Immersive Eye Tracking
Calibration in Virtual Reality Using Interactions with In-game Objects

LUDWIG SIDENMARK
**ABSTRACT**
This thesis aims to investigate an eye tracking calibration method in virtual reality where users’ visual attention from eye-hand coordination is used when interacting with in-game items. This could potentially allow eye tracking calibration without interrupting the virtual experience, in comparison with traditional eye tracking calibration which is cumbersome, disruptive, and requires the user’s full attention.

A user study was conducted with 15 participants in which they were tasked to complete three different interactions. The interactions tested were a knob, a slider and a liftable cube. Where in the virtual environment the participants were looking during the interactions was recorded and processed to enable comparison. The processed data was analysed to find factors that influenced the calibration method. Additionally, the results was analysed to find at what point during the interactions that has the most consistent eye tracking fixations on the interacted item, and therefore most potential for eye tracking calibration.

The results showed that when the participant is interacting with the item and the interacted item is stationary, we received a fixation for around 60% of all trials at any time. When the interacted item was moving the results indicated a lower percentage. To increase this number, the gaze data should be filtered instead of using raw gaze data in order to avoid flickering from the eye tracker.

Regarding factors that influence the calibration method, the choice of interaction has a big impact on the method’s success where interactions in which the interacted item is stationary has more potential. Additionally, interactions that take longer time and requires precision in order to complete the interaction positively influences the potential of the calibration method. The surrounding virtual environment also has an influence, as a more distracting environment can negatively impact the calibration method.

---

**Immersion Eye Tracking Calibration in Virtual Reality Using Interactions with In-game Objects**

**Ludwig Sidenmark**  
KTH, Royal Institute of Technology  
Stockholm, Sweden  
ludwigsi@kth.se

**Degree Project in Computer Science and Communication, Second Cycle, DA222X**  
**Supervisor:** Anders Lundström  
**Examiner:** Ann Lantz  
**CSC, KTH, 19/06/2017**

**IMMERSIV ÖGONSPÅRNINGSKALIBRERING I VIRTUELL VERKLIGHET MED HJÄLP AV INTERAKTIONER MED IN-GAME OBJEKT**

**REFERAT**
Denna avhandling ämnar att att undersöka en kalibreringsmetod för ögonspårning i virtuell verklighet där användarnas visuella uppmärksamhet från ögon och hand koordination används när man interagerar med objekt i den virtuella verkligheten. Detta kan möjliggöra kalibrering av ögonspårning utan att avbryta den virtuella upplevelsen, i jämförelse med traditionell kalibrering som är besvärlig, störande och kräver användarens fulla uppmärksamhet.


Resultaten visade att när deltagaren interagerar med objektet och det interativa objektet är stillstående, fick vi en fixering för omkring 60% av alla försök under godtycklig tidpunkt. När det interaktiva objektet rörde sig, visade resultaten en lägre procentandel. För att öka antalet, ska blickdata filtreras i stället för att använda rå blickdata för att undvika att flicker från ögonspåraren.

När det gäller faktorer som påverkar kalibreringsmetoden har valet av interaktion stor inverkan på metodens framgång, där interaktioner där det interativa objektet är stationärt har större potential. Dessutom påverkar interaktioner som tar längre tid och kräver precision för att slutföra interactionen kalibreringsmetodens potential positivt. Den omgivande virtuella miljön har också inflytande, eftersom en mer distraherande miljö kan negativt påverka kalibreringsmetoden.
Immersive Eye Tracking Calibration in Virtual Reality Using Interactions with In-game Objects

Ludwig Sidenmark
KTH, Royal Institute of Technology
Stockholm, Sweden
ludwigsi@kth.se

ABSTRACT
This thesis aims to investigate an eye tracking calibration method in virtual reality where users’ visual attention from eye-hand coordination is used when interacting with in-game items. This could potentially allow eye tracking calibration without interrupting the virtual experience, in comparison with traditional eye tracking calibration which is cumbersome, disruptive, and requires the user’s full attention.

A user study was conducted with 15 participants in which they were tasked to complete three different interactions. The interactions tested were a knob, a slider, and a liftable cube. Where in the virtual environment the participants were looking during the interactions was recorded and processed to enable comparison. The processed data was analysed to find factors that influenced the calibration method. Additionally, the results were analysed to find at what point during the interactions that has the most consistent eye tracking fixations on the interacted item, and therefore most potential for eye tracking calibration.

The results showed that when the participant is interacting with the item and the interacted item is stationary, we received a fixation for around 60% of all trials at any time. When the interacted item was moving the results indicated a lower percentage. To increase this number, the gaze data should be filtered instead of using raw gaze data in order to avoid flickering from the eye tracker.

