On the Verge: Voluntary Convergences for Accurate and Precise Timing of Gaze Input

Dominik Kirst
Max Planck Institute for Informatics
Saarbrücken, Germany
kirst@mpi-inf.mpg.de

Andreas Bulling
Max Planck Institute for Informatics
Saarbrücken, Germany
bulling@mpi-inf.mpg.de

Abstract

The problem of triggering input accurately (with a small temporal offset) and precisely (with high repeatability) at a specific point in time has so far been largely ignored in gaze interaction research. We explore voluntary eye convergences as a novel interaction technique for precise and accurate timing of gaze input and a solution to the "Midas touch" problem, i.e. the accidental triggering of input when looking at an interface. We introduce a novel clock paradigm to study input timing and demonstrate that voluntary convergences are significantly more accurate and precise than common gaze dwelling. Our findings suggest that voluntary convergences are well-suited for applications in which timing of user input is important, thereby complementing existing gaze techniques that focus on speed and spatial precision.

Author Keywords

Vergence Eye Movements; Input Timing; Voluntary Convergences; Dwell Time; Clock Paradigm

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces – Evaluation/Methodology

CHI'16 Extended Abstracts, May 07-12, 2016, San Jose, CA, USA
ACM 978-1-4503-4082-3/16/05
http://dx.doi.org/10.1145/2851581.2892307

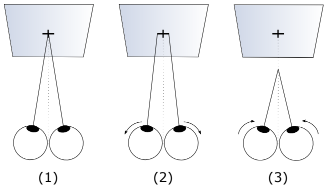

Figure 2: Common gaze techniques involve focussing on the display plane (1). When voluntarily diverging (2) or converging (3) the focus is behind or in front of the display, respectively.

Introduction

Gaze is a compelling modality for human-computer interaction given that the eyes can be positioned with both high speed and spatial precision [8, 11]. A number of different gaze interaction techniques have been proposed, such as eye gestures and voluntary eye blinks [5, 7], gaze dwelling [18], smooth pursuit movements [16], or left-right movements [20]. However, existing interaction techniques mainly exploited the high speed and/or spatial precision of gaze input, such as for pointing [19] or object selection [13]. In contrast, input timing, i.e. the problem of triggering input accurately and precisely at a specific point in time, has been largely ignored by the research community. This is despite the fact that input timing is important for many interactive applications. Similar to the general definition of accuracy and precision, temporal accuracy refers to the temporal offset of gaze input to the target position while temporal precision refers to the repeatability, or reproducibility of the input, calculated as the standard deviation of the temporal offset (see Figure 1 for an overview of the different time measures used in gaze interaction research).

In this work we explore voluntary vergence eye movements as a novel gaze interaction technique for accurate and precise input timing. Involuntary vergences are regularly performed in daily life but can also be performed voluntarily, for example, to view autostereograms or for crossing of the eyes. Although the neurology and control of vergences are well understood [14] they have so far only been used to estimate gaze depth [6, 15] or to detect the attention location of near-eye displays [17]. Kudo et al. investigated divergence movements as a gaze input technique [9]. We show that divergence movements are less favourable and instead propose convergences, which are movements of both eyes in inward direction (see Figure 2).

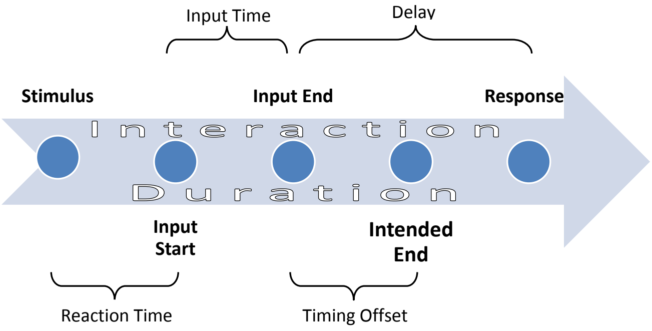

Figure 1: Time measures used in gaze interaction research. Exposed to a stimulus, a user performs some input, usually a specific movement. This might not trigger a response at the time it was intended by the user. The difference between interaction end and aimed end is the timing offset. It may overlap with the delay, which denotes the time until the response of the interface. Note that the order of the last three events may vary.

