Pursuit Calibration: Making Gaze Calibration Less Tedious and More Flexible

Ken Pfeuffer
Lancaster University
Lancaster, United Kingdom
k.pfeuffer@lancaster.ac.uk

Mélodie Vidal
Lancaster University
Lancaster, United Kingdom
m.vidal@lancaster.ac.uk

Jayson Turner
Lancaster University
Lancaster, United Kingdom
j.turner@lancaster.ac.uk

Andreas Bulling
Max Planck Institute for Informatics
Saarbrücken, Germany
andreas.bulling@acm.org

Hans Gellersen
Lancaster University
Lancaster, United Kingdom
hwg@comp.lancs.ac.uk

ABSTRACT
Eye gaze is a compelling interaction modality but requires user calibration before interaction can commence. State of the art procedures require the user to fixate on a succession of calibration markers, a task that is often experienced as difficult and tedious. We present *pursuit calibration*, a novel approach that, unlike existing methods, is able to detect the user's attention to a calibration target. This is achieved by using moving targets, and correlation of eye movement and target trajectory, implicitly exploiting smooth pursuit eye movement. Data for calibration is then only sampled when the user is attending to the target. Because of its ability to detect user attention, pursuit calibration can be performed implicitly, which enables more flexible designs of the calibration task. We demonstrate this in application examples and user studies, and show that pursuit calibration is tolerant to interruption, can blend naturally with applications and is able to calibrate users without their awareness.

Author Keywords
Eye gaze calibration; Gaze interaction; Gaze interfaces; Eye tracking; Smooth pursuit eye movement.

ACM Classification Keywords
H.5.2. Information interfaces and presentation: User Interfaces

Figure 1: *Pursuit calibration* is based on moving targets (a), tolerates interruptions (b), can blend with application tasks (c) and calibrate users even when they are not aware (d).

INTRODUCTION
Eye gaze is long established for human-computer interaction [29, 7, 8]. As input device, eye gaze has been shown to be fast and highly accurate while requiring no acquisition time [20]. Our eyes are always at the ready and naturally involved in attentive processes and selection of what we are interested in [24]. We can use our eyes to interact over a distance [23], and in combination with other modalities [22, 31]. However, although pervasive in usability labs, eye gaze is still rarely used in everyday interaction with computers. This is attributed to the calibration problem, as users are required to register the input space of their eye movement with the output space of a display before they can start to interact.

State of the art eye tracking systems are based on remote gaze estimation using the pupil center and corneal reflections [5]. The movement of the pupil relative to reflections on the eye can be accurately tracked with a single camera, but a user calibration is necessary to register estimated gaze direction with the target interaction space. The registration requires sampling of eye gaze at known points, for which the state of the art procedure is to present the user with a succession of targets to fixate. Theoretically fewer points would suffice to establish the input-output mapping, but in practice five or more points are used, spread over the target space. Unfortu-

UIST'13, October 8–11, 2013, St. Andrews, United Kingdom.

http://dx.doi.org/10.1145/2501988.2501998

nately, this is a process that presents users with a task that is of poor usability [14], often experienced as difficult [19], and frequently described as tedious [28, 16, 33].

We address the calibration problem by introducing *pursuit calibration* as a method that overcomes the rigidity of existing procedures. The method is based on the smooth pursuit eye movement that we naturally perform when we follow a moving object with our eyes [18]. Recent work demonstrated that smooth pursuits can be used to identify moving objects a user is looking at, by correlating the user's eye movement with the trajectory of the objects [26, 25]. We adapt this approach by basing calibration on presentation of a moving target to the user and continuous sampling of calibration points while the user follows the target. Because our method is based on correlation of eye and target movement, we can reliably determine when a user is attending to the target. Automatic detection of user attention to a known point is a fundamental advance over existing procedures, as it enables more flexible calibration task designs – there is no rigid start and end, calibration can stop and resume, and users need not be explicitly aware of the task.

Figure 1 illustrates pursuit calibration, and some of the advances the method provides over existing procedures. Instead of having to dwell on static points, users follow a moving calibration target (Figure 1a) during the whole procedure. The method is intelligent in collecting gaze data only when the user actually attends, and this can be done iteratively. If a user is interrupted during the calibration, they can simply resume and do not have to start all over again (Figure 1b). There is no need to make start or end of the calibration explicit, as calibration data can be automatically collected as soon as a user's attention is detected, and for as long as deemed necessary for a given application purpose. The calibration can thus be implicit, and blended into an application experience. Figure 1c illustrates the creative possibilities of leading from interaction with moving objects (here, shooting stars) smoothly to interaction with static content (star constellations). As calibration is implicit, it is not necessary that the user actively cooperates, and interfaces can be designed to calibrate users unawares. Figure 1d illustrates a waiting screen that provides an illusion of idle waiting while the system actually calibrates the user.

Our work makes four contributions. First, we introduce pursuit calibration and the concept of using correlation of eye and target movement for detecting user attention. Second, we provide a performance characterisation of the method against a baseline of state of the art static point calibration that shows that it is possible to achieve the same level of gaze estimation accuracy in potentially less time. Third, we demonstrate the method in three application examples that illustrate the flexibility gained for designing the calibration task. Fourth, we show that the pursuit calibration method is interruption-tolerant, and able to calibrate users without their awareness.