Regarding factors that influence the calibration method, the choice of interaction has a big impact on the method’s success where interactions in which the interacted item is stationary has more potential. Additionally, interactions that take longer time and requires precision in order to complete the interaction positively influences the potential of the calibration method. The surrounding virtual environment also has an influence, as a more distracting environment can negatively impact the calibration method.

Author Keywords
Virtual Reality, Eye Tracking, Eye-hand Coordination, Calibration

INTRODUCTION
Virtual reality (VR) is a big area of research at this time and new technology is developed rapidly. Among these technologies is eye tracking within VR. However, due to the fact that every person’s eyes are shaped differently there is a need for calibration before using an eye tracker. Calibration for eye tracking devices is traditionally performed by users looking at a number of predefined positions on the field of view or on the desktop screen that the eye tracker will use for reference. This process has to be done prior to every use to mitigate factors such as light conditions and make-up and becomes time consuming and tedious.

Previous research has looked at making calibration more enjoyable by making it unobtrusive [12] or by gamifying the process [2]. Research has also investigated alternative calibration methods [8,10,15,16,17]. Recently, a popular approach has been to predict where users are looking within a scene by using visual saliency models and automatically calibrate the eye tracker [8,15]. This approach relies heavily on the accuracy of the saliency models which are not yet accurate enough for widespread use. Another popular alternative method for calibration is matching movements on the screen to the eye movements of the user [10,16,17]. However, this approach severely limits the design space as objects moving a similar direction would not work for calibration. Therefore, most commercial eye trackers still use the traditional calibration method due to its reliability. No one has, however, looked into if hand interactions done in a VR environment can be used for eye tracking calibration. Most modern VR devices (e.g., HTC Vive and Oculus Rift) uses hand tracking motion controllers that can be used to emulate hand interactions in-game. This is interesting to look into as the coordination between the user’s hands and eyes when interacting with an environment can potentially be used for immersive and implicit in-game eye tracking calibration.

This thesis investigates a calibration method where the user's hand interactions with in-game items in a virtual environment together with the visual attention related to these interactions are used for calibration. In the study, participants have to complete a set of interactions within a virtual environment with three different interactive elements; a cube, a knob and a slider. During the interactions the gaze fixations on the interacted items and their positions in the visual field is recorded. If the time
spent fixating on an item is consistent and the positions of the items are spread over the visual field then there is a potential to use these interactions for calibration.

An example of an interaction in the virtual environment is pressing a button. If you can safely assume that the user is looking at the button while pressing it you can compare the point of the button in the world space with the point that the eye tracking device is pointing at. If those points differ, then you could use that data to adjust the calibration to be more accurate. Additionally, making the user interact with a calibration item multiple times during a virtual experience will hinder the calibration to deteriorate due to factors such as the VR Head-mounted display (HMD) moving with respect to the head. The in-game method can also potentially be hidden inside the virtual environment, making the process implicit compared to the traditional calibration method’s explicit nature and thereby not disturbing immersion in the virtual world. This method has potential to give freedom for creators to make a fun, immersive and seamless calibration without the cumbersome and tedious calibration process.

Problem Statement
The thesis aims to investigate whether eye tracking calibration in VR using in-game objects and interactions is a feasible method. To answer this, the following research questions will be investigated: Can we consistently get fixations from eye-hand coordination on the interacted item in VR? What factors influence the calibration method?

BACKGROUND
All eyes are shaped differently and in order to be able to track them accurately calibration is needed to account for each user’s unique eyes and characteristics. The point-of-gaze (POG) is the point in space that is imaged on the center of the highest acuity region of the retina (fovea) of each eye. The POG can be used for many applications such as studies of attention and as an input modality in human-computer interfaces. The most common approach to estimating the POG is to via cameras estimate the centers of the pupil and one or more corneal reflections. The corneal reflections are virtual images of light sources (usually infrared) that illuminate the eye [3]. The POG is formally defined as the intersection of the visual axes of both eyes with the 3D-scene. The visual axis is the line connecting the center of the fovea with the center of the eye’s lens. The optic axis of the eye passes through the pupil center and the center of the corneal curvature. The optic axis is what is tracked by the eye trackers via reflections. Since in the human eye, the visual axis deviates from the optic axis [18] and it is the optical axis that is recorded by the eye trackers the visual axis needs to be reconstructed from the optic axis. Some of the parameters required in order to reconstruct the visual axis accurately are user-specific. This is why calibration is needed and used.

Calibration for eye trackers is traditionally performed by asking the subject to fixate their eyes at different predefined points for a specified time depending on the eye tracker. Usually, the number of calibration points is usually between 5-9. At each calibration point, the eye tracker decides when the eyes are fixating and directed towards the target. The eye tracker then detects a number of eye image features and associates their positions in the eye image with the position of the target. The decision to then accept a calibration target as successful is based on the assumption that the eye is more or less stationary over a minimum period of time when it is fixating the target [9].

