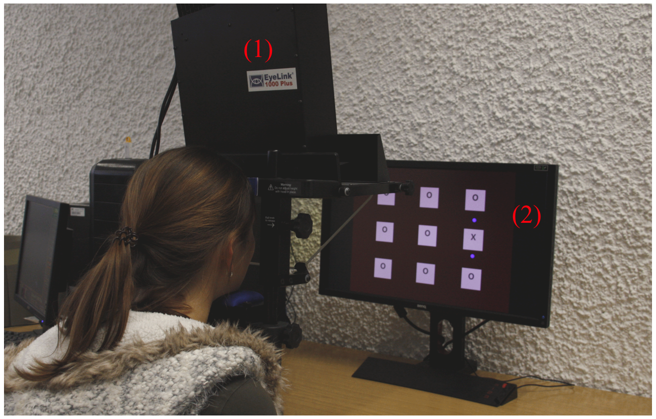


Figure 2: Experimental setup: Tower-mounted gaze tracker (1), and the monitor showing the selection scenario (2).

built-in smoothing or saccade or blink detection was used. The targets were shown on a 23-inch monitor with a resolution of 1920x1080 pixels at a viewing distance of 68 cm. The experiment was implemented in C# using the WPF framework and ran at 60 fps.

Selection targets had a size of 163 pixels (3°) and were spaced 237 pixels (4.4°) apart. The different answer choices of the multiple choice task had a size of 438x163 (width x height) pixels (8°x3°) and were spaced 163 pixels (3°) apart. Figure 2 shows the experimental setup with the selection scenario running.

Gestures
We detected SP gestures using Support Vector Machines (SVM) with Gaussian kernels. In a sliding window of 300 ms the velocity of the vertical gaze component was computed with finite differences, followed by low-pass filtering. These values were used as feature vectors for the SVMs. Windows that did not contain saccades were classified using a one-versus-one multi-SVM approach with three different SVMs – one for each gesture (SP up, SP down, and fixation). SVMs were trained with data from a prestudy under laboratory conditions (80,000 training vectors from hand-labelled data) and achieved a precision and recall of about 85%.

Since we wanted to use the same dwell time for both scenarios, we chose the longest yet still user-friendly dwell time suggested in [5], i.e. 500 ms. For SPOCK, we used blue discs with a diameter of 28 pixels (0.5°) as pursuit targets. They moved with a constant speed of 108 $\frac{\text{pixels}}{sec}$ ($2\frac{\text{deg}}{sec}$) for a maximum eccentricity of 163 pixels (3°).

Scenarios
Since neither position nor shape influence the selection performance (see [13]), we chose a 3x3 layout of square targets for the selection scenario (Figure 3). The subject had to activate the target marked with an *X*, whereas all other targets were marked with *O*s. These targets were placed randomly, but in the same order for all participants, and each target position had to be activated twice, leading to a total of 18 target activations.

In the multiple-choice question scenario, one question at a time was shown at the top of the screen. Four different answer choices were given below, arranged in a 2x2 grid (Figures 4 and 5). All questions and answers were taken from the German *'Who wants to be a millionaire?'* board game's 50€ – 300€ questions. If necessary, answers were slightly modified to have similar syllable count[1]. The correct answers were placed randomly, but at the same position for each participant. Each position held the correct answer

[1] Note that the screenshots in Figure 4/5 show a translated version.

three times, for a total of 12 questions. At the end of the session, the participants were asked to indicate for each question whether they had known the right answer, or simply guessed it.

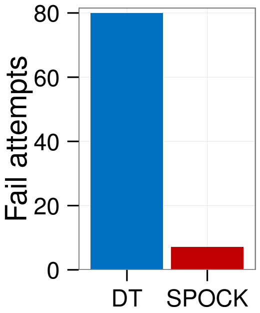


Figure 6: The total number of fail attempts for each activation method in the multiple choice scenario.

