Gaze Behaviour on Interacted Objects during Hand Interaction in Virtual Reality for Eye Tracking Calibration

Ludwig Sidenmark
Lancaster University
Lancaster, United Kingdom
l.sidenmark@lancaster.ac.uk

Anders Lundström
KTH, Royal Institute of Technology
Stockholm, Sweden
andelund@kth.se

ABSTRACT

In this paper, we investigate the probability and timing of attaining gaze fixations on interacted objects during hand interaction in virtual reality, with the main purpose for implicit and continuous eye tracking re-calibration. We conducted an evaluation with 15 participants in which their gaze was recorded while interacting with virtual objects. The data was analysed to find factors influencing the probability of fixations at different phases of interaction for different object types. The results indicate that 1) interacting with stationary objects may be favourable in attaining fixations to moving objects, 2) prolonged and precision-demanding interactions positively influences the probability to attain fixations, 3) performing multiple interactions simultaneously can negatively impact the probability of fixations, and 4) feedback can initiate and end fixations on objects.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **Virtual reality**.

KEYWORDS

Virtual Reality, Eye Tracking, Hand-eye Coordination, Calibration, Empirical Study

ACM Reference Format:

Ludwig Sidenmark and Anders Lundström. 2019. Gaze Behaviour on Interacted Objects during Hand Interaction in Virtual Reality for Eye Tracking Calibration. In *Proceedings of 2019 Symposium on Eye Tracking Research and Applications (ETRA '19).* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Virtual reality (VR) is a continuously growing area of research, and new technology is developed rapidly. Motion controllers that emulate real-world hand interaction in VR have become more established as the primary interaction modality. Additionally, eye tracking technology has recently been introduced for multiple VR platforms. The availability of both hand and eye tracking allows researchers to study how users coordinate their hands and eyes during interaction in VR to inform interaction behaviour and design and may also be leveraged for novel interaction techniques.

Hand-eye coordination during interaction has been extensively studied in neuroscience in both controlled [Ballard et al. 1992, 1995; Helsen et al. 1998; Johansson et al. 2001; Pelz et al. 2001; Smeets et al. 1996] and real-world contexts [Land et al. 1999; Land and Tatler 2012]. The combined research has shown that humans regularly use their eyes to guide and coordinate their hands when interacting. Additionally, humans generally fixate at interacted objects. These findings have been foundational for practical applications such as gaze prediction [Huang et al. 2012], gaze-enhanced input [Zhai et al. 1999] and multimodal interaction techniques that integrate eyes and hands in desktop [Pfeuffer et al. 2014, 2015] or VR settings [Pfeuffer et al. 2017; Zeleznik et al. 2005]. However, gaze patterns during hand interaction in VR have not yet been extensively investigated. As such, significant differences in gaze patterns during hand interaction between real-world and VR contexts due to the limitations of VR technology may remain uncovered. Further knowledge could inform VR design and open up the design space for novel VR interaction techniques.

A possible application area for hand interaction and related gaze patterns in VR is to leverage the expected gaze patterns on interacted objects for eye tracking re-calibration. Calibration is achieved by using the interacted object as a substitute for a conventional predefined calibration point. Traditional eye tracking calibration has shown to be tedious and time-consuming for users [Flatla et al. 2011]. Therefore, alternative eye tracking calibration methods has become an extensive research field. Techniques that make the process unobtrusive or gamified [Flatla et al. 2011; Gomez and Gellersen 2017; Renner et al. 2011], using visual saliency models [Judd et al. 2009; Sugano and Bulling 2015], smooth pursuit eye movements [Gomez and Gellersen 2017; Pfeuffer et al. 2013; Tripathi and Guenter 2016] or by selecting optimal calibration points have been introduced [Gomez and Gellersen 2018; Lander et al. 2016]. However, most commercial eye trackers still use the traditional calibration method due to its reliability.

Gaze patterns on interacted objects during hand interaction is an interesting opportunity with several potential benefits for VR eye tracking calibration. First, it is ubiquitous and can be hidden within the environment, and is, therefore, less interruptive compared to the traditional calibration. Second, it can be used continuously, thus mitigating the risk of calibration deterioration. We envision that the method would be suitable for a seamless and implicit re-calibration that continuously updates the current calibration during a VR session. However, questions such as what interaction type and at what point during an interaction would be most suitable for

ETRA '19, June 25–28, 2019, Denver , CO, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

re-calibration needs to be investigated. Additionally, the general gaze behaviour during hand interaction in VR has to be established.

In this paper, we evaluate the prevalence of eye fixations on interacted objects during hand interaction in VR as a means for seamless and implicit eye tracking re-calibration. In the evaluation, participants completed a set of interactions in VR using three different interactive objects; a cube, a knob and a slider. Gaze data was recorded during the preformed interactions and analysed for fixations on the interacted objects.

2 BACKGROUND

2.1 Hand-eye Coordination

Hand-eye coordination has long been researched in neuroscience to understand the relationship between the eyes and hands. Eye movement generally precedes hand movement [Land et al. 1999; Land and Tatler 2012] and is extensively used to guide the hands by fixating on the object the user wants to interact with [Johansson et al. 2001]. Additionally, humans use their eyes to incrementally acquire the information required to perform the current task just before action [Ballard et al. 1992, 1995]. If the eyes are not available for target selection or guidance, then the hands will wait until they are available [Pelz et al. 2001]. The gaze is then directed towards locations that appear critical for the task, and the kinematics of the tasks determine when users shift their gaze between landmarks [Johansson et al. 2001; Land et al. 1999].

As such, several types of fixations used for different purposes has been discovered. Land et al. defined four types of fixations that are frequently used when performing a task [Land et al. 1999]:

- **Locate:** fixation on an object that will be used later in the process.
- **Direct:** fixation on a location or a part of an object which is about to be approached and contacted by the hand, or by an object held by the hand.
- **Guide:** implies one, or frequently several, fixations between two objects that are approaching each other.
- **Check:** fixation at a location where the state of some variable is being assessed.

Additionally, Land et al. found that the hands, objects the hands have made contact with, and objects with a familiar manipulation are rarely gazed upon during interaction [Land et al. 1999]. Work has also shown that people will sometimes look away from a target before it has been acquired, an effect known as the "Late-Triggers error" [Kumar et al. 2008].

Hand-eye coordination has also been investigated for virtual metaphors such as the cursor and found similar results. The cursor generally lags behind the gaze which is used to direct the cursor [Huang et al. 2012; Smith et al. 2000]. Additionally, if the target location is predictable, top-down knowledge can enable users to initiate pointer movements before target fixation [Bieg et al. 2010].