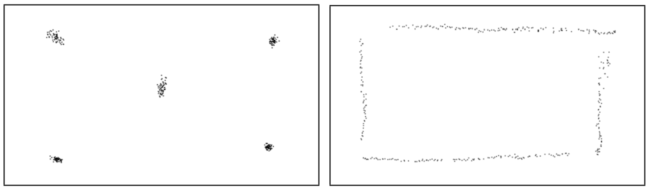
RELATED WORK

The requirement for user calibration is well recognised as a critical problem for wider adoption of eye gaze. Morimoto and Mimica, in their 2005 survey, concluded "although it has been thought for long that [eye gaze tracking devices] have the potential to become important computer input devices as well, the technology still lacks important usability requirements that hinders its applicability", pointing specifically at calibration as a task of poor usability [14]. It is not natural for users to dwell on static calibration markers, and the protocol of following through a succession of markers presented on the screen is dull and tiring for the eye. Schnipke and Todd's account of "trials and tribulations of using an eye-tracking system" [19] might be dismissed as dated but there is hardly any publication of eye-tracked user studies that does not report calibration problems. Villanueva et al., and others in eye tracking research, have described the calibration task as "most tedious" [28].

Consequently, calibration is a significant focus in eye tracking research. A common thrust is to develop advanced geometric models to reduce the number of calibration points that need to be sampled to achieve good accuracy (a generally agreed accuracy target is 1° of visual angle), for example to only two points [28, 16]. Commercial suppliers advertise "child-friendly versions" where only two points are used, and researchers have proposed single-point multi-glint calibration also with infants in mind [6]. However, although a single point with multiple reflections is theoretically sufficient, good accuracy generally requires additional points [27]. Pursuit calibration contrasts the quest for fewer calibration points, as it involves continuous collection of point samples while the user follows a moving target.

A range of work are aimed at using the eyes for interaction while avoiding calibration altogether. Eye gestures use relative eye movements to encode commands for computer input and thus do not require any registration with the output space [2]. Lightweight computer vision techniques have been used to detect attention to displays [21] and to classify whether users look straight at a display or to the left or right [32]. *Pursuits* correlate relative eye movements with the movement of objects on a display, and thus enable gaze selection of moving objects without estimating gaze direction [26, 25]. All these efforts have in common that they compromise expressiveness, whereas pursuit calibration is aimed at accurate registration of eye gaze with the output space. Other approaches aim to achieve good accuracy without calibration but rely on additional instrumentation [13], or longitudinal training for probabilistic gaze estimation [1]. Pursuit calibration, in contrast, does not avoid calibration but is aimed to make the process less tedious and more flexible.

Calibration Games [4] share our motivation to make calibration a more natural task. The system transforms the calibration into a game to make it more pleasant. Users perceived the calibration as more enjoyable while calibration data quality was not compromised by the game design. However they examine calibration in general while we focus on gaze calibration which they did not specifically address. Renner et al. [17] demonstrated playful integration of gaze calibration in a virtual reality environment. A dragonfly flew into the view of the user every time a calibration was required and directed the user from point to point. To ensure that the user



(a) Point clouds of a 5-point calibration. (b) Samples of a moving-target calibration.

Figure 2: Samples of a point-to-point and a moving-target calibration. Contrary to a point-to-point method, a moving-target calibration cannot average each point, however the samples are more equally distributed across the screen.

truly follows this target, the gaze estimation of the previous calibration was regarded. In a comparison to a standard point-to-point calibration, users found the guided approach more enjoyable.

A few papers present moving targets for calibration. Kang and Malpeli [9] calibrate their system to cats with moving targets, because fixations on stationary targets are often unreliable. Here, they use horizontally moving targets to calibrate the vertical eye position and vice versa. This resulted in an accurate calibration and emphasized the advantage of not relying on successful fixations. They also point out that this method is applicable to any species and any recording system. Kondou and Ebisawa [10] evaluated a gaze detection system with user head movements during calibration. Two point-to-point and one moving-target calibration methods were compared. The calibration with moving targets yielded the highest accuracy. These efforts show that calibration with moving targets is feasible and producing good results. However, our approach introduces automatic detection of user attention as an additional feature, achieved by correlation of eye and target movement.

PURSUIT CALIBRATION

The goal of a gaze calibration method is to collect samples of eye gaze at known points in the output space, in order to calculate an accurate gaze-to-display mapping. The pursuit calibration method approaches sampling in a novel way, by displaying a smoothly moving target to the user. The method automatically detects when the user is following the target with their gaze, by correlating the eye movement trajectory with the trajectory of the target. Only points on the display through which the target passes while the user is looking at it, are collected as samples for calibration.

Figure 2 illustrates how sampling of gaze points with pursuit calibration fundamentally differs from sampling with a conventional calibration procedure. A point-to-point calibration collects multiple gaze samples at each of a small number of points, typically at the corners and the centre of the display (Figure 2a). In contrast, pursuit calibration collects one sample for each display point at which the user followed the target. Since the target moves continuously, a larger number of display points can be sampled but each only once (see Figure 2b).

