

# Idiosyncratic Feature-Based Gaze Mapping

Pieter Blignaut

University of the Free State  
Bloemfontein, South Africa

It is argued that polynomial expressions that are normally used for remote, video-based, low cost eye tracking systems, are not always ideal to accommodate individual differences in eye cleft, position of the eye in the socket, corneal bulge, astigmatism, etc. A procedure to identify a set of polynomial expressions that will provide the best possible accuracy for a specific individual is proposed. It is also proposed that regression coefficients are recalculated in real-time, based on a subset of calibration points in the region of the current gaze and that a real-time correction is applied, based on the offsets from calibration targets that are close to the estimated point of regard.

It was found that if no correction is applied, the choice of polynomial is critically important to get an accuracy that is just acceptable. Previously identified polynomial sets were confirmed to provide good results in the absence of any correction procedure. By applying real-time correction, the accuracy of any given polynomial improves while the choice of polynomial becomes less critical. Identification of the best polynomial set per participant and correction technique in combination with the aforementioned correction techniques, lead to an average error of  $0.32^\circ$  ( $SD = 0.10^\circ$ ) over 134 participant recordings.

The proposed improvements could lead to low-cost systems that are accurate and fast enough to do reading research or other studies where high accuracy is expected at frame-rates in excess of 200 Hz.

---

**Keywords:** Low-cost eye tracking, polynomial regression, calibration, accuracy, idiosyncratic

---

## Introduction

The output from eye tracking devices varies with individual differences in the shape or size of the eyes, such as the corneal bulge and the relationship between the eye features (pupil and corneal reflections) and the foveal region on the retina. Ethnicity, viewing angle, head pose, colour, texture, light conditions, position of the iris within the eye socket and the state of the eye (open or closed) all influence the appearance of the eye (Hansen & Ji, 2010) and therefore, the quality of eye tracking data (Holmqvist et al., 2011). In particular, the individual shapes of participants' eye balls, and the varying positions of cameras and

illumination require all eye trackers to be calibrated. The primary question in this paper is whether calibration suffices to cater for all individual differences. If not, the question arises how to accommodate idiosyncrasies to suit individual participants.

Besides accommodating idiosyncrasies, the focus of this study is also on the improvement of low-cost eye tracking to such an extent that it can be used for studies where a high level of accuracy is needed. Lack of accuracy, also known as systematic error, may not be a problem in usability studies when the areas of interest are large and are separated by large distances (Zhang & Hornof, 2011), but in studies where the stimuli are closely spaced as in reading (Rayner et al., 2007), uncertainty of as little as  $0.5^\circ - 1^\circ$  can be critical in the correct analysis of eye tracking data. Rayner et al. (2007, p. 522) states that "...there can be a discrepancy between the word that is attended to even at the beginning of a fixation and the word that is recorded as the fixated word." Accuracy is also of great importance for gaze-input systems (Abe et al., 2007).


Received March 1, 2016; Published April 21, 2016.

Citation: Blignaut, P.J. (2016). Idiosyncratic Feature-Based Gaze Mapping. *Journal of Eye Movement Research*, 9(3):2, 1-17.

Digital Object Identifier: 10.16910/jemr.9.3.2

ISSN: 1995-8692

This article is licensed under a [Creative Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)

[International license.](https://creativecommons.org/licenses/by/4.0/) 

Systematic errors may result from bad calibrations, head movements, astigmatism, eye-lid closure or other sources strongly dependent on the particular characteristics of the individual participant (Hornof & Halverson, 2002). Systematic errors can be several degrees of visual angle, which may have a serious impact on results that refer to the number of fixations or the amount of time spent on a specific area of interest.

Several attempts to correct for systematic error, both in real-time and post-hoc, were made in the past (cf. Buscher et al., 2009; Hornof & Halverson, 2002; Hyrskykari, 2006; Zhang & Hornof, 2011; Blignaut et al. 2014) and details of these are discussed below. This paper explains a three-step procedure where

- (i) a gaze mapping polynomial set will be identified for individual users, instead of using a one-size-fits-all set for everybody;
- (ii) regression coefficients will be recalculated in real-time, based on a subset of calibration points in the region of the current gaze; and
- (iii) a real-time localized correction will be done based on calibration targets to the same region.