Knowledge gained from fundamental hand-eye coordination studies has been used for gaze prediction from manual movements [Huang et al. 2012], gaze-enhanced manual input [Zhai et al. 1999] and multimodal techniques that integrate hand and eye input [Pfeuffer et al. 2014, 2015]. Techniques that combines hands and eyes have also been developed for VR contexts [Pfeuffer et al. 2017; Zeleznik et al. 2005]. However, gaze behaviour during hand interaction in VR remains unexplored and more knowledge may inspire further interaction techniques. We surmise that gaze patterns on interacted objects as a result of hand-eye coordination may be leveraged for implicit and continuous eye tracking re-calibration.

2.2 Eye Tracking Calibration

All eyes are shaped differently, and calibration is needed to account for each user's unique characteristics and track them accurately [Guestrin and Eizenman 2006]. Eye tracking calibration is traditionally performed by users fixating their eyes at predefined points (usually 5-9 points). At each calibration point, the eye tracker extracts eye image features and associates their positions in the eye image with the position of the calibration point. Traditional calibration is generally associated with challenges such as light conditions, eye anatomy, glasses, mascara, and the distance between the user and the eye tracker – all affecting the result of the calibration [Holmqvist et al. 2011]. These factors may change between and during sessions, and the calibration may deteriorate over time. It is, therefore, preferable to calibrate before each session and sometimes several times during a session. The calibration procedure can therefore easily become time-consuming and tedious [Flatla et al. 2011]. Another challenge is the calibration process itself. Fixating on a point for an extended time is for many considered unnatural [Chen and Ji 2011; Tripathi and Guenter 2016] as humans usually perform three to four fixations per second [Pfeuffer et al. 2013].

Research has been made to make the process less tedious by gamifying the calibration process [Flatla et al. 2011] or integrating the process into the environment [Gomez and Gellersen 2017; Renner et al. 2011] and found the calibration to be more enjoyable and more comfortable to use. Research has also been done on pursuit calibration where the user follows moving stimuli on the screen, and the positions of the eyes are mapped to the positions of the stimuli [Pfeuffer et al. 2013; Tripathi and Guenter 2016]. Smooth pursuits calibration has the advantage that it is more natural for humans to follow moving targets than to perform extended fixations. However, the method requires moving targets to follow and requires concurrent objects to move in different trajectories for accurate calibration/selection [Vidal et al. 2013].

Gaze calibration can also deteriorate and become inaccurate over time [Nyström et al. 2013]. Work has therefore been done into re-calibration methods that asses the current calibration drift and only re-calibrate the parts that are needed, using smooth pursuits [Gomez and Gellersen 2018] or standard calibration points [Lander et al. 2016]. However, while these methods are less time-consuming than a full calibration, they still require the user to pause their current interaction to perform the needed re-calibration points or movements. While no study to our knowledge has studied the extent of gaze calibration deterioration in VR settings, the gaze calibration might be affected if the HMD moves with respect to the head while the user is moving around or if the user takes off the HMD and puts it on again.

3 HAND INTERACTION FOR EYE TRACKING RE-CALIBRATION IN VR

Modern VR devices (HTC Vive [HTC 2016], Oculus Rift [Oculus VR 2016] and, PlayStation VR [Sony Computer Entertainment 2016]) come with hand-held motion controllers that are designed so that users can interact via their hands in the virtual world via pushing, pulling, picking objects up, and so on naturally as in the real world. Research has shown that hands and eyes are coordinated during hand interaction and that humans look at objects when interacting with them. Additionally, studies have indicated that visual processing [Perry et al. 2016], as well as oculomotor performance [Steinbach and Held 1968], are improved close to the hand. As such the gaze patterns associated with such interactions can potentially be used for eye tracking re-calibration in VR by treating the interacted object as a calibration point. The calibration process can then be completely hidden inside the VR environment, implicitly running while users interact with the environment. The proposed method can also be performed continuously in a seamless manner as users interact with objects. The proposed method will generally only have one calibration point during a single interaction (the interacted object). As such, multiple interactions would be required to gather calibration data across the whole field of view (FOV). Therefore, we argue the method would be most suitable for continuous re-calibration of an already existing calibration. However, the question remains whether the proposed re-calibration method is feasible for practical use. Hand interaction in VR differs from real-world hand interaction as a hand-controller that acts as a proxy for the hand is used. How the hand-controller affects users' gaze on interacted objects has to our knowledge not been investigated. Additionally, the question remains whether fixations gathered on interacted objects are consistent enough to warrant hand-eye coordination as a re-calibration method.

When evaluating a calibration method, three factors need to be taken into account. For calibration to be successful, the eye tracker needs a sufficient amount of time fixating on the target to estimate all the needed parameters, generally less than one second. It is also important that the user's gaze is fixated at the calibration point since they will be used for future reference. The maximum angular difference is device-specific, but generally at 1°. Additionally, the calibration points should cover the whole FOV for accurate tracking in all directions [Holmqvist et al. 2011]. Traditional calibration can reliably provide good results for these factors among many users and is therefore widely used. Therefore, measuring these fixation parameters on interacted objects will tell whether the coordination method has potential for calibration. Our analysis focuses on these three factors for eye tracking calibration. However, the results are also applicable to other interaction paradigms that leverage gaze patterns during hand interaction in VR.

4 METHOD

Multiple different interactions were internally prototyped and tested to find suitable interactions for the evaluation. The interactions chosen were a draggable slider, a turnable knob and a liftable cube that all are common 3D interactions which require assumingly long enough input and precision to be completed successfully (see Fig. 1). The required precision would help avoid the "Late-Triggers error" and maximise fixations. Furthermore, they were all single targets which should increase the probability of attaining gaze fixations. When interacting with the slider, participants were instructed to adjust the slider to the correct position, showcased by the handle

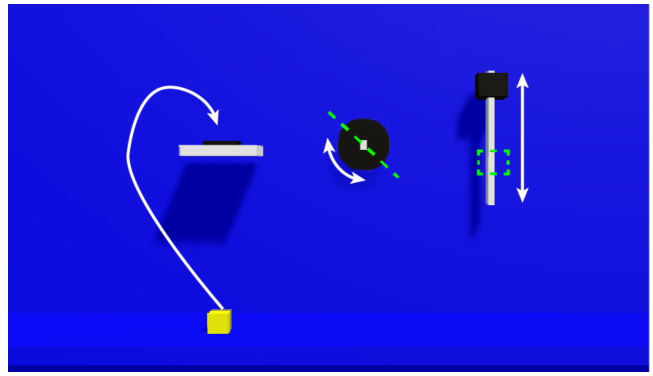

Figure 1: The three interactions used for the study. Left: a cube that participants placed on the drop area. Middle: a knob that participants rotated to the correct position. Right: a slider that participants moved to the correct position.