Traditional calibration is associated with several challenges. Factors such as light conditions, the anatomy of the user’s eyes, glasses, makeup as well as the distance between the user and the eye tracker all affect the result of the calibration [6]. Since these factors may change between uses the calibration will deteriorate over time and need to be redone. It is preferable to calibrate at each use due to these factors. If the eye tracker is used for a longer period of time recalibration can sometimes also be necessary. Due to this the calibration procedure can easily become time-consuming and tedious [2]. Modern eye trackers use software to keep profiles for users that save their calibration. This works well for private use but not as well when the eye tracker is used by multiple people. Another challenge is the calibration process itself. Fixating on a point for an extended time is considered for many unnatural [1,16] since humans usually perform three to four fixations per second [10].

Research has been made in order to make the process less tedious. For instance, Flatta et al. made games out of the traditional calibration process and found that users experienced it to be more enjoyable and easier to use [2]. Renner et al. tried to make the calibration less obtrusive by having a dragonfly fly across the screen in a game setting and use that as a calibration stimuli in order to make the calibration process less immersion-breaking and found similar results [12]. Research has also been made in order to find alternative calibration methods. Pursuit calibration has been an alternative where the user follows a moving item on the screen and the positions of the eyes are mapped to the positions of the moving stimuli on the screen [10,16,17]. This method has the advantage that it is more natural for humans to follow moving targets rather than fixate on static targets. However, this method is limited by the fact that it only works if there is a single item that moves in a certain direction.

In VR, eye tracking has been recently introduced and it has some key differences in comparison to more traditional eye trackers which track the user's gaze on a regular computer screen. First, since the eye tracker is attached to the VR HMD it is not as affected by head movements as static
desktop eye trackers. The tracker might, however, be affected if the HMD moves with respect to the head while the user is moving around. Additionally, since the HMDs cover the eyes, external light conditions has minimal effect on the eye tracking. The eye tracker is also closer to the eyes compared with regular desktop eye trackers which lead to easier gaze tracking.

Modern VR devices (e.g. HTC Vive and Oculus Rift) also come with hand tracking motion controllers that can be used for interactions with the hands. Research has shown that the eyes are used for guiding the hands while interacting with the world [5,7,14]. The hand tracking motion controllers of modern VR devices are designed so that users can use their hand similarly as they would do in the real world. The user can interact via their hands in the virtual world via pushing, pulling picking objects up, etc. naturally just as in the real world. The visual attention related to these interactions in-game can potentially be used for eye tracking calibration in VR since you can assume that a user is looking at an item while interacting with it.

The method of calibrating using hand and eye coordination is interesting because there are many unknown aspects. There are some potential key differences between calibrating using hand and eye coordination which will from here on be called the in-game method and the traditional calibration method. The in-game method can potentially be hidden inside the virtual environment, making the process implicit instead of the traditional calibration method’s explicit nature. Another key difference is the fact that when doing a single interaction with an item you will rarely fixate at multiple positions in your field of view consistently. This indicates that the in-game method would be more suited for a continuous calibration. The in-game method will receive more calibration data as the users perform more interactions, in comparison to the traditional method where all calibration data is collected at the same time. The traditional calibration method is designed so that it is the only stimuli present in the stimuli area. This is different in comparison to the in-game method where multiple interactions may be present at the same time, depending on the in-game environment. Users may behave differently if there are different amount of possible interactions present.

When evaluating a calibration method there are two factors that should be taken into account: The time spent fixating on each calibration point and the angular difference between the gaze vector and the vector from the eye to the calibration point [6]. For a calibration to be successful the eye tracker needs a sufficient amount of time, depending on the sophistication of the eye-tracker used but generally less than one second to gather enough fixation data at each calibration point in order to accurately estimate all the needed parameters [6]. Depending on the amount of points and the eye tracker used the whole process takes about 5 seconds to complete. Additionally, it is important that the gaze is actually fixated at the calibration point since the eye tracker will use these points and the corresponding eye position for reference. The angular difference threshold is different depending on the application but usually, 1° is sufficient [6]. Another aspect that is interesting to look into is the spread of the calibration points on the user’s field of view. The calibration points should cover the area of stimuli in order to accurately track the eyes [6]. When there is no close reference point the eye tracker will interpolate between existing reference points. The better the spread of reference points the eye tracker has the more accurate these interpolations will be [6]. The main reason to why the traditional calibration is so widely used is because it can provide good results for these factors reliably among many users. Therefore when evaluating the in-game method, measuring these three will tell whether the in-game method has potential to be a working calibration method.