Method implementation

Pursuit calibration builds on recent work in which eye movement was correlated with the movement of screen objects, in order to determine which of a number of objects a user is looking at [25]. As in that work, our method takes uncalibrated gaze coordinates from the eye tracker as input, in the form of a synchronised time series. A moving window is used over which the gaze coordinates are compared with the coordinates of the moving target. The correlation is computed by using Pearson's product-moment correlation for the horizontal and the vertical gaze position. If the returned correlation coefficient lies above a defined threshold, the user must have followed the trajectory of the moving target with their eyes, and eye gaze is sampled at the current position of the target for later calibration. Both the size of the moving window and the threshold for correlation are controlled by the designer, and for this work have been based on Vidal *et al.*'s analysis of the Pursuits method [25].

Once sampling has been completed, the collected gaze points are fitted to a homography model, representing a perspective projection of gaze and screen plane. This is implemented with the homography computation of OpenCV, using RANSAC for outlier elimination with default parameter settings [3].

Design considerations

Attention detection. Automatic detection of user attention is the distinguishing feature of pursuit calibration. The design choices influencing attention detection are the length of time window and the threshold set for correlation of gaze and target movement. The confidence that a user is following a moving target increases with evidence over a longer time window, and with a higher threshold, but as a trade off will effect the sampling rate as more samples will fail the correlation test.

Path. A moving target for pursuit calibration can follow any path across the display. A main design consideration is to sample the target interaction space in a representative manner to obtain a well-balanced mapping. A pilot study in which we investigated the influence of the target's path indicated that the most accurate results are obtained when a path along the borders of the display is followed (see Figure 3). This is similar to a point-to-point calibration where the static points are often near the screen's edge.

Sample size. The more samples are collected, the more accurate the calibration will be, but also the more time consuming. The calibration task can be designed for calibration in a fixed amount of time, during which as many samples as possible are collected. Alternatively, target movement could continue until a desired number of samples have been collected.

Target velocity. Because the target is moving, its velocity is relevant as well. The velocity of a human smooth pursuit eye movements is limited to up to $30°/s$ [30], and greater speeds lead to catch-up eye movements. However, the detection rate for smooth pursuits also decreases when targets are too slow [26]. The speed of the target also affects the duration of the calibration: if the target moves slowly, it will take longer to sample the output space.

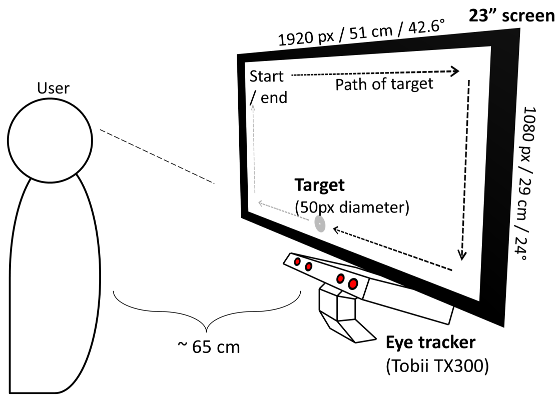


Figure 3: Moving target calibration setup during the accuracy evaluation.

ACCURACY EVALUATION

From a usability point of view, the most important factors for a successful calibration are its accuracy and how time-consuming it is. We therefore investigate the accuracy of pursuit calibration depending on its duration and target velocity. In order to make sense of the results, we compare it to a baseline point-to-point calibration procedure. We use the Tobii 5-points calibration to represent a standard of calibration for the eye-tracking industry. It takes 19 seconds overall, during which the user has to fixate on four points at screen corners and one at the center, and involves a moving target to guide the user from one point to another.

The calibration stimulus we evaluate is that of a target that moves along the borders of the screen in a rectangle (see Figure 3). We test two types of design: one where the speed of the moving target is constant, and one where it accelerates on straight lines and slows down when approaching a corner, for a smoother transition. The shorter the duration of the stimulus, the quicker the target travels. We conducted a pilot study to determine which range of speeds is feasible: four users performed each calibration, calibration duration was varied, speed varied with respect to duration and path of target remained rectangular. Data analysis resulted in the following ranges of durations/speeds for our study evaluation: for the constant speed case durations of $\{5, 10, 15, 20, 25, 30\}$ seconds, which in visual angles equates to a target speed of $\{23.3, 11.6, 7.6, 5.8, 4.7, 3.9\}°/s$, and for the accelerated speed durations of $\{5, 7.5, 10, 12.5, 15, 17.5\}$ seconds (target speed of $\{23.3, 15.5, 11.6, 9.3, 7.8, 6.6\}°/s$). Accelerated speed has lower durations because calibration was satisfying with shorter durations during the pilot study. The acceleration reached a peak speed of $35.77°/s$ and slowed down to $1.1°/s$ at the corners.