## Video-based Eye Tracking

### *Gaze mapping*

Video-based eye tracking is based on the principle that near-infrared (NIR) light shone onto the eyes is reflected off the different structures in the eye to create four Purkinje reflections (Crane & Steele, 1985). The transformation from eye-position to point of regard (PoR) can be either model-based (geometric) or regression-based (Hansen & Ji, 2010). With model-based gaze estimation, a model of the eye is built from the observable eye features (pupil, corneal reflection, etc.) to compute the gaze direction. See Hansen and Ji (2010) for a comprehensive overview of possible transformations.

Regression-based systems use polynomial expressions to determine the point of regard as a function of the pupil-glinton vector in the eye image, using a least squares estimation to minimize the distances between the observed points and the actual points (Hoorman, 2008). Other examples of 2-dimensional interpolation schemes can be found in McConkie (1981) as well as Kliegl and Olson (1981), while a cascaded polynomial curve fit method is described by Sheena and Borah (1981). Sesma-Sanchez et al. (2016) describes a procedure using Gaussian regression.

Polynomial models should include two independent variables ( $x$  and  $y$  components of the pupil-glinton vectors) which may or may not interact with each other for each one of the dependent variables ( $X$  and  $Y$  of the point of regard) separately. Corrections for head movement can be done by normalising the pupil-glinton vector in terms of the distance between the glintons (if there are more than one IR source) or between the pupils (inter-pupil distance (IPD)).

A set of  $n$  points can be approximated with a polynomial of  $n$  or fewer terms

$$\begin{aligned} X &= \sum_{k=0}^{n-1} a_k x^i y^j, i \in [0, k], j \in [0, k] \\ Y &= \sum_{k=0}^{n-1} b_k x^i y^j, i \in [0, k], j \in [0, k] \end{aligned} \quad (1)$$

where  $x$  and  $y$  refer to the normalised camera  $x$  and  $y$  components of the pupil-glinton vector of a specific eye at a specific point in time, and  $X$  refers to the  $X$ -coordinate of the PoR for the specific eye on the two dimensional plane of the screen. A similar model can be used for the  $Y$ -coordinate of the PoR for the specific eye.

The coefficients  $a_k$  and  $b_k, k \in [0, n-1]$ , are determined through a calibration process that requires the user to focus on a number of dots (also referred to as calibration targets) at known positions while storing the positions of the feature points (pupils and glintons) (Abe et al., 2007; Kliegl & Olson, 1981; Tobii, 2010). A least squares regression is then applied separately for each eye to determine the coefficients such that the differences between the reported gaze coordinates and the calibration targets are minimised. During live gaze recording, these coefficients are used to calculate an interpolated point of regard as the average of the  $(X_{left}, Y_{left})$  and  $(X_{right}, Y_{right})$  coordinates.

Ideally, the mapping of feature points on the camera sensor to gaze coordinates on the stimulus plane should remove any systematic error, but the limited number of calibration points that are normally used, limits the accuracy that can be achieved. Typical calibration schemes require 5 or 9 pre-defined points, and rarely use more than 20 points (Borah, 1998).

While it is true that the calibration procedure should not distract from the main study and should preferably not be time consuming (Brolly & Mulligan, 2004), it is also important that a short calibration should not be conducted at the cost of accuracy. This study proposes a procedure of 45 dots of which some are used for regression and the full set is used to validate and improve accuracy. It is acknowledged that this is an unusual high number of points, however, the specific way in which the points are presented

assists to moderate the tedious nature of the task. Furthermore, the entire process can be completed in about 30 seconds – a small price to pay for the advantage of improved accuracy.