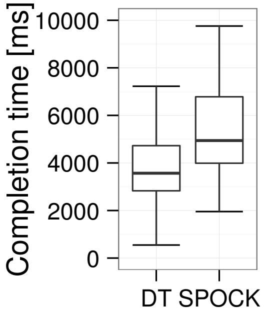


Figure 7: The overall completion time for the multiple choice scenario. Three outliers not shown.

Results

Some of the selection targets were dropped during the selection scenario due to recording difficulties, and questions for which the subject indicated they had only guessed the answer were discarded from the analysis as well. Overall, we analysed 323 selection and 193 question targets (out of 324 and 216 targets, respectively) for SPOCK; for dwell time, these numbers were 293 selection and 189 question targets.

Neither fail attempts nor completion time met the normality assumption of ANOVA. Instead, we used the fail attempt rate per subject per target (PSPT) for testing, i.e. the mean of all fail attempts for one subject. Similarly, we computed mean completion time for each subject over all targets per scenario. Both fail attempt rates and mean completion times were not normally distributed, so Wilcoxon signed-rank tests were used for significance testing.

The fail attempt rate was very low in the selection scenario with 0.01 fail attempts PSPT for SPOCK and 0.02 fail attempts PSPT for dwell time (no significant difference, $p \approx 0.58$). In the question scenario, the fail attempt rate increased for dwell time, while SPOCK still produced very low fail attempt rates (0.45 and 0.03 fail attempts PSPT, respectively). The Wilcoxon test confirmed this observation with the fail attempt rate being significantly smaller for SPOCK ($p << 0.01$). Figure 6 shows the fail attempts for the different activation methods in the multiple choice scenario.

The mean completion time was lower for dwell time in both scenarios. The Wilcoxon test confirmed significance in the selection and the question scenarios (both $p << 0.01$). Figure 7 shows the overall completion time for both scenarios.

Discussion and Conclusion

More and more low-cost eye trackers are currently becoming available, but gaze-based interfaces have not become commonplace yet; besides technical issues such as calibration quality and robustness, one of the major tripping stones is the Midas touch problem, i.e. the difficulty of distinguishing between unintentional and intentional gaze gestures. The current state of the art uses dwell time, i.e. long fixations, to detect intentional activations, but only with mixed success because of the sensitivity of fixation duration to factors such as mental workload. Recently, methods based on smooth pursuit eye movements have been proposed, but these require custom-tailored dynamic interfaces and thus cannot be used for general purposes.

In this paper, we proposed a new method for gaze-based interaction using smooth pursuit gestures that could be added to any general desktop UI by simply introducing a superimposed layer of dynamic stimuli. In a case study with a simple selection scenario and a more complex multiple-choice scenario, we compared our method to the dwell time method.

Results confirmed that dwell time can be a very fast and accurate activation method for simple tasks that require only short pointing actions. However, when an additional workload was introduced that required subjects to process complex visual information at the target location and to answer a question, the fail attempt rate increased dramatically by a factor of 20. The SPOCK method, however, yielded low fail attempt rates even under these difficult circumstances. With a fail attempt rate of only 0.03 (compared to 0.45 for dwell time) in the multiple-choice scenario, we were able to eliminate the Midas touch problem almost completely.

In the current state, SPOCK is relatively slow compared to dwell time. Even in the more complex multiple choice scenario, dwell time was faster than SPOCK, despite an error rate of 0.45 fail attempts per trial. However, introducing a moderate penalty for each misclick, e.g. having to re-take the trial (an average penalty of 1.6 s), would have equalized performance. Also, the current detection algorithm based on SVMs leaves room for improvement. Considering that the fastest completion time was achieved by SPOCK (335 ms versus 512 ms), improving detection rate could greatly improve the completion time.

Overall, we presented a gaze-based activation method that can easily be applied to existing mouse-based UIs and that successfully solved the Midas touch problem. Thus, it can be used in scenarios where the cost of misclicks is high.