The specific contributions of this work are two-fold. We first introduce timing of user input as a complement to established performance measures of gaze interaction techniques, such as input speed or spatial precision. We introduce a new experimental setup that uses a clock paradigm to measure accuracy and precision of input timing in a principled way. Second, we propose voluntary convergence eye movements as a novel interaction technique and demonstrate that convergences are significantly more accurate and precise for timing input than well-established gaze dwelling.

Pilot Study

We first conducted a pilot study to better understand the potential of voluntary vergence movements for gaze interaction. The pilot study had two objectives: (1) to obtain initial insights into if and how voluntary vergence movements could be used and (2) to collect data on which we could develop methods to robustly detect convergence and divergence.

Setup and Procedure

We recorded 40 voluntarily triggered vergence movements (20 convergences, 20 divergences) of seven members of our research group and asked for feedback. The stimulus was a central cross hair and participants were instructed to arbitrarily begin a vergence. Whenever the eyeball pose was changed enough with respect to a certain threshold (see below), the time was measured until the eyes returned to the starting position. Participants were seated in front of a 29.7” display (resolution 2560×1600 px) at a distance of about 60 cm. The display was connected to a desktop computer running the experiment software. The software was implemented in Python using the PyGaze [4] and PsychoPy [12] libraries. Gaze data was recorded at a sampling rate of 300 Hz using a Tobii TX300 remote eye tracker placed under the display.

Usability of Vergence Movements

All participants were able to perform convergence movements after a few minutes of training. During training we told them the trick to focus on a point on their nose tip. The average duration for performing a convergence from begin to end was 287 ms ($SD = 251$ ms). In contrast, given that here are no comparable aids to perform a divergence, participants felt not confident while diverging and required a mean duration of 1695 ms ($SD = 5583$ ms). Consequently, the majority of participants reported that divergences were difficult to perform and not suited for a desktop scenario. In particular the necessary concentration to perform the divergence was reported to slow down the technique. We therefore decided to focus on convergence movements.

Detection of Convergence Movements

We investigated two methods for detecting convergence eye movements. Input to both methods is binocular gaze data, i.e. data recorded simultaneously from both eyes, as readily provided by state-of-the-art stationary eye trackers. The first method uses the 3D eye positions to compute the angle between both gaze vectors. An increase of this value indicates a convergence while a decrease indicates a divergence, respectively. The advantage of this method is that it is precise and robust since no transformation of the gaze points into the display coordinate system is necessary. However, this method requires high-quality gaze estimates and the angle between the eyes depends on the distance from head to display. This means that head movements might be misinterpreted as vergence movements. The second method avoids these problems by using the offset of the x-coordinates of on-screen gaze positions in pixels. This value is close to zero whenever the user focuses on the screen and is independent from the head-display distance. We therefore used the first algorithm for convergence detection, an angular threshold of 3°, and a temporal threshold of 150 ms. In uncontrolled settings in the wild, the second algorithm will most likely be more robust.

Main Study

Based on the pre-study findings, we designed a controlled laboratory study to evaluate the performance and usability of voluntary convergence movements. To be able to study learning effects, the study was conducted in three recording sessions over the course of several days. Specifically, we wanted to investigate the following hypotheses:

H1: Convergence is temporally more accurate than dwelling.

H2: Convergence is temporally more precise than dwelling.

Measuring Timing Accuracy and Precision

We used a clock paradigm to measure temporal accuracy and precision of voluntary convergence with the most established gaze technique, namely gaze dwelling (with a dwell-time set to 300 ms). Our paradigm is inspired by the Libet clock [10] as used by Coyle et al. to study agency [3]. It consists of a clock-like circle with a clock hand (indicated by a black line) rotating around the centre (see Figure 3). The task is to trigger input and thereby stop the hand as close as possible to a target position indicated by a triangular marker at the clock's outer rim. The hand stopped whenever the gaze cursor reached the inner (blue) belt of the clock or when a convergence movement was performed. We measured the temporal offset to the target time when the hand reached the marker. While in this work we focus on gaze-based interaction, we believe the proposed paradigm has potential also for other input modalities and interaction techniques for which input timing is important.