**METHOD**

To investigate eye-hand coordination when interacting with in-game items in VR a user study was conducted. The study was conducted with 15 participants (9 male and 6 female). 2 of the participants had extensive experience with VR while the rest had none or only tried it on occasion. An HTC Vive set with a built in 120 hz Tobii eye tracker with < 1° precision was used for the study. The participant first had to answer questions regarding factors such as eye sight and the participant’s dominant hand and eye. They also signed a consent form which informed them on the nature of the test and their right to withdraw. Since not all participants had previous experience with VR the participants were also allowed to get used to the experience by freely playing around in a virtual environment until they felt comfortable enough to start. This was done at most for 5 minutes. At the beginning of the test, the participant would perform a traditional eye tracking calibration. Each participant was then presented with interactions which they were tasked to complete.

A rapid prototyping stage was conducted in order to choose which type of interactions would have the most potential for the study. The interactions chosen were a draggable slider, a turnable knob and a cube that the participant can pick up (see Figure 1). The interactions chosen have in common that they require continuous input and attention during the whole interaction and precision in order to complete successfully. When interacting with a slider the participant was instructed to adjust the slider until the slider handle was glowing a different color. When interacting with a knob the participant was instructed to turn the knob until the knob was glowing in a different color. When interacting with the liftable cube the participant was tasked to pick up the cube and place it at a marked area. No additional feedback was presented for the cube. The participants were allowed to
freely move to their discretion in order to complete the interactions.

Figure 1. One combination of the three interactions used for the study.

The participant’s visual axis and the vector between the item and the participant’s eyes were recorded when interacting with the in-game items. The order in which the participant decided to interact with the different items was recorded. The screen of the participant's point of view was also recorded during the test. The participant was asked to think aloud during the test about the experience. After the participant had completed the test an interview was conducted with questions focusing on easiness, quickness, the joy of use and disturbance during the test.

Figure 2. A participant is completing a cube interaction by placing it on the drop zone. The blue line is a visualization of the participant’s gaze.

The study was arranged so that the participant was presented with a combination of the three interactions that they were instructed to complete. When all interactions were completed the participant would then be presented with a new combination of interactions. Each combination was different in the amount of interactions presented, ranging from 1 to 3 and the relative position of each interaction in relation to each other. There was at most one instance of each interaction type present in a single combination. In total the participant would interact with 15 different combinations. The order of the combinations was randomly selected. When grabbing an item the virtual model of the hand was turned invisible in order to not occlude the item, see Figure 2.

One way repeated measures ANOVAs was conducted to determine if there were any significant differences in total fixation duration, the number of fixations and duration of each interaction between the three interactions. Eventual extreme outliers were replaced with less extreme values. All distributions were assessed via Q-Q plot. The assumption of sphericity was assessed via Mauchly’s test of sphericity and if violated a Greenhouse-Geisser correction was then applied. Post hoc analysis to compare group differences was made with a Bonferroni adjustment.

In order to investigate the probability of fixations during an interaction all interactions were computed to a common time that had been normalized such that each interaction has been scaled to the median time of that interaction. Each interaction was initiated at time 0. The probability of getting a fixation at a specific time during the interaction was calculated by counting the percentage of the normalized interactions that had a fixation at that time (computed for 100ms bins).

RESULTS
During the initial rapid prototyping stage, several different types of interactions were tested by the experimenter. The results from this stage pointed towards two factors that are needed for an interaction to be suitable for calibration. The first factor is that the interaction should have continuous input, meaning that the interaction should require a longer continuous motion to complete. This leads to a longer interaction which gives the eye tracker more time to gather sufficient eye tracking data. The second factor is that the interaction should require precision in order to successfully complete. This could entail dragging a slider to or placing a cube at the correct position. This factor seemed to lead to the participant paying more attention towards the calibration target. The three interactions used in the experiment all have these two factors in common and was therefore chosen.

The interactions were separated into three different phases; reach, interaction and release. The reach phase starts when the participant’s gaze is targeted towards the interacted item and ends when the participant grabs the item. The interaction phase starts when the participant grabs the interacted item and ends when the participant releases the item. The release phase is defined to start when the participant releases the interacted item and ends when the participants gaze is no longer targeted towards the item.

Fixations were filtered after the test using the raw gaze data that was recorded during each interaction of the experiment. For this, an ID-T algorithm [13] was used with a duration threshold of 100ms and a dispersion threshold of 1° of the visual angle. Additionally for the fixation to be counted as
fixating towards the interacted item the fixation’s average point should be within 3° of the visual angle to the interacted item’s center. This angle was chosen through data testing.