Participants performed 13 calibrations: first the Tobii baseline, then the twelve other in randomised order. In order to calculate the accuracy of each stimulus, after each calibration procedure we presented a test where participants had to fixate 16 equally distributed, static points appearing in random order, for two seconds each. Calibrated gaze data was collected for each of these points in order to calculate the accuracy by evaluating the difference between the mapped gaze and the real point position. During analysis, the first 0.8s and the last 0.2s were discarded as to compensate for the time participants' eyes need to travel to reach the target and for anticipation movements. After each task the user was instructed to a pause of at least 90 seconds to avoid eye fatigue effects. In total, the study lasted approximately 35 minutes.

15 paid participants between 21 and 32 years ($\mu = 25.6$, $\sigma = 3.97$, 4 female) took part in the study. Participants sat in front of a display coupled with a Tobii X300 remote eye tracker in a dimmed room (see Figure 3). Data was collected at 60Hz.

Results

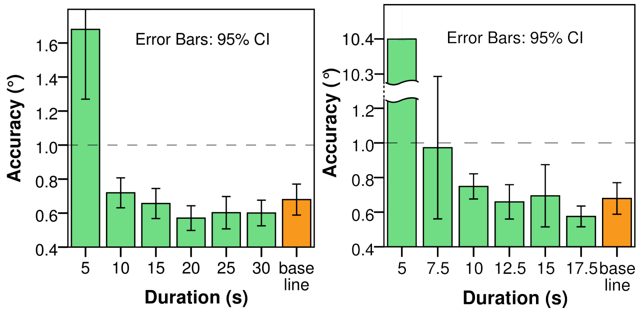
Figure 4 shows the measured difference between the true position of points and the gaze collected, averaged over all 16 points. For the constant speed calibration (Figure 4a), all durations ≥ 10 seconds show marginal differences and achieved an accuracy under $1°$. The five second calibration procedure showed an accuracy 2.15 times higher than the one that lasted ten seconds, indicating that a target speed of $23.3°/s$ is less feasible when users calibrate with our rectangular path design.

Results for the accelerating target calibration (Figure 4b) confirm this observation, the five second procedure also being the worst with 9.96 times higher difference than the 7.5 seconds procedure. Here again, from duration $10s$ and above, tests show an accuracy of under $1°$. The baseline 5-points calibration resulted, as expected from the industry standard, also in less than a $1°$ error.

In order to evaluate how well the correlation method detected the participants' gaze, we plotted its detection rate in Figure 5. For both target movement types, more than half of the gaze data was detected by the correlation algorithm at ten seconds and above. The five seconds procedure yielded the lowest detection, which indicates that the target was moving too quickly to ensure robust detection. Undetected correlation is caused by too fast target movement, blinks of participants as well as sharp direction changes at the corners. At corners the user was often required to perform catch-up eye movements on the target, causing the eye and target correlation to be lost for a short time as shown in Figure 2b.

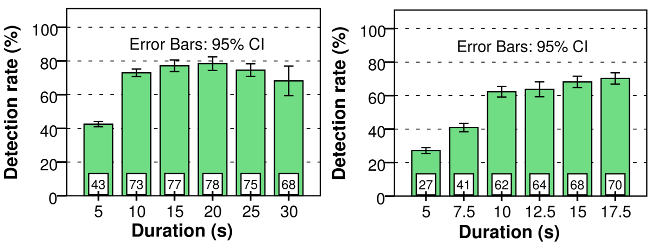
We also conducted a post-hoc analysis of how accuracy iteratively improves as the target travels across the screen. For the analysis we calculated the accuracy for each 100 ms, i.e. fitting all samples available so far into a mapping. Figure 6 represents this progressive accuracy for each constant speed procedure. It shows that the accuracy was not computable before at least 40% of its path was completed, which corresponds to the target being halfway between the top right and bottom right corner of the screen. This indicates the necessity to collect enough samples on both vertical and horizontal axis of the screen. This is rational, since samples on only one axis cannot produce an accurate mapping over two dimensions.

Shortly after the vertical axis samples are collected, accuracy drastically improves and stabilizes. The final accuracy is reached around 66% of its whole path, which corresponds to a location slightly after having moved over the bottom right



(a) Constant target calibration. (b) Accelerated target calibration.

Figure 4: Calibration accuracy results (in degree of visual angle) for each movement type and each duration.



(a) Constant target calibration. (b) Accelerated target calibration.

Figure 5: Correlation detection rate: the ratio of detected gaze samples to overall possible samples in a calibration procedure.

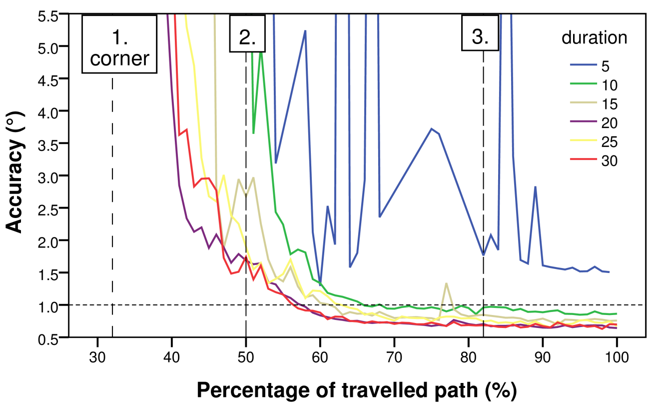


Figure 6: Analysis on how accuracy incrementally improves over time, for procedures with constant speed. Rectangles depict when the target encounters corners of the target path.

corner of the screen. This shows that the calibration procedure could have stopped at this point and thus significantly shortened the procedure duration, e.g. the ten second procedure could have already stopped at seven seconds.