### Distortions

In ideal circumstances, an affine mapping of the pupil-glint vector on the camera sensor to gaze coordinates on the stimulus plane should suffice. The simplest model would be to map the gaze coordinates in terms of a linear relationship with the normalised pupil-glint vector without considering interactions between the two dimensions:

$$\begin{aligned} X &= a_0 + a_1x \\ Y &= b_0 + b_1y \end{aligned} \quad (2)$$

or adding a term for the other dimension:

$$\begin{aligned} X &= a_0 + a_1x + a_2y \\ Y &= b_0 + b_1x + b_2y \end{aligned} \quad (3)$$

With these mappings, the parameters,  $a_0$  and  $b_0$  represent a shift of the coordinates in the horizontal or vertical directions while  $a_1/b_1$  and  $a_2/b_2$  express the rotation of the points about the axis perpendicular to the plane.

Unfortunately, there are several factors that undermine the effectivity of an affine transformation for a remote video-based eye tracker:

- The flat display means that points at the left and right edges are further from the eyes than points in the centre of the display. This problem can be alleviated through the use of curved display units, but these are not yet commonly used.
- Depending on the participant's seating position, points at the top edge of the display could be nearer to the eyes than points at the bottom edge or vice versa. See Figure 3 for example.
- The mounting of the camera below the bottom edge of the screen causes varying gaze angles across the display.
- Remote video-based eye tracking assumes that the eye is a sphere that only rotates around its centre (Morimoto & Mimica, 2005). If the camera and light source(s) are fixed, the position of the corneal reflection(s) do not move with the eye rotation, and therefore can be used as a reference point. However, the eye is seldom perfectly spherical and differs from one partic-

ipant to the next, depending on physiological differences or vision correction such as glasses, contact lenses or surgery.

- Habitual behaviour can also affect the transformation. For example, some participants tend to keep their heads still while watching targets at various positions across the display. Others tend to move their heads slightly from side to side with consequently less rotation of the eye balls.

Due to the above mentioned factors, an affine transformation as in Equation 2 can cause distortions such as the pin-cushion, barrel or moustache effects (Van Walree, 2015) (cf Figure 1). Using the mapping polynomial set of Equation 3, the distortions are clearly demonstrated in Figure 2.

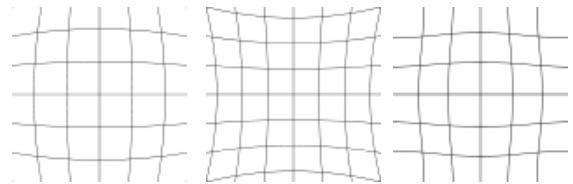


Figure 1: Barrel, pin-cushion and moustache distortion effects (Source: Wikipedia)

### Specific polynomials

Figure 2 clearly shows that no translational or rotational transformation will succeed in correcting the offsets and therefore interactions and higher order terms should be considered. Several polynomials have been proposed in the past, with varying success. Hennessey [2008] proposed the addition of a single interaction term to the affine transformation of Equation 3:

$$\begin{aligned} X &= a_0 + a_1x + a_2y + a_3xy \\ Y &= b_0 + b_1x + b_2y + b_3xy \end{aligned} \quad (4)$$

Zhu and Ji (2005) adapted the Hennessey model somewhat with respect to the Y-coordinate:

$$\begin{aligned} X &= a_0 + a_1x + a_2y + a_3xy \\ Y &= b_0 + b_1x + b_2y + b_3y^2 \end{aligned} \quad (5)$$

A second order polynomial in  $x$  and  $y$  with first order interactions are used by Mitsugami, Ukita and Kidode (2003), Morimoto and Mimica (2005) and Cerrolaza et al. (2012). This model can also be extended to include second order terms (in brackets):

$$\begin{aligned} X &= a_0 + a_1x + a_2x^2 + a_3y + a_4y^2 + a_5xy (+ a_6x^2y^2) \\ Y &= b_0 + b_1x + b_2x^2 + b_3y + b_4y^2 + b_5xy (+ b_6x^2y^2) \end{aligned} \quad (6)$$