**Observations of general participant behaviour**
During the first trials of the interaction, the participants appeared to be more careful as they were trying to figure out how to complete the interactions. After some trials they seemed to be more consistent in their interactions as they became experienced in how to complete the interactions. of the participants started experimenting and testing the boundaries by using, for example, both hands to complete multiple interactions at the same time. However users would still focus their attention on completing the interactions rather than playing around. Participants mainly used their dominant hand when completing the interactions. After several trials some participants started to switch hands more often. Participants would interact with the items at approximately an arm’s length. The participants would take steps toward the interactions in order to comfortably grab the interacted item. The cube had a mean distance between the participant’s eyes and the cube at 0.56 meters. The knob and slider had a mean distance of 0.66 meters and 0.63 meters respectively. No significant differences could be found between the participants based on previous VR experience.

**Probability of fixations**

![Figure 3](image_url)

**Figure 3.** Bar height indicates the probability of a fixation starting within 3° of the interacted item center during each interaction phase of an interaction.

Figure 3 shows, for each interaction, the probability of starting a fixation on the calibration target within 3° of the calibration target’s center during a trial. Probability in this context entails the percentage of all trials that fit the criteria. For all three interactions, the participant would fixate at the calibration target during the interaction phase for more than 95% of the trials. 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 an even lower percentage for all interactions ranging from 50% for the cube and knob to 40% for the slider.

All three interactions had different timings on when it was most likely a participant would fixate on the interacted item as shown in Figure 4. During the reach phase, all three interactions had a higher probability of fixating on the interacted item when closer to the interaction phase. The cube interaction had a generally lower probability during this phase, reaching 40% compared to the knob and slider who at the end of the phase reached 60%. This coincides with the results shown in Figure 3.

Figure 4 also shows that during the interaction phase all three interactions had largely different probability curves as can be explained by the fact that the fixations themselves are significantly different. The cube showed a large dip during the start of the interaction. During this part of the interaction phase, the participant’s attention would shift from the cube towards the cubes drop zone and the interacted item would also shift towards the drop zone. The probability of the participant fixating would then steadily increase as the cube got closer towards the drop zone reaching an 80% probability towards the end.

The knob’s interaction phase had a relatively stable fixation probability at 60% during the whole phase due to the fact that the knob is only rotating and not moving during the interaction. At the end of the interaction phase the probability falls, this coincides with when the knob has been rotated into the correct angle and is highlighted. At this point, the participants would often shift their attention away from the knob and release it.

During the slider’s interaction phase we can see a large dip at the start of the interaction. This coincides with general participant behavior where they would start by very quickly dragging the slider through its whole range, to find the correct position. When the general correct position of the slider had been found the participant would then more carefully drag the slider into the correct position, leading to an increase in the probability of fixating at the interacted item. At the end of the interaction, the probability would fall similarly to the knob when the slider had been dragged into the correct position and the participant would often shift their attention away from the slider.

All three interactions had generally similar curves during the release phase, gradually falling off until the participant had shifted their attention towards the next interaction.
before moving on. This can be highlighted by the cube’s initial spike in fixation probability as the participant would double check when releasing the cube that it had landed within the drop zone. This can be compared to the knob and slider where the participant would not release the interaction item until it was highlighted as completed. Note that the cube’s initial spike’s value is higher than the probability of starting a fixation during the release phase as shown in Figure 3. This is due to the fact that most of the fixations are started during the interaction phase and continued on to the release phase.

**Fixation parameters across interactions**

Some significant differences could be found when investigating the total gaze fixation duration. Results (Figure 5A) showed that both the knob (.571 ± .579 s, p < .0005) and slider (.461 ± .425 s, p < .0005) had a significantly higher total fixation duration than the cube (.188 ± .209) during the reach phase. During the interaction phase the knob (1.153 ± .751) had a significantly higher total fixation duration compared to both the cube (0.981 ± .534 s, p = .046) and slider (0.647 ± .474 s, p < .0005). The cube had a significantly higher total fixation duration than the slider (p < .0005). During the release phase the slider’s total fixation duration (0.080 ± .123) was significantly lower than the cube (.164 ± .223 s, p < .0005) and knob (.174 ± .272 s, p < .0005).

When comparing the number of fixations for each phase (Figure 5B) some similar results was found. During the reach phase the knob (2.935 ± 2.584, p < .0005) and slider (2.532 ± 2.261, p < .0005) had significantly higher number of fixations compared to the cube (1.218 ± 1.131). During the interaction phase, the knob (5.635 ± 3.510) had a significantly higher number than both the cube (4.428 ± 2.368, p = .002) and slider (4.055 ± 2.700, p < .0005). The release phase showed that the knob (1.111 ± 1.404) had a significantly higher number than the slider (.516 ± 0.807, p < .0005). No significant difference could be found for the cube (.778 ± 1.065).