These results show that a calibration with moving targets is a feasible alternative to a point-to-point calibration. Pursuit calibration yields an accuracy as high as the industrial baseline for the same time or even shorter durations, as short as seven seconds. Constant and accelerated target movement showed minor differences in accuracy. The study also showed that

	ATM	**Stargazing**	**Waiting Screen**
Settings			
Participants count (calibrated / all)	20/20 (12F)	30/37 (5F)	13/14 (5F)
Study duration (min)	5-8	3-5	3-5
Target speed (deg/s)	6.6	∼17.7	6.6
Target design	Simple circle	Shooting star	Text
Correlation threshold	0.3	0.7	0.3
Window size (ms)	160	80	80
Sample size threshold	200	100	200
Threshold per target	200 (1 target)	25 (4)	50 (4)
Results			
Calibration time(s)	7.8,12.4, 10.8	30.9(median)	13.6
Calibration time SD (s)	2.1, 3.8, 3.9	28	5.3

Table 1: Settings and results of the three application evaluations. For the ATM case, results indicate the calibration time for no interruption, on-screen interruption and off-screen interruption, consecutively. The calibration times presented for the on-screen and off-screen distractions exclude the time during which users were looking at the distractions.

the target does not need to travel across the entire screen, but rather needs to cover the full range of the two dimensions of the display.

APPLICATIONS

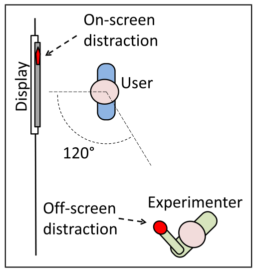
We designed three realistic applications that each emphasize different capabilities of pursuit calibration: the Automated Teller Machine (ATM) application demonstrates the interruption-proof capability of the method, the stargazing scenario shows the smooth blend between calibration and interaction, and in the waiting screen study users are calibrated without being conscious of it. All three applications also demonstrate how pursuit calibration does not require a clear start or end but adapts to the user. We implemented them to show the feasibility of the method and collect the time needed for users to calibrate on each system. For this purpose, we used a fixed number of samples as a calibration completion criteria, instead of a fixed duration as in the accuracy study. A summary of the settings and results of the three studies is shown in table 1.

ATM

Kumar et al. [11] suggested an eye-based ATM as it significantly increases security. This is based on accurate eye tracking to select the digits, enabled by a user calibration procedure. In real ATM situations users can get distracted, e.g. from passers-by, vehicles or barking dogs. During a normal calibration procedure, any eye movement to see outside the screen, caused by reflex, would void the calibration and lengthen the process. We evaluate the performance of pursuit calibration when these distractions occur (see Figure 7b).

We created an interface where an explicit calibration stimulus is shown: in realistic settings, users wanting to use an eye-based interface would be aware they have to undergo a calibration procedure. The moving target of the calibration follows a circular path within the interface (around the white circle at fig. 7a). Once calibrated a numeric keypad appears in the area where the calibration stimulus was located, allowing

(a) A user calibrating to the ATM application by looking at the circular moving target.



(b) Distraction positions when users interact with the interactive ATM application.

Figure 7: User study with the ATM application.

the user to enter a four-digits PIN by dwelling on the numbers for one second.

We evaluated the time needed to calibrate and complete the task for three cases: when no distraction interrupted the user, with an on-screen distraction and with an off-screen distraction. The on-screen distraction was a red circle shown for three seconds appearing when the calibration reached 100 valid samples, which was 50% of the total number of samples needed to complete the procedure. Users were instructed to voluntarily look at the target. The off-screen distraction was created by the experimenter, who gave an oral sign to the user to turn towards them and tell the number of fingers the experimenter was showing with their hand before returning to the calibration. This distraction created an interruption of 2.6 seconds (SD=1.5s) on average.

20 paid users aged between 19 to 31 years (mean=22.4, 12 female) took part in the user study. Participants stood in front of a 40" display with a Tobii X300 eye-tracker placed underneath (see Figure 7a). Data was collected at 60Hz. Each condition was repeated twice.

Results

The user study showed that all users calibrated without distraction in 7.8s, and needed an additional 4.6 seconds to complete the calibration in the case of an on-screen distraction, and 3 seconds for the off-screen case (see Table 1 for calibration times). These times correspond to the distraction durations and that users had to re-adjust to the calibration stimulus. The study showed that the method is robust against distractions during the calibration: users can look somewhere else for some time and resume calibration without having to start the procedure all over again.