changing colour. For the knob, participants were instructed to turn the knob to the correct position, shown via feedback. The goal position for both objects was random and not visible beforehand. Hence the participant needed to search for the correct position. When interacting with the cube, participants were tasked to place the cube in a marked area. The cube had no additional visual feedback indicating completion, but the target was clearly visible. The lack of feedback was chosen to see if the lack of feedback would influence the prevalence of "Check" fixations. All objects had a size of 2.5-3.5° at 1 metre's distance. The objects' visual size was thus dependent on the distance to the user. We wanted to investigate if no additional transformations were needed to attain usable fixations on the object for calibration. An HTC Vive with a Tobii integrated eye tracker with a reported 0.5° accuracy and frequency of 120Hz was used for the study [Tobii 2018]. The HTC Vive tracking system was used for users positional tracking while performing the interactions. Previous research has highlighted the limitations of the HTC Vive [Niehorster et al. 2017], however we are not interested in the measurements that are considered to be problematic and the participants are unlikely to move across a large area as they will be standing in front of the interactions.

One HTC Vive hand-controller for each hand was used to interact with the objects, and a model of the controller was used to show its position in VR. Participants had to touch an object with the controller and hold down the trigger button to grasp a target. Previous research has shown that fixation positions may be affected when an object is occluded [de Grave et al. 2008; Vishwanath et al. 2000]. Thus, to avoid occlusion, the virtual model of the controller was turned invisible when the object was grabbed, presumingly making fixation positions on the objects more predictable.

4.1 Procedure

The evaluation was conducted with 15 volunteers (9 male, 6 female, 30.74 ±4.65 years old) to investigate the fixation behaviour on interacted objects during hand interaction in VR. Participants first signed a consent form and answered a demographic questionnaire.

Two participants reported daily VR experience. The remaining participants reported none or occasional experience. Six participants reported occasional eye tracking experience, while the rest reported daily or weekly experience with eye tracking. Fourteen participants reported the right hand to be their dominant hand. One person reported their left hand to be dominant. The participants then explored a VR environment to get accustomed to VR for 5 minutes or until they felt comfortable enough to start. A 5-point eye tracking calibration was performed before starting the experiment. The experiment was arranged in steps, where one to three interactions were presented to the participants at a time. Participants were instructed to move around freely in standing position and perform the presented tasks to their discretion in whatever order they wished to simulate natural interaction behaviour. When all interactions were completed, a new combination of interactions was presented. Each combination was different in the number of interactions presented, ranging from 1 to 3 and the position of each interaction relative to each other. No combination had more than one interaction of the same type. In total, the participants interacted with 15 different combinations ordered randomly. The evaluation ended with a semi-structured interview to get their thoughts on their approach to solve the tasks during the study.

4.2 Analysis

The analysis was performed between participants and interactions. The interactions were separated into three phases for data analysis. The *reach phase* starts when participants' gaze first targets the interacted object and ends when participants grab the object. The *manipulation phase* starts when participants grab the interacted object and ends when it is released. The *release phase* starts when participants release the interacted object and ends when the participants' gaze no longer targets the object.

We conducted a two-way repeated measures ANOVAs to discover differences in total fixation duration, the number of fixations and total phase duration with the interactions the phases as independent factors. Eventual extreme outliers found via box-plots were replaced with the largest non-extreme value. All distributions were assessed via a Q-Q plot. The assumption of sphericity was assessed via Mauchly's test of sphericity and if violated a Greenhouse-Geisser correction was applied. Post hoc analysis to compare group differences was made with a Bonferroni adjustment.

All interactions were scaled to the median time of each object to calculate the occurrence of a fixation at any time during an interaction. The fixation percentage at a specific time during the interaction was calculated via the percentage of the normalised interactions that had a fixation at that time (computed for 100ms bins). An I-DT algorithm with a duration threshold of 100ms and a dispersion threshold of 1° of the visual angle was used to filter fixations [Salvucci and Goldberg 2000].

5 RESULTS

To evaluate gaze patterns on objects during hand interaction we will here go through relevant data from our study. The analysis showed no significant differences in results between participants. Therefore, the results presented in this section derives from all recorded interactions. We will begin by presenting our data on the

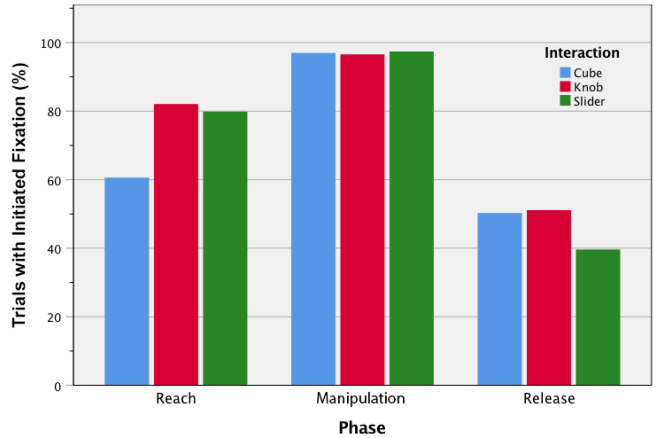

Figure 2: The percentage of all interactions that started at least one fixation during each phase of interaction.

occurrence of attaining fixations, various important parameters associated with fixations, and the positions of the fixations relative to the object and the FOV. We end our results with a short section on general observations on the participant behaviour and responses.

5.1 Fixation Occurrence

Fig. 2 shows, for each object, the occurrence of starting a fixation on the object. The manipulation phase had the highest percentage of trials for all objects, reaching over 95% for all. The reach phase had a lower percentage of trials with fixations ranging between 62% for the cube to 80% for the knob and slider. The release phase had the lowest percentage for all interactions ranging from 50% for the cube and knob to 40% for the slider.

All three interactions had different points for when it was most common that a participant would fixate on the interacted object (Fig. 3). During the reach phase, it was more common to acquire a fixation on the interacted object for all three interactions when closer to the manipulation phase. The cube had an overall lower occurrence of fixations during the end of the reach phase (40%) compared to the knob and slider (60%), coinciding with the results in Fig. 2, where the percentage for the cube in the reach phase was considerably lower.