When comparing the durations of each phase of the interactions (Figure 5C) some significant differences could be found. During the reach phase both the knob (1.191 s ± .750, p < .0005) and slider (.967 ± .626 s, p < .0005) had significantly longer phase duration compared to the cube (.578 ±.346 s). This compared to the similar results from total fixation duration and the number of fixations indicates interacted item placement affects these three factors. During the interaction phase both cube (1.706 ± .666 s, p < .0005) and knob (1.858 ± .924 s, p < .0005) had substantially longer phase durations than the slider (1.150 ± .652 s). During the release phase the knob (.590 ± .563 s) had a significantly longer phase duration than the cube (.418 ± .320 s, p = .007) and slider (.330 ± .277 s, p < .0005).

![Figure 4. Probability of fixation within 3° of the interacted item center (computed for 100ms bins) during each respective interaction as a function of time.](image-url)
When investigating where on the interacted items the participant would fixate, it was found, as shown in Figure 6, that there was a skewness in the y-axis, meaning that there was a tendency of the fixations to be 1° above the interacted item center. In the x-axis, however, most fixations were positioned at the center of the interacted item. No significant differences could be found between the different interactions nor the different phases.

Fixations in field of view

Results indicated that the choice of hand to interact with the item impacts where on participant’s the field of view the fixations occur. When using the left hand the fixations occurred 86% on the left side of the field of view and 14% on the right side. When using the right hand the fixations occurred 79% on the right side and 21% on the right side. No significant differences could be found between the three interactions nor the different phases.

Interview results

When the participants were asked what they felt about the experiment and the interactions the general consensus among all participants was that the interactions felt natural and easy. P4 stated, “The interactions felt natural and were easy to understand, it felt intuitive on how you should interact with the items”. P13 commented on the cube interaction that “When dropping the cube it felt uneasy as you lost control of it when you release and would, therefore, pay more attention towards it”.

When asked how it felt as the participant had made multiple trials of the interactions the participants stated that they got more comfortable after more trials. P1 stated, “During the first couple of trials you were more careful and tried to figure out what to do”. Some participants also stated that they started to experiment more with the interactions when they got more comfortable with them. P6 stated, “I would
try and complete two interaction at the same time when I was more comfortable with the interactions”.

When asked if they felt if they behaved any differently if presented with multiple interactions compared to only one there were mixed responses. Some participants didn’t feel like there was any difference. Some participants stated however that they were acting differently. P13 stated, “I switched hand more regularly when there were multiple interactions”. Some participants commented that they would try to complete multiple interactions at the same time if they were able to.

**DISCUSSION**

In this study I have investigated eye-hand coordination for three different interactions that require continuous input and precision, and compared them in the context of eye tracking calibration. Results from the study indicate that even though they have these two factors in common the interactions are still widely different when comparing user's visual attention between the interactions and that these two factors alone are not enough to determine whether an interaction is suitable for calibration purposes.

**Fixations during interaction phases**

There are some common occurrences when comparing the three interaction phases. First, the knob has a stable fixation probability during the whole interaction phase. This is likely due to the fact that the knob is stationary during the interaction and only rotates around its own axis. Secondly, the cube’s interaction phase has a significantly higher fixation probability towards the end of the interaction phase when the interacted item is the drop zone, which is stationary. Thirdly, the slider is similar in that it has a higher fixation probability during its later stages when the user is closer to the correct position and is dragging the slider slower in order to fine tune it to the correct position. This indicates that the user is more likely to fixate on the target when more precision is required. Additionally, it points towards that the participant’s intentions needs to be taken into account when considering an interaction for calibration purposes. In general, people tend to be less accurate and focused at the start while at the end of the interaction, when more precision is needed, they are more focused. The three interactions together indicate that there is a higher likelihood of getting a fixation when the interacted item is stationary during the interaction. In other words this indicates that interactions that have a stationary interaction phase is more suitable for eye tracking calibration compared to interactions that do not.

During the end of the interaction phase the probability of a fixation increases for the cube while it falls for both the knob and slider. This may be due to the fact that the knob and slider provided feedback when in the correct position, while the participant is still grabbing onto the interacted item. For the cube however, there was no direct feedback. This may indicate that having more disruptive interactions with less feedback could be better from a calibration perspective as users potentially would pay more attention in order to know that they have completed an interaction. In this study the feedback was shown on the interacted item itself so that the participants would keep their attention towards the item during the whole interaction. It would be interesting to see how participants would react if this feedback was shown elsewhere. For example on a lamp next to the interacted item.

During the reach phase, the cube had a lower probability of the participant’s fixating on the interacted item as well as a significantly lower total fixation duration, number of fixations, and phase duration, than both the knob and the slider. The cube was placed on a shelf underneath its drop zone while the knob and slider had similar height in the virtual environment (Figure 1). This meant that the cube was closer to the participant's hands when the hands were in in their rest position and the participants did not have to reach as far in order to grab it. The knob and slider which were placed at a similar height did not have any significant differences. This indicates that placement of the interacted item is important in the reach phase in order to maximize the probability of fixations.