Stargazing

Stagazing presents an interesting application, because it naturally includes static elements (stars) and moving elements (shooting stars). We use the shooting stars as the calibrating targets, while the stars are the objects users can interact with once calibrated. This is an example of how smoothly pursuit calibration can be integrated in an application, without separating it as a phase of its own.

We implemented an interface where a user can see stars glowing but cannot interact with them yet. Shooting stars appear randomly on four trajectories (see Figure 8b) and the user's eye is naturally attracted by them. The system detects when a user looks at a shooting star and collects samples for this area of the screen, while simultaneously notifying the user that something is happening by highlighting the shooting star in red. This means the calibration part was subtle but still noticeable. Once the calibration is complete, stars become gaze-aware and present additional information when looked at, such as their name and constellation lines (see Figure 8c).

The system was set up in a public area of the computer science department of a university (Figure 8a). It uses the same hardware and setup as the ATM study. Passers-by were encouraged to take part in the study by an experimenter that assisted them in placing themselves when needed.

Results

Based on application and video log we counted 247 passers-by of which 37 interacted with the system and 30 of them calibrated successfully (81%) in a median calibration time of 30.9 seconds (SD=28s, MIN=14.5s, MAX=120.8s). We report the median, as few users calibrated much longer because of poor eye detection caused by eye physiology [15], lots of head movements and interface design.

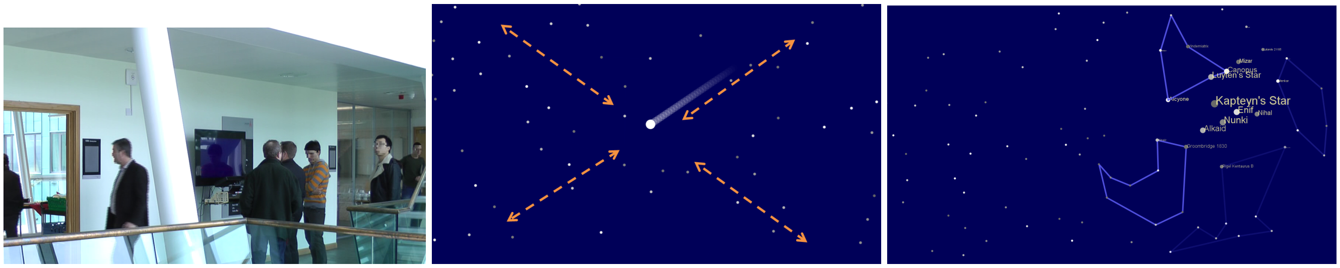
In general, users instantly understood that they could explore the stars and constellations by moving their eyes. Despite inexperience with eye-tracking, most users were able to calibrate with little assistance or instructions. This application demonstrates how calibration can be seamlessly integrated into an application without having a sharp transition or explaining to the user what their task is.

Waiting screen

This application investigates the feasibility of calibrating users when they are completely unaware that they are being calibrated. We aimed for the calibration to blend in the interface so that it is virtually invisible and the user does not have the impression that a task is required of them to enable the interaction. The methodology is based on Latane and Darley [12], in which users are in a waiting room waiting to be called for an arbitrary task while in reality the actual experiment investigates what happens during the waiting time.

In order to entice participants to use the system naïvely, we randomly invited passers-by in the university foyer of the management department. We described that users were to test an eye-based interface where they had to "click" on coloured circled by looking at them. They were told the goal of the experiment was to measure how fast people can select targets with their eyes and to perform as few head movement as possible therefore, but calibration was never mentioned in any way. They were shown a short example of the task so they could visualise it. After being placed in front of the system, a waiting screen appeared and the experimenter pretended to be "starting the system".

The waiting screen was designed to be the actual calibration procedure. It consisted of a study title and floating words enticing the user to wait (see figure 9). Words read "Please



(a) Setup of the system and during the field study. The experimenter encouraged passers-by to participate and assisted in placing them when needed.

(b) Calibration phase. The arrows show the trajectory of all possible shooting stars (arrows itself not part of the application). During runtime, one shooting star is shown at a time.

(c) Interaction phase. Constellations and star names become visible when the user looks at them. The size of the visuals increased around the gaze direction of the user.

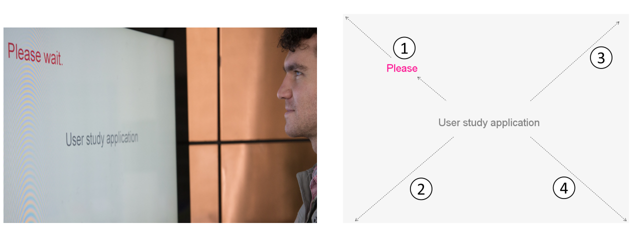
Figure 8: With the stargazing application, we design a seamless transition between calibration and interaction phase. Users first look at shooting stars and get calibrated. When the calibration finishes, the system switches to interaction mode. Users interact with a gaze aware interface that smoothly displays names and constellations wherever they look at.

wait" in five different languages, in order to distract the user enough during the waiting phase and subtely encourage them to keep reading. Our aim was to display a waiting screen that has movement in it for calibration, but is subtle enough so users do not get suspicious that it is part of the study. Words slowly faded in and out to enable a smooth experience. When 100 calibration points were collected, the task started automatically.