Fig. 3 also shows that the three interactions had mostly different fixation occurrence curves during the manipulation phase. The cube showed a large dip during the start of the interaction when the participant's attention would shift from the cube to the marked area and would after that stay on the marked area. The probability of the participant fixating would then steadily increase as the cube got closer towards the drop area reaching 80% towards the end of the manipulation phase.

The knob's manipulation phase had a relatively stable fixation occurrence at 60% during the whole phase, likely because the knob was stationary. The fixation occurrence falls at the end of the manipulation phase, coinciding with when the participant gets visual feedback that the knob has been rotated into the correct position. Participants would then often shift their attention away from the knob before releasing it.

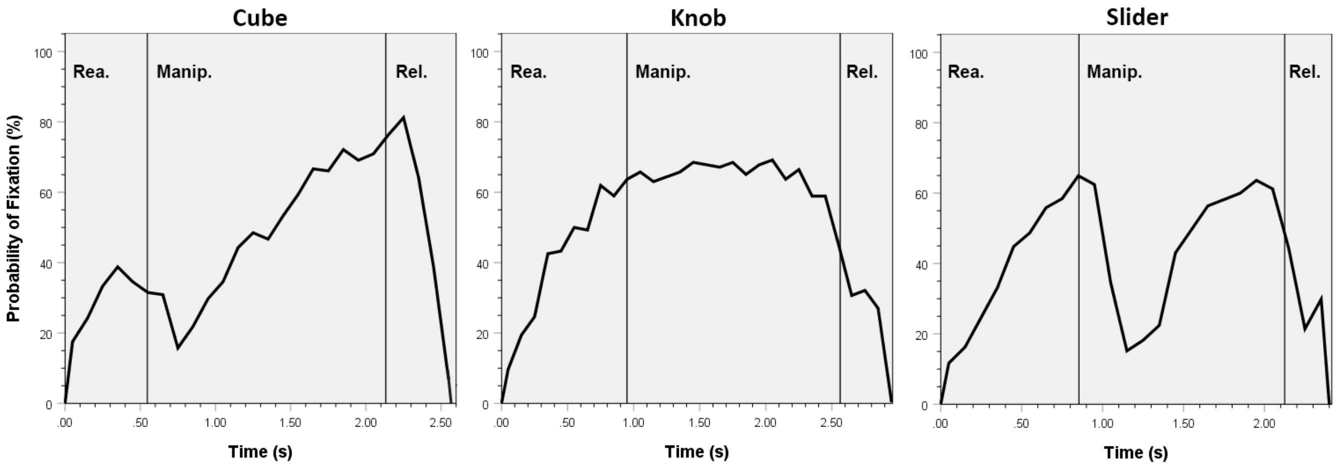

Figure 3: The percentage of all normalised interactions with a fixation on the interacted object at any given time.

A notable dip in fixations was found at the start of the slider's manipulation phase as participants would often first quickly drag the slider through its whole range to find the approximate correct position. Fixations are then hard to attain on the object as the participant's gaze is following the sliders fast movement. When the approximate correct position had been found, the participant dragged the slider more carefully into the correct position, leading to an increase in fixations. At the end of the manipulation phase, fixations rapidly dropped similarly to the knob.

During the release phase, the slider and knob had an earlier fixation drop than the cube. The cube's initial spike in fixations is due to the participant more frequently double-checking that the cube was placed inside the drop area after releasing it due to the lack of other feedback. The knob and slider, however, had a lower need for double-checking as the participant received clear feedback when the interaction was completed.

5.2 Fixation Characteristics

Table 1 show the fixation characteristics for each interaction and phase. Two-way repeated measures ANOVA showed significant differences in fixation duration ($F(3.08, 379.24) = 22.66$, $p < .001$), fixation count ($F(3.27, 401.85) = 8.60$, $p < .001$), and phase duration ($F(3.22, 395.97) = 12.21$, $p < .001$). Further analysis showed significant differences in all characteristics when comparing within each object (Table 2) and each phase (Table 3).

Table 1: Mean gaze fixation characteristics with standard deviations for each respective interaction and phase.

Phase	Total Fixation Duration (s)	Number of Fixations (#)	Total Phase Duration (s)
Cube Reach	.19 ± .21	1.22 ± 1.13	.58 ± .35
Knob Reach	.57 ± .58	2.94 ± 2.58	1.19 ± .75
Slider Reach	.46 ± .43	2.53 ± 2.26	.97 ± .63
Cube Manip.	.98 ±.53	4.43 ± 2.37	1.71 ± .67
Knob Manip.	1.15 ±.75	5.64 ± 3.51	1.86 ± .92
Slider Manip.	.65 ±.47	4.06 ± 2.70	1.15 ± .65
Cube Release	.16 ±.22	.78 ± 1.07	.42 ± .30
Knob Release	.17 ±.27	1.11 ± 1.40	.59 ± .56
Slider Release	0.08 ±.12	.52 ± .81	.33 ± .28

When comparing the phases within each object, post-hoc analysis showed that for all objects, the manipulation phase had a significantly higher fixation duration (all $p < .001$), fixation count (all $p < .001$), and phase duration (all $p < .001$) than the reach and release phase. Additionally, for the knob and slider, the reach phase had a significantly higher fixation duration (all $p < .001$), fixation count (all $p < .001$) phase duration (all $p < .001$) than the reach phase. The cube showed showed a significant difference in fixation count ($p = .012$) and phase duration (p = .001), but no significance in fixation duration ($p = 1.000$). The results indicates that the manipulation phase is the longest phase with the most fixations.

When comparing the items within each phase, post-analysis showed that both the knob and slider had significantly higher values compared to the cube (all $p < .001$) during the reach phase, indicating that the initial position of the object has a significant effect on reaching time and on their fixation behaviour. During the manipulation phase, the knob had significantly higher fixation duration than the cube ($p = .046$) and slider ($p < .001$). Additionally, the cube had significantly higher fixation duration compared to the slider ($p < .001$). The knob also had a significantly higher fixation count than the cube ($p = .002$) and slider ($p < .001$). Finally, both the

Table 2: ANOVA of fixation characteristics with the object as the independent factor.