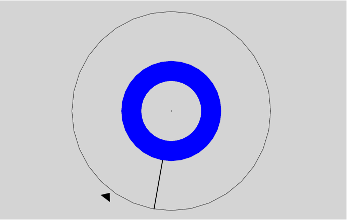

Figure 3: Clock paradigm to study input timing performance. The clock hand rotates and has to be stopped by participants as close as possible to the target position indicated by the triangular marker.

Participants

We recruited 14 participants aged between 22 and 27 years with mixed experience in eye tracking studies (six had no experience, six had one or two, and two had four or more participations). They were paid 20 EUR for compensation.

Procedure

We chose a constant angular velocity of 36°/s for the hand. Thus one full rotation took 10 s, a time we considered reasonable since it was short enough to not demotivate when the target was missed and long enough to allow to detect offsets up to 5 s magnitude. Since we measured the temporal offset and not the angle between hand and marker, we believe that the angular velocity does not significantly influence the results as long as it does not leave this reasonable range. The radii were 150 px / 3.5° and 250 px / 6° for the inner belt and 500 px / 12° for the full clock. Hence the interactive part filled the foveal area of the visual field while the moving hand was visible in the periphery.

The experiment was conducted in separate recording sessions on three different days, each taking approximately 15 minutes. The recording system was the same as in the pre-study. In the first recording, participants were introduced to the eye tracking system and the interaction techniques. After a first calibration, they were allowed to practise vergence interaction in a training scenario. The training interface simply changed the screen colour whenever a vergence was successfully performed. After training, the eye tracker was recalibrated and gaze estimation accuracy validated. Participants then performed the timing task with 30 trials for dwell-time and convergence interaction in randomised order. The position of the triangular marker and the starting angle of the clock hand were also randomised in each trial. The second and third recording session followed the same procedure where 40 and 50 trials of the timing task were performed, respectively. After the first two recordings, we asked participants for feedback on both techniques using a standard System Usability Scale (SUS) questionnaire on a five-point Likert scale ranging from "strongly disagree" to "strongly agree" [2]. In addition, participants were asked for informal feedback.

Results

Results for both methods and all three recording sessions of the timing task are shown in Figure 5. We measured an average temporal offset of 748 ms for dwell-time interaction and of 515 ms for convergence interaction. The respective standard deviations were 997 ms and 760 ms, respectively. We conducted a two-way repeated measures ANOVA with the recording mean of the timing offset as the dependent variable. As shown in Figure 5, we found a significant main effect of method on the temporal offset

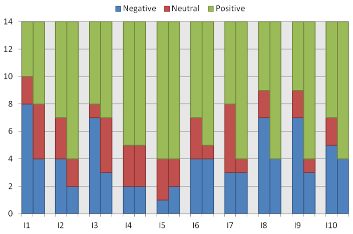

Figure 4: The results of the SUS-questionnaires for items I1 through I10 regarding convergence interaction. The two respective columns depict first and second recording. The length of the coloured intervals indicates how many participants answered equally. An answer is positive whenever the participant disagreed or strongly disagreed with a negative item or agreed or strongly agreed with a positive item (opposite for negative answers).

($F_{1,13} = 6.39, p < .05, \eta^2 = .33$). As mentioned above, the offset for convergence interaction ($M = 515, SD = 760$) is smaller than the offset of dwell-time interaction ($M = 748, SD = 997$). Thus hypothesis H1 can be accepted. Also the effect of the recording became significant ($F_{2,26} = 14.34, p < .05, \eta^2 = .52$). More specifically, only the comparison of recording 2 to 3 did not show a significant effect on the performance ($p > .05$). The interaction effect of both method and recording on the performance was not significant ($F_{2,26} = 1.06, p > .05, \eta^2 = .08$).

A second two-way repeated measures ANOVA with the recording standard deviation of the timing offset as the dependent variable unveiled evidence for hypothesis H2. Both the method ($F_{1,13} = 6.4, p < .05, \eta^2 = .33$) and the recording ($F_{2,26} = 9.49, p < .05, \eta^2 = .42$) had a significant effect on the variable. The average standard deviation of convergence interaction was 760 ms and 997 ms for dwell-time interaction.