The setup was the same as the accuracy evaluation, where users were sitting in front of a display. We recruited 14 participants (5 female). The font size of the words was 30px, and the calibration point it was calibrating for was the exact centre of the word. As the user can gaze at any position within the word's bounding box, inaccuracy up to 30px in height and 30px$\times$wordlength in width can occur. The task users performed after calibration is selecting 9 circular objects with a width of 250px.

Results

The last application showed that 13 of 14 users were calibrated in 13.6 seconds on average (SD=5.3s) and performed the circle task easily, without being aware of a calibration procedure. By using floating words, we were able to successfully catch the attention of the user to a moving target. However, as words are not exact points, the calibration accuracy decreased, which is why we used a coarse task after the calibration. This can be avoided by using different target design.



(a) (b)

Figure 9: A user calibrates by looking at a floating word (a). These words move through four trajectories (b). Arrows, circles and numbers are not part of the interface but illustrate path (arrows) and order of appearance (numbers) of attention attracting floating words. From 1 to 4, it consecutively shows 'Please', 'wait.', 'Bitte', 'warten.'.

DISCUSSION

The detection of user attention in pursuit calibration presents a significant advance over previous methods, both in terms of calibration and usability. From a calibration point of view, a key advantage is that samples for calibration are only collected at points the user is known to have looked at. In conventional calibration, outliers can be filtered from a point cloud of samples at each display point, but the entire point cloud can be off target, resulting in a poor calibration.

The usability advantages lie primarily in the flexible calibration design enabled by our approach, where the user is not required to follow a rigid procedure. This renders eye based interfaces more realistic to use in everyday settings where distractions and interruptions are common. The method enables designers to introduce users to the application through calibration, rather than showing a separate necessary step to undergo before being able to interact. One can envision this kind of introduction with any kind of naturally moving objects: flying birds before discovering the highlights of a nature trail, racing cars followed by information on different car models, snowflakes falling and leading to the latest information on ski trails status, etc.

Pursuit calibration can be beneficial in application areas where users can not be assumed to cooperate actively toward calibration. An example may be infant eye-tracking studies: while it is hard to keep an infant's gaze focussed on one point, their attention can easily be caught by moving coloured objects. In eye-tracking studies, experimenters can more generally benefit from our method because it removes the need to manually check the accuracy of collected samples, which is now commonly required in the process. Our method can also calibrate users unawares which opens opportunities for casual interaction with public devices, as well as for studies in which it may be required that participants are not conscious of their eye movement being analysed.

Limitations

We used a high-range remote eye-tracker to evaluate the accuracy and usability of the method. However, pursuit calibration is not limited to such device and the method is generic. Because it is based on eye movement, it could work similarly with a mobile eye-tracker and off-the-shelf trackers. The requirement is that both eye data and screen coordinate system evolve in the same direction for the correlation to work.

In order for pursuit calibration to be integrated into applications, it requires careful interface design. The accuracy study showed that a target that is too quick is hard to follow for the eye, while one that is too slow avoids detection of correlation. In addition, the target's path needs to be equally distributed along the screen as well as cover both horizontal and vertical directions to attain high accuracy. This can be automated by using the iterative characteristic of pursuit calibration. Additionally, the calibration target should be as small as possible. Small targets offer an exact point the user can gaze at, while larger targets, such as words used in the waiting screen study, lead to more approximation and a less accurate calibration.

During the accuracy evaluation, we only tested a rectangle path along the borders of the screen. It would be interesting to conduct in-depth investigations of other paths, for example a circle path as used in the ATM application. Another interesting way to extend the accuracy evaluation would be to create a model of the accuracy obtained depending on the target path used. Such a model would allow the procedure to estimate when the desired accuracy has been reached, instead of calibrating until a certain number of samples are collected. This could further optimise the calibration target placement and shorten the calibration duration.

CONCLUSION

In this paper we introduced pursuit calibration as a novel method for gaze calibration. The method represents a radical departure from conventional approaches, as it calibrates eye gaze against a moving target, exploiting smooth pursuit eye movement to determine when the user is following the target. This leads to novel strategies for more intelligent sampling of calibration points, as well as usability advantages in tolerating user distraction and facilitating casual calibration. Pursuit calibration is simple in its core design but lends itself to implement gaze calibration in creative and innovative ways.