Object	Total Fixation Duration	Number of Fixations	Total Phase Duration
Cube	$F(1.45, 178.14) = 207.31$ $p < .001$	$F(1.64, 201.93) = 168.09$ $p < .001$	$F(1.59, 195.73) = 288.95$ $p < .001$
Knob	$F(1.63, 200.85) = 103.05$ $p < .001$	$F(1.75, 214.84) = 102.46$ $p < .001$	$F(1.83, 225.35) = 101.16$ $p < .001$
Slider	$F(1.55, 190.57) = 72.53$ $p < .001$	$F(1.58, 193.69) = 89.63$ $p < .001$	$F(1.68, 206.56) = 142.63$ $p < .001$

Table 3: ANOVA of fixation characteristics with the interaction phase as the independent factor.

Phase	Total Fixation Duration	Number of Fixations	Total Phase Duration
Reach	$F(1.83, 182.96) = 21.10$ $p < .001$	$F(2, 202) = 14.64$ $p < .001$	$F(1.67, 206.60) = 28.26$ $p < .001$
Manip.	$F(1.81, 249.95) = 26.22$ $p < .001$	$F(1.87, 255.74) = 24.50$ $p < .001$	$F(1.89, 308.48) = 11.83$ $p < .001$
Release	$F(1.32, 77.85) = 12.16$ $p < .001$	$F(1.23, 73.65) = 11.72$ $p < .001$	$F(1.60, 198.38) = 14.70$ $p < .001$

cube ($p < .001$) and knob ($p < .001$) had a significantly higher phase duration than the slider. The results suggest that stationary objects are favourable in acquiring fixations on the object. Additionally, a larger phase duration also seems significant in acquiring fixations. For the release phase, the knob had a significantly higher fixation duration, fixation count, and phase duration than the slider (all $p < .001$). The knob also had a significantly higher phase duration than the cube ($p = .007$). Finally, the cube's fixation duration was significantly longer than the slider ($p < .001$).

5.3 Position of Fixations

Fixation position data showed a tendency of the fixation centroids to be 0.5-1° above the object centre (Fig. 4). However, fixations were mainly positioned at the centre of the interacted object in the x-axis. Both axes showed a normal distribution around its centre value, where fixations on objects could reach up to 4° away from the centre value. No significant differences were found between the interactions, nor phases.

The results also indicated that the choice of hand to interact with impacted where on the participant's FOV fixations would occur. We found that 86% of the fixations were on the left side of the FOV when using the left hand. When using the right hand, 79% of the fixations occurred on the right side in the FOV. No significant differences were found between the interactions.

5.4 Participant Behaviour and Interviews

Participants appeared to be more careful during the first interactions and became more consistent in their interactions as they learned. Participants mainly used their dominant hand when completing interactions, although after several trials some participants also used their other hand. Furthermore, three participants started experimenting with the interactions by using both hands to complete multiple interactions at the same time. Participants generally interacted with the objects at approximately an arm's length and actively positioned themselves to comfortably grab the objects. The data showed a mean distance of 0.56 meters between the participants' eyes and the cube. The knob and slider had similar mean distances of 0.66 meters and 0.63 meters respectively.

The consensus among the participants was that the interactions felt natural and easy. However, P13 commented *"It felt uneasy to drop the cube as it gave a sense of lost control and therefore it required more attention"*. Many participants stated (9/15) that they were more careful in the beginning and got quicker and more comfortable after more trials. Furthermore, some participants (3/15) claimed

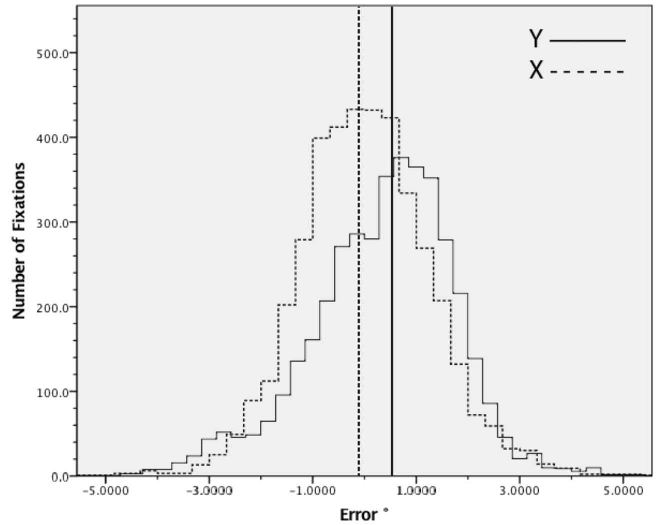

Figure 4: Angular difference in the horizontal (x) and vertical (y) plane between the fixation centroid and object centre for all fixations. Vertical lines represent medians.

that they attempted to complete two interactions simultaneously further into the study. This behaviour indicates that interaction behaviour may change over time. Most participants did not report any noticeable difference in behaviour depending on the number of present interactions. However, a few participants (4/15) stated that they acted differently, for instance by switching hands more often.

6 DISCUSSION

In this paper, we report on an evaluation of fixations on interacted objects for three different interactions that required sufficient time and precision to be completed. The participants' visual attention between the three interactions showed different results, which indicates that multiple factors affect gaze behaviour on interacted objects and needs to be taken into consideration when evaluating if an interaction is suitable for eye tracking calibration. We will in this section discuss these found factors.

6.1 Fixations During the Interaction Phases

During the reach phase, the cube had a lower occurrence fixations as well as a significantly lower total fixation duration, number of fixations, and phase duration than the knob and the slider. The cube was placed on a shelf underneath its drop area while the knob and slider were placed at the drop area's height (Fig. 1). The cube was, therefore, closer to participants' hands when the arms were in a resting position, and the participants did not have to reach as far to grab it. Additionally, no significant differences were found between the knob and slider, indicating that placement of the interacted object is important during the reach phase to maximise the probability of fixations. However, the general fixation occurrence during the reach phase was low. A possible explanation could be that "Locate", and "Direct" fixations at the start of an interaction are not only located on the interacted object but also on the surrounding environment as the participants gather initial information.

The knob had a stable fixation occurrence during the manipulation phase, likely due to the knob being stationary during the interaction, only rotating around its axis. The cube had a higher fixation occurrence towards the end of the manipulation phase when the participant is focusing towards the stationary drop area. Additionally, the slider had a higher fixation occurrence during its later stages when participants were closer to the correct position and were dragging the slider slower to fine tune the slider to the correct position. The combined results indicate that users are more likely to fixate on an object when more precision is required. Additionally, the results indicate that participants' general task completion approach should be taken into account when considering an interaction for calibration purposes. For example, participants tended to be less accurate and focussed on the interacted object at the start of an interaction while they were fixating more on the object towards the end when more precision was needed. The three interactions also indicate a higher likelihood of attaining fixations for stationary objects during the interaction. In other words, a stationary manipulation phase is more suitable for a fixation based calibration compared to a moving interaction phase. However, moving objects could be used for pursuits based calibration during hand-interaction. Previous research has shown that users' oculomotor performance is enhanced for pursuits [Chen et al. 2016] or vergence [Maiello et al. 2018] when the hands control the target. As such, with the slider as an example, it could be possible to calibrate the eye tracker if the user performs a pursuit movement following the slider while dragging the slider. The difficulty lies in designing the interaction to ensure that the object's movement is possible to follow with a pursuit movement during an interaction. Further work should investigate this type of pursuit-based calibration.