Questionnaires
Note that scores in the SUS [2] range from 0 (worst usability) to 100 (best usability). In our data collection, the average SUS score of convergence interaction was 53,93 ($SD = 23,87$) in the first and 65,36 ($SD = 23,33$) in the second recording, while dwell-time interaction was scored with an average of 86,17 points ($SD = 12,74$). According to [1], the results for convergence interaction correspond to marginal acceptability, which ranges from a score of 50 to 70. Following the standardised vocabulary, the usability of convergence interaction was “OK” while usability of dwell-time interaction was “excellent”.

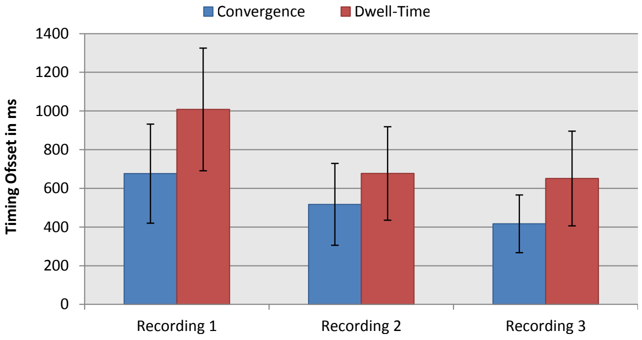

Figure 5: Average timing offset and standard error for each recording session and method in the timing task.

Figure 4 summarizes the SUS scores for convergence interaction. For simplicity, the five-point Likert scale was reduced to three classes by merging the two agree and disagree scores. Notably, the number of participants (strongly) disagreeing with positive items (odd indexes) decreased while also fewer participants (strongly) agreed with negative items (even numbers) in the second recording. In the first recording only 5 out of 14 participants reported that they felt confident with 10 in the second recording. The opposite group reduced from 7 to 3. The informal feedback regarding convergence interaction ranged from “completely awkward” to “surprisingly easy”. Most of the participants stated they had problems triggering the input first but identified a clear turning point when they had learned the technique. None of the participants reported any headache, eye pain or other symptoms of fatigue.

Discussion

Our findings suggest that voluntary convergences are a temporally accurate and precise interaction technique. The technique opens up new perspectives on the design and implementation of a new class of gaze-based interfaces that rely on convergences either as the sole interaction technique or as a means to complement existing techniques. A key advantage over most widely used gaze dwelling is that convergence movements allow the user to precisely decide when input should be triggered. This makes convergences particularly appealing for real-time control, gaming or musical interfaces. Beyond such special-purpose interfaces, convergences also encourage to break out of conventional thinking with respect to which other eye movements future gaze-based interfaces might leverage and how they could better exploit gaze data recorded using binocular eye trackers.

In addition, convergences have the inherent advantage of robustness against false positives and thereby address the so-called “Midas touch” problem, i.e. the problem of distinguishing intentional gaze input from involuntary fixations performed regularly to find and look at content on the interface [8]. This is because unlike dwelling, confirming selections via convergences is a discriminable act with which random visual skimming of the interface and selecting can be clearly distinguished. The clock paradigm proved well-suited to measure the temporal accuracy and precision of gaze input. Given that the paradigm was applicable to two different interaction techniques, we believe that it could be used also for other techniques and modalities.

While voluntary convergences outperformed gaze dwelling, divergence movements appear to be less usable. This was unexpected given previous work on divergences [9]. This might be, at least in part, due to the difficulty of teaching participants how to perform such movements. Focusing on the nose instead of imagining to fixate on some point behind the display plane is easier and results in convergence interaction to be more natural and faster to learn. For other interfaces, e.g. see-through head-mounted or stereoscopic displays, we believe divergence movements have the potential to perform comparably or even better.

While our preliminary study suggests that voluntary convergences have potential, they face similar usability problems as existing voluntary gaze interaction techniques, such as gaze gestures or blinks. For one, while convergences are regularly performed in daily life, their voluntary control may sound uncomfortable, hard, or even impossible to perform. Second, given that convergence interaction involves shifting the focus of visual attention away from the display plane, display content cannot be perceived any more. It will be interesting to investigate how to address these usability challenges, for example by using on-screen visual cues and/or feedback.