REFERENCES

1. Chen, J., and Ji, Q. Probabilistic gaze estimation without active personal calibration. In *Proc. of CVPR '11* (2011), 609–616.
2. Drewes, H., and Schmidt, A. Interacting with the computer using gaze gestures. In *Proc. of INTERACT '07*, Springer (2007), 475–488.
3. Fischler, M. A., and Bolles, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM 24*, 6 (1981), 381–395.
4. Flatla, D., Gutwin, C., Nacke, L., Bateman, S., and Mandryk, R. Calibration games: Making calibration tasks enjoyable by adding motivating game elements. In *UIST '11* (2011), 403–412.
5. Guestrin, E., and Einzenmann, M. General theory of remote gaze estimation using the pupil center and corneal reflections. In *IEEE Trans. on Biomedical Engineering*, vol. 53 (2006), 1124–1133.
6. Guestrin, E. D., and Eizenman, M. Remote point-of-gaze estimation requiring a single-point calibration for applications with infants. In *Proc. of ETRA '08*, ACM (2008), 267–274.
7. Hutchinson, T., White, K.P., J., Martin, W., Reichert, K., and Frey, L. Human-computer interaction using eye-gaze input. *IEEE Trans. on Systems, Man, and Cybernetics, Part C 19*, 6 (1989), 1527 –1534.
8. Jacob, R. J. K. What you look at is what you get: eye movement-based interaction techniques. In *Proc. of CHI '90*, ACM (1990), 11–18.
9. Kang, I., and Malpeli, J. G. *Behavioral calibration of eye movement recording systems using moving targets*, vol. 124. Journal of Neuroscience Methods, 2003.
10. Kondou, Y., and Ebisawa, Y. Easy eye-gaze calibration using a moving visual target in the head-free remote eye-gaze detection system. In *Proc. of VECIMS '08* (2008), 145 –150.
11. Kumar, M., Garfinkel, T., Boneh, D., and Winograd, T. Reducing shoulder-surfing by using gaze-based password entry. In *Proc. of SOUPS '07*, ACM (2007), 13–19.
12. Latane, B., and Darley, J. M. Bystander "apathy". *American Scientist 57* (1969).
13. Model, D., and Eizenman, M. User-calibration-free remote eye-gaze tracking system with extended tracking range. In *Proc. of CCECE '2011* (2011), 1268–1271.
14. Morimoto, C. H., and Mimica, M. R. M. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding 98*, 1 (2005), 4–24.
15. Nyström, M., Andersson, R., Holmqvist, K., and van de Weijer, J. The influence of calibration method and eye physiology on eyetracking data quality. *Behaviour Research Methods* (2012).
16. Ohno, T., and Mukawa, N. A free-head, simple calibration, gaze tracking system that enables gaze-based interaction. In *Proc. of ETRA '04*, ACM (2004), 115–122.
17. Renner, P., Lüdike, N., Wittrowski, J., and Pfeiffer, T. Towards continuous gaze-based interaction in 3d environments - unobtrusive calibration and accuracy monitoring. In *GI VR/AR workshop*, Shaker Verlag (2011), 13–24.

18. Robinson, D. The mechanics of human smooth pursuit eye movement. *The Journal of Physiology 180*, 3 (Oct. 1965), 569–591.

19. Schnipke, S. K., and Todd, M. W. Trials and tribulations of using an eye-tracking system. In *Proc. of CHI EA '00*, ACM (2000), 273–274.

20. Sibert, L. E., and Jacob, R. J. K. Evaluation of eye gaze interaction. In *Proc. of CHI '00*, ACM (2000), 281–288.

21. Smith, J. D., Vertegaal, R., and Sohn, C. Viewpointer: lightweight calibration-free eye tracking for ubiquitous handsfree deixis. In *Proc. of UIST '05*, ACM (2005), 53–61.

22. Stellmach, S., and Dachselt, R. Look & touch: gaze-supported target acquisition. In *Proc. of CHI '12*, ACM (2012), 2981–2990.

23. Turner, J., Alexander, J., Bulling, A., Schmidt, D., and Gellersen, H. Eye pull, eye push: Moving objects between large screens and personal devices with gaze & touch. In *Proc. of INTERACT '13* (2013).

24. Vertegaal, R. Attentive user interfaces. *Commun. ACM 46*, 3 (2003), 30–33.

25. Vidal, M., Bulling, A., and Gellersen, H. Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proc. of UbiComp '13*, ACM (2013).

26. Vidal, M., Pfeuffer, K., Bulling, A., and Gellersen, H. Pursuits: Eye-based interaction with moving targets. In *Proc. of CHI EA '13*, ACM (2013).

27. Villanueva, A., and Cabeza, R. A novel gaze estimation system with one calibration point. *IEEE Trans. on Systems, Man, and Cybernetics, Part B 38*, 4 (2008), 1123–1138.

28. Villanueva, A., Cabeza, R., and Porta, S. Eye tracking system model with easy calibration. In *Proc. of ETRA '04*, ACM (2004), 55–55.

29. Ware, C., and Mikaelian, H. H. An evaluation of an eye tracker as a device for computer input. *SIGCHI Bull. 17*, SI (1986), 183–188.

30. Young, L., and Sheena, D. Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation 7* (1975), 397–429.

31. Zhai, S., Morimoto, C., and Ihde, S. Manual and gaze input cascaded (magic) pointing. In *Proc. of CHI '99*, CHI '99, ACM (1999), 246–253.

32. Zhang, Y., Bulling, A., and Gellersen, H. Sideways: A gaze interface for spontaneous interaction with public displays. In *Proc. of CHI '13*, ACM (2013).

33. Zhu, Z., and Ji, Q. Eye gaze tracking under natural head movements. In *Proc. of CVPR '05*, IEEE Computer Society (2005), 918–923.