During the end of the manipulation phase, the probability of a fixation increased for the cube while dropping for the knob and slider. These results are likely due to that the knob and slider provided feedback when they were in the correct position, while participants were still holding the interacted object. For the cube, however, there was no direct feedback and participants did not know that the interaction was completed until they had released the cube and could see that it was placed correctly within the drop area. As such the lack of feedback forced participants to perform more "Check" fixations during the interaction. From a calibration perspective, it suggests that the calibration should take into account how and when feedback is given to the participant. Additionally, the results may also indicate that having more disruptive interactions with less feedback could lead to attaining more fixations as users have to perform more "Check" fixations. This indication is further supported by comments from the participants who stated that they felt like they needed to keep more attention towards the cube because they did not receive any clear feedback.

The release phase had a lower fixation occurrence for all interactions compared to the reach and manipulation phase. During the release phase, it was common that participants moved their visual attention away from the interacted object before releasing it. For the cube, however, most fixations that were seen in the spike in Figure 4 were fixations that started in the manipulation phase and were carried over to the release phase. These results indicate that the release phase is difficult to use for calibration purposes if clear feedback is given.

For all three phases, the interaction with the longest phase duration also had the longest total fixation duration. Additionally, the interaction with the shortest phase duration had the shortest total fixation duration. These results hint that it may be beneficial to use sufficiently long interactions to extend the total fixation time and attain more fixations. All three interactions had a higher probability of fixating on the object (Fig. 3), longer total fixation duration, the number of fixations and phase duration (Fig. 5) during the manipulation phase compared to their reach phase and release phase counterparts. These results further indicate that it is most beneficial to conduct an eye tracking calibration during the manipulation phase and not when reaching for or releasing an object.

6.2 Fixation Positions

Results showed that participants were fixating slightly above the objects' centre instead of fixating directly at their centre (Fig. 4). A possible explanation could be the placement of the interacted objects. The objects were placed at a 1.3 meters height so that participants could comfortably reach the objects leading to participants looking down on the interacted objects. Additionally, Fig. 4 shows that the fixations were spread out on the objects. While the majority of fixations were located within a 1° centre, a large part could also be found up to 3-4° distance, which is larger than what is required for an accurate eye tracking calibration (0.5-1°). The results can be explained by considering the size of the objects together with the average interacted distance of ~0.6 metres. Object sizes at those distances spanned up to 7-8° of the FOV which lead to a larger spread of fixations on the objects, thus indicating that the size of the interacted object is vital to the calibration method's success.

Several approaches could be made to guarantee the appropriate object size. A naive approach would be to set the target size to be 1° or less at the regular interaction distance. A second approach could be to dynamically change the object size relative to users' viewport, by for example adjusting the object's distance to the user or by changing the object size. Finally, objects can be placed at the edge of users' reach to guarantee that objects remain within a certain size or be placed outside users' reach and thus only be reachable via for example amplified hand movement [Wilson et al. 2018].

The results also showed that using the left or right hand had an effect on which side of the x-axis on the FOV fixations occur when interacting with an object. Participants would rarely shift their heads entirely towards an interacted object but rather slightly towards it. This behaviour implies that it is possible to attain fixations that are spread over a large part of the FOV. Further investigation on how the object position relative to the participants' head affects fixation position on the FOV needs to be conducted.

6.3 Participant Behaviour

Participants expressed that they were more careful and took more time to complete the interaction due to its unfamiliarity during the first trials of each interaction. This feedback may indicate that the first time performing an interaction will have the highest probability of attaining hand-eye coordinated fixations. Similarly, performing interactions that are familiar to the user may lead to fewer fixations or the "Late-Triggers error".

A potential pitfall in using interactions for calibration is the need to ensure that the user is paying attention to the current interaction. The design of the interaction is one aspect, but there are also other factors to consider. Some participants tried to complete multiple interactions at the same time using one hand for each interaction. Participants would then constantly switch their visual attention between the two interactions; thus the likelihood of attaining useful fixations becomes smaller. To avoid this behaviour, an interaction that is used for calibration should not be in the direct vicinity of another possible interaction.

Additionally, the surrounding environment may have an impact on participants' attention towards an interacted item. Using visual tricks to exploit humans' automatic reactions to events could potentially be used to increase the fixation probability [Irwin et al. 2000]. For example, if dragging a slider that suddenly blinks, the user would instinctively fixate at the slider and data could be sampled. Also, making sure the interacted object sticks out in the VR environment through colour contrasts or other mediums may help to keep users' attention on the object. Another aspect to consider is that the participants in this evaluation were put under no time pressure to complete the interactions. In a game setting where users may be under time pressure, it is possible that the user's visual attention may shift more often in order to plan or look out for enemies.

6.4 Practical Application

Our results show that it is possible to attain fixations on interacted objects during hand interaction. While the fixation occurrence may seem low and therefore unreliable, we believe that is partly due to our strict fixation criteria. With looser requirements as in other studies, the number should be overall higher [Johansson et al. 2001]. However, our study was performed in a lab environment where eye tracking accuracy was assured during the whole experiment. A practical implementation of the proposed re-calibration method should be able to identify when users fixate on an item during an interaction, even if the calibration is not perfect.

Looking at the data, there seem to be some viable options. The most straightforward approach would be to treat the object as a regular calibration point and only record gaze data when it is most likely to acquire a fixation, thereby assuming that the user is fixating on the object. For example after a disruptive event or feedback, or when an interacted object is close to its goal position. This method, however, may still be unreliable as it is difficult to fully guarantee that the user is fixating at the object at a specific time. Additionally, the interacted object size has to be considered as we found that not all fixations were located on the object centre.