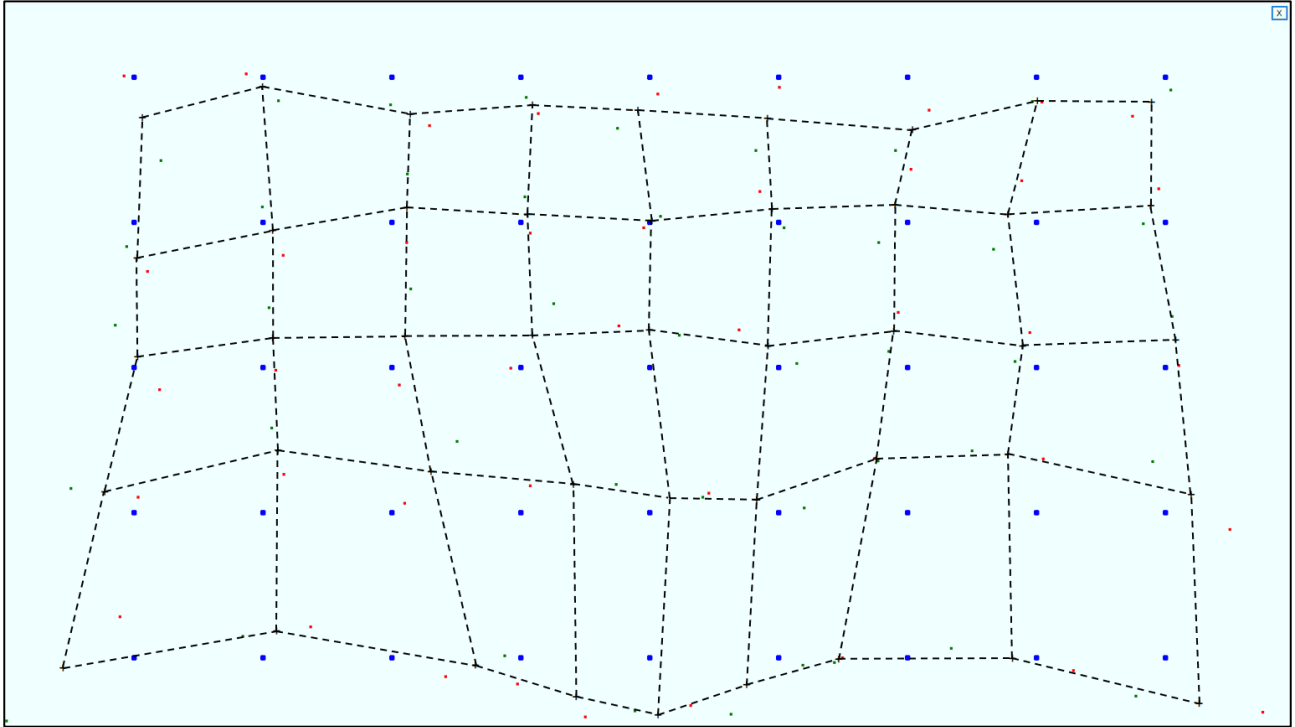


Figure 2. Visualisation of mapped gaze coordinates in relation to 45 calibration targets (indicated in blue) using an affine linear transformation from pupil-glint vector on the camera sensor to gaze coordinates on the stimulus. The green dots represent the gaze positions of the left eye and the red dots that of the right eye. The average gaze positions of the two eyes are connected to illustrate the trends in offsets across the display. The combination of the pincushion and moustache effects is clearly visible at the bottom of the display.

Cerrolaza and Villanueva (2008) generated a large number of mapping functions, varying the degree and number of terms of the polynomial. They found that, apart from some of the simplest models, increasing the number of terms or the order of the polynomial had almost no effect on accuracy. A preferred model was chosen as one that showed good accuracy in addition to having a small number of terms and being of low order:

$$\begin{aligned} X &= a_0 + a_1x + a_2x^2 + a_3y \\ Y &= b_0 + b_1x^2 + b_2y + b_3xy + b_4x^2y \end{aligned} \quad (7)$$

In a previous study (Blignaut & Wium, 2013), the accuracy of 625 polynomials for a simple one light, one camera configuration was examined and the following model was found to provide the best results for all participants as long as at least 8 calibration points are used:

$$\begin{aligned} X &= a_0 + a_1x + a_2x^3 + a_3y^2 + a_4xy \\ Y &= b_0 + b_1x + b_2x^2 + b_3y + b_4y^2 + b_5xy + b_6x^2y \end{aligned} \quad (8)$$