In future work the detection rate, i.e. precision and recall, of the smooth pursuit gestures has to be improved in order to reduce completion time and make the system more user-friendly. Also, the size and movement speed of the visual stimulus are important factors for the usability of our system and have to be optimized.

Acknowledgments

This research was supported by the Elite Network Bavaria, funded by the Bavarian State Ministry for Research and Education.

References

[1] Tuhin Chakraborty, Sayan Sarcar, and Debasis Samanta. 2014. Design and Evaluation of a Dwell-free Eye Typing Technique. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. ACM, 1573–1578.

[2] Dietlind Helene Cymek, Antje Christine Venjakob, Stefan Ruff, Otto Hans-Martin Lutz, Simon Hofmann, and Matthias Roetting. 2014. Entering PIN codes by smooth pursuit eye movements. *Journal of Eye Movement Research* 7, 4 (2014), 1–11.

[3] Michael Dorr, Martin Böhme, Thomas Martinetz, and Erhardt Barth. 2007. Gaze beats mouse: a case study. In *The 3rd Conference on Communication by Gaze Interaction (COGAIN '07)*. 16–19.

[4] Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology (UIST '15)*. ACM, 457–466.

[5] Jens R. Helmert, Sebastian Pannasch, and Boris M. Velichkovsky. 2008. Influences of dwell time and cursor control on the performance in gaze driven typing. *Journal of Eye Movement Research* 2, 4 (2008), 1–8.

[6] Anke Huckauf and Mario H. Urbina. 2011. Object selection in gaze controlled systems: What you don't look at is what you get. *ACM Transactions on Applied Perception* 8, 2 (2011), 13:1–13:14.

[7] Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-computer Interaction Techniques: What You Look at is What You Get. *ACM Transactions on Information Systems* 9, 2 (1991), 152–169.

[8] Jari Kangas, Deepak Akkil, Jussi Rantala, Poika Isokoski, Päivi Majaranta, and Roope Raisamo. 2014. Gaze Gestures and Haptic Feedback in Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, 435–438.

[9] Otto Hans-Martin Lutz, Antje Christine Venjakob, and Stefan Ruff. 2015. SMOOVS: Towards calibration-free text entry by gaze using smooth pursuit movements. *Journal of Eye Movement Research* 8, 2 (2015), 1–11.

[10] Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. 2009. Fast Gaze Typing with an Adjustable Dwell Time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, 357–360.

[11] Julio C. Mateo, Javier San Agustin, and John Paulin Hansen. 2008. Gaze Beats Mouse: Hands-free Selection by Combining Gaze and EMG. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08)*. ACM, 3039–3044.

[12] Emilie Mollenbach, John Paulin Hansen, Martin Lillholm, and Alastair Gale. 2009. Single Stroke Gaze Gestures. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems (CHI EA '09)*. ACM, 4555–4560.

[13] Atsuo Murata, Raku Uetsugi, and Daichi Fukunaga. 2014. Effects of target shape and display location on pointing performance by eye-gaze input system. In *Proceedings of the SICE Annual Conference*. 955–962.

[14] Abdul Moiz Penkar, Christof Lutteroth, and Gerald Weber. 2012. Designing for the Eye: Design Parameters for Dwell in Gaze Interaction. In *Proceedings of the 24th Australian Computer-Human Interaction Conference (OzCHI '12)*. ACM, 479–488.

[15] Linda E. Sibert and Robert J. K. Jacob. 2000. Evaluation of Eye Gaze Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*. ACM, 281–288.

[16] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing (UbiComp '13)*. ACM, 439–448.

[17] David J. Ward and David J. C. MacKay. 2002. Fast Hands-free Writing by Gaze Direction. *Nature* 418, 6900 (2002), 838–838.

[18] Steven Yantis. 1996. Attentional capture in vision. *Converging operations in the study of visual selective attention* (1996), 45–76.