Another approach could be to exploit the fixation positions on the interacted objects. In Fig. 4 we saw a normal distribution of the angular difference between the gaze fixation and the object position for all fixations of the experiment. Additionally, similar normal distributions were also found for each participant separately. If the normal distribution can be found uncalibrated, then it may be possible to adjust the calibration of the eye tracking with the help of the normal distribution. This proposed technique assumes that the distribution centre (where the users look the most) during the interaction with the object is correct; in other words, that they are looking at the object. It might be necessary to accumulate data from a series of hand-eye coordinated interactions to generate a sufficient data-set for the normal distribution curve. However, this direction needs further research.

6.5 Study Limitations

Finally, we consider a few limitations of the user study. First, a possible limitation of the results is the I-DT algorithm mislabelling of pursuit eye movements as fixation which may affect the results. However, the eventual pursuit movements are limited to a short distance. Second, the study only considers a small number of interactions. Our interactions were based on interactions that are common in VR and that we believed would be suitable for eye tracking calibration. However, more types of common interactions should be studied. Third, the environment used in our study was a simple environment without any distracting elements except for the interactions themselves. It would be reasonable to believe that distracting environments would affect gaze behaviour. Fourth, further studies should be conducted on how different sized objects and dynamic object sizes affect gaze patterns. Longitudinal studies on gaze pattern changes would also be interesting since qualitative results showed that participants got more confident after performing interactions multiple times. Additionally, matching the gaze data with additional kinematic data from the hand could provide further insights. Finally, the proposed re-calibration method has to be implemented and evaluated to assess its feasibility fully. Nevertheless, our study is a first step in investigating gaze during hand interaction in VR.

7 CONCLUSION

This study aimed to investigate gaze fixations on interacted objects during hand interaction in VR and the feasibility of attaining related fixations as a method for eye tracking re-calibration. First, the results show that the manipulation phase is most suitable for calibration with a relatively high probability of attaining fixations. Second, it is more likely to attain fixations when interacting with stationary objects compared to moving objects. Third, interactions which required sufficient precision and time to complete positively influence the probability of attaining fixations. Fourth, carefully designed feedback and disruptive elements of interactions may momentarily increase the probability of fixations, but the probability often drops rapidly after task completion. Fifth, there was a tendency among participants to complete tasks using their dominant hand in their dominant hand's part of the FOV. This behaviour needs to be counteracted in order to achieve accurate eye-tracking calibration covering the whole FOV. Lastly, the surrounding VR environment may have an impact on participant behaviour and thereby fixation probability, a less distracting environment with less reachable interactions may be beneficial to attain fixations on the currently interacted object.

REFERENCES

Dana H. Ballard, Mary M. Hayhoe, F. Li, Whitehead S. D., J. P. Frisby, J. G. Taylor, and R. B. Fisher. 1992. Hand-Eye Coordination during Sequential Tasks. *Philosophical Transactions of the Royal Society B: Biological Sciences* 337, 1281 (sep 1992), 331–339. https://doi.org/10.1098/rstb.1992.0111

Dana H. Ballard, Mary M. Hayhoe, and Jeff B. Pelz. 1995. Memory Representations in Natural Tasks. *Journal of Cognitive Neuroscience* 7, 1 (dec 1995), 66–80. https://doi.org/10.1162/jocn.1995.7.1.66

Hans-Joachim Bieg, Lewis L. Chuang, Roland W. Fleming, Harald Reiterer, and Heinrich H. Bülthoff. 2010. Eye and Pointer Coordination in Search and Selection Tasks. In *Proceedings of the 2010 Symposium on Eye Tracking Research & Applications (ETRA '10)*. ACM, New York, NY, USA, 89–92. https://doi.org/10.1145/1743666.1743688

Jiu Chen and Qiang Ji. 2011. Probabilistic gaze estimation without active personal calibration. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*. IEEE, 609–616. https://doi.org/10.1109/CVPR.2011.5995675

Jing Chen, Matteo Valsecchi, and Karl R. Gegenfurtner. 2016. LRP predicts smooth pursuit eye movement onset during the ocular tracking of self-generated movements. *Journal of Neurophysiology* 116, 1 (2016), 18–29. https://doi.org/10.1152/jn.00184.2016

Denise D. J. de Grave, Constanze Hesse, Anne-Marie Brouwer, and Volker H. Franz. 2008. Fixation locations when grasping partly occluded objects. *Journal of vision* 8, 7 (may 2008), 1–11. https://doi.org/10.1167/8.7.5

David R. Flatla, Carl Gutwin, Lennart E. Nacke, Scott Bateman, and Regan L. Mandryk. 2011. Calibration Games: Making Calibration Tasks Enjoyable by Adding Motivating Game Elements. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 403–412. https://doi.org/10.1145/2047196.2047248

Argenis Ramirez Gomez and Hans Gellersen. 2017. GazeBall: Leveraging Natural Gaze Behavior for Continuous Re-calibration in Gameplay. *Journal of Eye Movement Research* 10, 6 (2017), 3. https://bop.unibe.ch/index.php/JEMR/article/view/4182

Argenis Ramirez Gomez and Hans Gellersen. 2018. Smooth-i: Smart Re-calibration Using Smooth Pursuit Eye Movements. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA '18)*. ACM, New York, NY, USA, Article 10, 5 pages. https://doi.org/10.1145/3204493.3204585

Elias D. Guestrin and Moshe Eizenman. 2006. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering* 53, 6 (jun 2006), 1124–1133. https://doi.org/10.1109/TBME.2005.863952

Werner F. Helsen, Digby Elliott, Janet L. Starkes, and Kathryn L. Ricker. 1998. Temporal and Spatial Coupling of Point of Gaze and Hand Movements in Aiming. *Journal of Motor Behavior* 30, 3 (sep 1998), 249–259. https://doi.org/10.1080/00222899809601340

Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Jarodzka Halszka, and Joost van de Weijer. 2011. *Eye Tracking : A Comprehensive Guide to Methods and Measures*. Oxford University Press, Oxford, United Kingdom. 560 pages.