The release phase had a lower probability for all interactions of starting a fixation, total fixation duration, the number of fixations and phase duration, compared to the reach and interaction phases. In this phase it was common that users moved their visual attention away from the interacted item before releasing it, leading to a missing release phase due to no visual attention towards the item from the participant. For the cube, most fixations that can be seen in the spike in Figure 4 are fixations that started in the interaction phase and was carried over to the release phase as shown in Figure 3. In summary, these results indicate that the release phase is difficult to use for calibration purposes compared to the other two phases.

For all three phases, the interaction with the longest phase duration also had the longest total fixation duration and the interaction with the shortest phase duration had the shortest total fixation duration. In a calibration context, this would mean that it is important to make sure that the interaction duration is sufficiently long. This hints that it may be beneficial to use interactions that require continuous input as they generally take longer to complete compared to interactions that do not require continuous input. All three interactions had a higher probability to fixate on the target (Figure 3), longer total fixation duration, the number of fixations and phase duration (Figure 5) during the interaction phase compared to their reach phase and release phase counterparts. This indicates that it is most beneficial to conduct an eye tracking calibration during the interaction.
Fixation positions

Results showed that instead of the participants fixating directly towards the center of the interacted item there is instead a 1° skew in the y-axis which means that the participants were fixating slightly above the item’s center (Figure 6). A possible explanation for this may be due to the placement of the interacted items. The items were placed at a height of 1.3 meters in the virtual world. This was made so that the participants could comfortably reach for and interact with the items without much strain. This, however, leads to that the participants will look down on the interacted items instead of looking directly at it. These results indicate that it is important to consider the position of the interacted item in the y-axis in order to choose an accurate reference point as possible for eye tracking calibration. However, this aspect needs to be investigated further in order to be certain.

An unexpected find was the effect of which hand used when interacting with an item on the fixation position on the participant’s field of view. It is, however, reasonable as it is more effortless and natural to shift your eyes and head together towards the interacted item and therefore the fixation position is on one side of the field of view instead of only shifting the head fully towards the interacted item. These results indicate the need to ensure that the calibration points are equally distributed on the left and right side of the field of view to provide accurate eye tracking on both sides by forcing users to alternate hands when doing the calibration interaction.

User behaviour

Participants thought that the interactions felt natural and easy to do, this is important as one of the main theoretical advantages of the in-game calibration method is that it is implicit and not disturbing. Participants expressed that during the first trials of each interaction they were more careful and would take longer to complete the interaction because they were not completely confident on how they should do it. These comments coincide with the quantitative data were many of the outliers were the first trials of an interaction. From a calibration perspective, this may indicate that the first time doing an interaction is best suited for calibration as it gives the eye tracker more time to collect sufficient gaze data as hinted towards in the results from Figure 5.

A potential pitfall in using interactions as a calibration method is how to ensure that the user is paying attention towards the interaction that they are doing. The design of the interaction itself is one aspect of it but there are also other factors to consider. Some participants in the study would try and complete two different interactions at the same time if able, using one hand for each interaction. This lead to a participant behavior where they would constantly switch their visual attention between the two interactions making the likelihood of fixations less. In order to avoid this behavior, an interaction that is used for calibration should not be in the direct vicinity of another possible interaction and the surrounding environment should not be distracting the user when doing the interaction. This way the user should keep the focus on the task at hand.

Using visual tricks in order to exploit human’s automatic reactions on events could potentially be used in order to increase the probability of a fixation. For example, if dragging a slider and it suddenly blink, the user would instinctively fixate at the slider and it could be suitable to sample the data right after the slider blinking. Additionally, making sure the interacted item sticks out in the virtual environment through color contrasts or other mediums may helps keeping users’ attention on the interacted item. In this study, for example, the cube was chosen to be yellow so that it would not blend in with the background or with the drop zone. Additionally, in this study, the participants had as much time as they wanted and were put under no stress or time pressure. In a game setting where users may be under time pressure it is entirely possible that the user’s visual attention may shift more often in order to plan ahead or look out for enemies and it would therefore be more important to use these visual tricks to keep the attention towards the interacted item.

In this study the participant’s hands turned invisible when grabbing an item so that the hand would not occlude the item. During the initial prototyping phase, both versions were tested. When not turning the hand invisible people would often look at the part of the item which was not occluded by the hand. This made it so the visual attention was entirely dependent on how the user grabbed the item, making it harder to predict where on the item the user will look at it. Having the hands turn invisible made the visual attention more predictable and was therefore chosen for the study.

The results from investigating the mean distance between the eyes and the interacted item during and interaction show that participants will interact with an item somewhere between 0.6 to 0.8 meters. This information may be useful when choosing the size of the interacted item for calibration as the best case scenario is to have the visual size of the interacted item to be 1° of the visual angle.