Conclusion

In this work we introduced voluntary convergence eye movements as a novel interaction technique for gaze-based interfaces. As the convergences are voluntary, the technique is inherently insusceptible to the Midas Touch problem. We further proposed timing of user input as a novel performance measure and a clock paradigm to study input timing in a principled way. Results from our user study suggest that voluntary convergences can provide a robust means for temporally accurate and precise input timing, thereby outperforming established selection methods based on gaze dwelling. Voluntary convergences thereby complement existing gaze techniques that focus on speed and spatial precision.

Acknowledgements

This work was funded, in part, by the Cluster of Excellence on Multimodal Computing and Interaction (MMCI) at Saarland University, Germany.

References

[1] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An Empirical Evaluation of the System Usability

Scale. *International Journal of Human-Computer Interaction* 24, 6 (2008), 574–594.

[2] John Brooke. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[3] David Coyle, James Moore, Per Ola Kristensson, Paul Fletcher, and Alan Blackwell. 2012. I Did That! Measuring Users' Experience of Agency in Their Own Actions. In *Proc. CHI*. 2025–2034.

[4] Edwin S. Dalmaijer, Sebastiaan Mathôt, and Stefan Van der Stigchel. 2014. PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods* 46, 4 (2014), 913–921.

[5] Heiko Drewes and Albrecht Schmidt. 2007. Interacting with the computer using gaze gestures. In *Proc. INTERACT*. 475–488.

[6] Andrew T. Duchowski, Donald H. House, Jordan Gestring, Robert Congdon, Lech Świrski, Neil A. Dodgson, Krzysztof Krejtz, and Izabela Krejtz. 2014. Comparing Estimated Gaze Depth in Virtual and Physical Environments. In *Proc. ETRA*. 103–110.

[7] Henna Heikkilä and Kari-Jouko Räihä. 2012. Simple Gaze Gestures and the Closure of the Eyes As an Interaction Technique. In *Proc. ETRA*. 147–154.

[8] Robert JK Jacob. 1990. What you look at is what you get: eye movement-based interaction techniques. In *Proc. CHI*. 11–18.

[9] Shinya Kudo, Hiroyuki Okabe, Taku Hachisu, Michi Sato, Shogo Fukushima, and Hiroyuki Kajimoto. 2013. Input Method Using Divergence Eye Movement. In *Ext. Abstr. CHI*. 1335–1340.

[10] Benjamin Libet, Curtis A Gleason, Elwood W Wright, and Dennis K Pearl. 1983. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). *Brain* 106, 3 (1983), 623–642.

[11] Päivi Majaranta and Andreas Bulling. 2014. *Eye Tracking and Eye-Based Human-Computer Interaction*. Springer, 39–65.

[12] Jonathan W. Peirce. 2007. PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods* 162, 1-2 (2007), 8–13.

[13] Sophie Stellmach and Raimund Dachselt. 2013. Still Looking: Investigating Seamless Gaze-supported Selection, Positioning, and Manipulation of Distant Targets. In *Proc. CHI*. 285–294.

[14] FM Toates. 1974. Vergence eye movements. *Documenta Ophthalmologica* 37, 1 (1974), 153–214.

[15] Takumi Toyama, Jason Orlosky, Daniel Sonntag, and Kiyoshi Kiyokawa. 2014. A Natural Interface for Multi-focal Plane Head Mounted Displays Using 3D Gaze. In *Proc. AVI*. 25–32.

[16] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays based on Smooth Pursuit Eye Movement and Moving Targets. In *Proc. UbiComp*. 439–448.

[17] Mélodie Vidal, David H Nguyen, and Kent Lyons. 2014. Looking at or through?: using eye tracking to infer attention location for wearable transparent displays. In *Proc. ISWC*. 87–90.

[18] Colin Ware and Harutune H. Mikaelian. 1987. An Evaluation of an Eye Tracker As a Device for Computer Input. In *Proc. CHI*. 183–188.

[19] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and gaze input cascaded (MAGIC) pointing. In *Proc. CHI*. 246–253.

[20] Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2013. SideWays: A Gaze Interface for Spontaneous Interaction with Situated Displays. In *Proc. CHI*. 851–860.