HTC. 2016. HTC Vive. Retrieved February 20, 2019 from https://www.vive.com

Jeff Huang, Ryen White, and Georg Buscher. 2012. User See, User Point: Gaze and Cursor Alignment in Web Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1341–1350. https://doi.org/10.1145/2207676.2208591

David E. Irwin, Angela M. Colcombe, Arthur F. Kramer, and Sowon Hahn. 2000. Attentional and oculomotor capture by onset, luminance and color singletons. *Vision Research* 40, 10 (2000), 1443 – 1458. https://doi.org/10.1016/S0042-6989(00)00030-4

Roland S. Johansson, Göran Westling, Anders Bäckström, and J. Randall Flanagan. 2001. Eye–Hand Coordination in Object Manipulation. *Journal of Neuroscience* 21, 17 (sep 2001), 6917–6932. https://doi.org/10.1523/JNEUROSCI.21-17-06917.2001

Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. 2009. Learning to predict where humans look. In *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2106–2113. https://doi.org/10.1109/ICCV.2009.5459462

Manu Kumar, Jeff Klingner, Rohan Puranik, Terry Winograd, and Andreas Paepcke. 2008. Improving the Accuracy of Gaze Input for Interaction. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications (ETRA '08)*. ACM, New York, NY, USA, 65–68. https://doi.org/10.1145/1344471.1344488

Michael Land, Neil Mennie, and Jennifer Rusted. 1999. The Roles of Vision and Eye Movements in the Control of Activities of Daily Living. *Perception* 28, 11 (1999), 1311–1328. https://doi.org/10.1068/p2935

Michael Land and Benjamin Tatler. 2012. *Looking and Acting: Vision and Eye Movements in Natural Behaviour*. Oxford University Press, Oxford, United Kingdom. https://doi.org/10.1093/acprof:oso/9780198570943.001.0001

Christian Lander, Frederic Kerber, Thorsten Rauber, and Antonio Krüger. 2016. A Time-efficient Re-calibration Algorithm for Improved Long-term Accuracy of Head-worn Eye Trackers. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA '16)*. ACM, New York, NY, USA, 213–216. https://doi.org/10.1145/2857491.2857513

Guido Maiello, MiYoung Kwon, and Peter J. Bex. 2018. Three-dimensional binocular eye–hand coordination in normal vision and with simulated visual impairment. *Experimental Brain Research* 236, 3 (01 Mar 2018), 691–709. https://doi.org/10.1007/s00221-017-5160-8

Diederick C. Niehorster, Li Li, and Markus Lappe. 2017. The Accuracy and Precision of Position and Orientation Tracking in the HTC Vive Virtual Reality System for Scientific Research. *i-Perception* 8, 3 (2017), 2041669517708205. https://doi.org/10.1177/2041669517708205

Marcus Nyström, Richard Andersson, Kenneth Holmqvist, and Joost van de Weijer. 2013. The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods* 45, 1 (01 Mar 2013), 272–288. https://doi.org/10.3758/s13428-012-0247-4

Oculus VR. 2016. Oculus Rift. Retrieved February 20, 2019 from https://www.oculus.com/rift

Jeff Pelz, Mary Hayhoe, and Russ Loeber. 2001. The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research* 139, 3 (aug 2001), 266–277. https://doi.org/10.1007/s002210100745

Carolyn J. Perry, Prakash Amarasooriya, and Mazyar Fallah. 2016. An Eye in the Palm of Your Hand: Alterations in Visual Processing Near the Hand, a Mini-Review. *Frontiers in Computational Neuroscience* 10 (2016), 37. https://doi.org/10.3389/fncom.2016.00037

Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. Gaze-touch: Combining Gaze with Multi-touch for Interaction on the Same Surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 509–518. https://doi.org/10.1145/2642918.2647397

Ken Pfeuffer, Jason Alexander, Ming Ki Chong, Yanxia Zhang, and Hans Gellersen. 2015. Gaze-Shifting: Direct-Indirect Input with Pen and Touch Modulated by Gaze. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 373–383. https://doi.org/10.1145/2807442.2807460

Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17)*. ACM, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

Ken Pfeuffer, Mélodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. 2013. Pursuit Calibration: Making Gaze Calibration Less Tedious and More Flexible. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 261–270. https://doi.org/10.1145/2501988.2501998

Patrick Renner, Nico Lüdike, Jens Wittrowski, and Thies Pfeiffer. 2011. Towards Continuous Gaze-Based Interaction in 3D Environments - Unobtrusive Calibration and Accuracy Monitoring. In *Proceedings of the Workshop Virtuelle & Erweiterte Realität 2011*, Christian-A. Bohn and Sina Mostafawy (Eds.). Shaker Verlag, Aachen, Germany, 13–24.

Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA '00)*. ACM, New York, NY, USA, 71–78. https://doi.org/10.1145/355017.355028

Jeroen B. J. Smeets, Mary M. Hayhoe, and Dana H. Ballard. 1996. Goal-directed arm movements change eye-head coordination. *Experimental Brain Research* 109, 3 (jun 1996), 434–440. https://doi.org/10.1007/BF00229627

Barton A. Smith, Janet Ho, Wendy Ark, and Shumin Zhai. 2000. Hand Eye Coordination Patterns in Target Selection. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA '00)*. ACM, New York, NY, USA, 117–122. https://doi.org/10.1145/355017.355041

Sony Computer Entertainment. 2016. PlayStation VR. Retrieved February 20, 2019 from https://www.playstation.com/en-gb/explore/playstation-vr

Martin J. Steinbach and Richard Held. 1968. Eye Tracking of Observer-Generated Target Movements. *Science* 161, 3837 (1968), 187–188. https://doi.org/10.1126/science.161.3837.187 arXiv:http://science.sciencemag.org/content/161/3837/187.full.pdf

Yusuke Sugano and Andreas Bulling. 2015. Self-Calibrating Head-Mounted Eye Trackers Using Egocentric Visual Saliency. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 363–372. https://doi.org/10.1145/2807442.2807445

Tobii. 2018. Tobii Pro VR Integration. Retrieved February 20, 2019 from https://www.tobiipro.com/product-listing/vr-integration/

Subarna Tripathi and Brian Guenter. 2016. A Statistical Approach to Continuous Self-Calibrating Eye Gaze Tracking for Head-Mounted Virtual Reality Systems. *CoRR* abs/1612.06919 (2016), 9. arXiv:1612.06919

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, New York, NY, USA, 439–448. https://doi.org/10.1145/2493432.2493477

Dhanraj Vishwanath, Eileen Kowler, and Jacob Feldman. 2000. Saccadic localization of occluded targets. *Vision Research* 40, 20 (sep 2000), 2797–2811. https://doi.org/10.1016/S0042-6989(00)00118-8

Graham Wilson, Mark McGill, Matthew Jamieson, Julie R. Williamson, and Stephen A. Brewster. 2018. Object Manipulation in Virtual Reality Under Increasing Levels of Translational Gain. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 99, 13 pages. https://doi.org/10.1145/3173574.3173673

Robert C. Zeleznik, Andrew S. Forsberg, and Jürgen P. Schulze. 2005. *Look-That-There: Exploiting Gaze in Virtual Reality Interactions*. Technical Report. Brown University.

Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 246–253. https://doi.org/10.1145/302979.303053