Data reliability

Note that while each interaction phase had a very high probability of getting a fixation during the entire phase (Figure 3) the probability of getting a fixation at a specific time was not higher than 60% for both the knob and slider and at 80% for the cube. These numbers can be considered low, especially if you compare with the results from [7], which had made a similar experiment by investigating eye-hand coordination during interactions and reported
much higher percentages. It is worth questioning whether these results points towards the fact that the in-game method would be unreliable and would therefore not be suitable for calibration at all. There are some possible explanations for this results. One may be the definition of a fixation which may be considered harsh. In this report, a fixation is defined as being a minimum of 100ms long and the gaze points not being further than 1° of the visual angle away from each other. Compare this with [7] where their definition of a fixation is the period between two jumps of the eyes. Additionally, for this study the unfiltered gaze data was used and during the experiment some flickering from the eye tracker was found. If the gaze data was smoothed in order to combat the flickering then it could lead to better results in terms of fixations.

Another important difference is the simple fact that this study was done in VR while [7] was not and it is unclear whether this difference leads to a different participant behavior. Another consequence of this study being conducted in VR is the method of tracking the participant’s gaze data. The HTC Vive have lenses which are rounded, making it difficult to track gaze as traditionally done with desktop eye trackers by getting an x and y position on the stimuli space. Therefore, in this study, the gaze data was recorded via directional vectors inside the virtual world instead of the directional vector from the eyes to the screen. Due to this difference in data collection it is difficult to determine whether the results are possible to compare.

It is important to note that a calibration does not necessarily require the gaze data to be a fixation with a dispersion threshold of 1° of the visual angle and could possibly make a sufficiently accurate calibration with less strict requirements.

Another possible application of the in-game calibration method is to use it to compensate for the traditional calibrations weaknesses, which are breaking of immersion and possible drift during a virtual experience instead of completely replacing it. A possible scenario can be at a start of a virtual experience to calibrate with the traditional calibration to the get the important initial calibration. The in-game method could then be used to continuously check and adjust the eye tracker's calibration during the virtual experience in order to mitigate drifting and other factors. In this scenario, it is not vital for the in-game method to be as reliable and immersion will not be disturbed during the virtual experience. In this scenario a 60% chance of getting a valid calibration can be enough, while it may be a bit low if the eye tracker is completely uncalibrated beforehand.

**Future research**

A key benefit of using interactions for eye tracking calibration compared to the traditional calibration method is that the interactions do not necessarily break the immersion of the virtual environment. Therefore it is important to investigate the use of interactions in more immersive virtual environments and more intense scenarios such as in games. In this study, the environment was constructed so that the interactions themselves were the most interesting aspect of the environment. Additionally, the participants were instructed to complete the interactions. A more natural case would be having the users do as they please. Therefore it would be interesting to see if the user’s visual attention would be similar if they were given more freedom.

This study focused on interactions that required continuous input and precision to complete. These two factors were chosen through testing by the experimenter and for the sake of the in-game calibration method. It would be interesting to see how other types of interactions would perform when placed in front of a larger pool of users in order to test the limits of the in-game method. The next step of this work would be to implement and test an in-game eye tracking calibration algorithm whose calibration results could be compared with the traditional calibration method.

**CONCLUSION**

The aim of this study was to investigate the feasibility of using eye-hand coordination as a method for eye tracking calibration. In summary, regarding the calibration method’s feasibility the method has great potential, in particular for stationary interacted items during the interaction phase. However, the method is not entirely feasible in the way it was studied and tested, because of a relatively low consistency in the fixations at any specific time. The feasibility of the method is dependent on how much fixation data is required by the eye tracker to successfully calibrate, how critical it is to get a successful calibration at that time during the experience, how distracting the surrounding environment is and the design of the interaction itself.

The results showed that when the participant is interacting with the item and the interacted item is stationary, we get a fixation for 60% of all trials at any specific time. When the interacted item was moving the results indicated a lower percentage. In order to increase this number the gaze data should be filtered instead of using the raw gaze data in order to avoid flickering. Also avoid distracting elements in the virtual environment such as additional interactions that users can perform at the same time to improve the potential of the calibration method.

Regarding factors that influence the calibration method, the choice of interaction has a large impact on the methods success where interactions in which the interacted item is stationary has more potential. Additionally interactions that take longer time and requires precision in order to complete the interaction positively influences the potential of the calibration method. The surrounding virtual environment also has an influence, as a more distracting environment with for example multiple intractable items can negatively impact the calibration method.
ACKNOWLEDGEMENTS
The author would like to thank the supervisors of this thesis Anders Lundström from KTH, Ralf Biedert and Geoffrey Cooper from Tobii for their help and motivation as well as the company Tobii for supplying help and equipment for the thesis.

REFERENCES