In another previous study, Blignaut (2013) found the following model to provide good accuracy ( $< 0.5^\circ$ ) given

that enough calibration points were used to facilitate regression of the multi-term polynomials (Blignaut, 2013). In this paper, this model is referred to as the SAICSIT polynomials. This model was also tested by Sesma-Sanchez et al. (2016) and found to deliver an accuracy of  $0.29^\circ$  using a simulated user on a large screen.

$$\begin{aligned} X &= a_0 + a_1x + a_2x^2 + a_3x^3 + a_4y + a_5xy + a_6x^2y + a_7x^3y \\ Y &= b_0 + b_1x + b_2x^2 + b_3y + b_4y^2 + b_5xy + b_6x^2y \end{aligned} \quad (9)$$

Blignaut (2014) also tested the following model, but could not prove that it produced better results than the SAICSIT polynomials.

$$\begin{aligned} X &= a_0 + a_1x + a_2x^2 + a_3x^3 + a_4y + a_5y^2 + a_6y^3 + a_7xy \\ &\quad + a_8x^2y + a_9x^3y \\ Y &= b_0 + b_1x + b_2x^2 + b_3y + b_5xy + b_6x^2y \end{aligned} \quad (10)$$

*Post-hoc and real-time correction of offsets in gaze coordinates*

While it is common practice to recalibrate between trials, for example in reading research, it is not always possi-

ble or feasible, specifically in studies where task completion time is measured or where an interruption might impact on contextual information or the participant's thought processes. Furthermore, it has also been reported that eye trackers often maintain a systematic error even directly after careful calibration (Hornof & Halverson, 2002).

A possible approach to improve accuracy is to use the mouse to drag a fixation or group of fixations until they better match obvious salient objects on the stimulus – as can be done with the EyeLink system (SR-Research, 2007). Hyrskykari (2006) developed a method for reading in which inaccurate data is moved in real-time to the most probable line. Buscher, Cuttrell and Morris (2009) did a manual 9-point calibration after a recording to determine an offset that was then applied to all recorded data. In general, any correction procedure with a single central calibration point would be effective only if the error is of the same magnitude and direction across the entire display.

Hornof and Halverson (2002) reasoned that, because the systematic error is approximately constant within a region on the display, it is possible that the effect of systematic error on each recorded fixation can be offset by adjusting the fixation location based on the weighted average of the error vectors that are closest to that recorded fixation, with heavier weights assigned to closer error vectors. They introduced the idea of required fixation locations (RFLs) – screen locations that the participant must look at within a certain timeframe in order to successfully complete a task, but without explicit instructions to fixate on those locations. They showed how the disparity between the fixations recorded by the eye tracker and RFLs can be used to monitor the accuracy of the eye tracker and to automatically invoke a recalibration procedure when necessary. They also demonstrated how the disparity varies across screen regions and participants, and how each participant's unique error signature can be used to reduce the systematic error in the eye movement data that is collected.

Blignaut et al. (2014) focused on real-time adjustments (as opposed to post-hoc corrections of fixations) of raw eye tracking data, which is important for gaze-contingent or gaze-controlled systems. Five commercial eye trackers were used and every participant was calibrated using the manufacturer's calibration routine with recommended settings for best results. Thereafter, participants had to click on tiny blue dots that appeared in two sets in random order on an 8×5 grid. Based on the offsets between the actual and observed gaze positions recorded with the first set of 40 dots, a regression formula was applied in real-time while

recording data for the second set of 40 dots. The five closest points from the first set of dots were identified and a set of regression coefficients was calculated based on the offsets of the identified points. The regression approach succeeded to improve the accuracy for all participants although not to the same extent for each person and not always by a significant margin. On average, the approach succeeded to improve accuracy by about 0.3°-0.6° on each of the eye trackers that were tested.

## System Details

A system was developed as part of a larger project to implement the proposed procedure and capture calibration and real-time gaze data.

### Hardware

An average computer was used with a Duo-core, i5 CPU, 4 GB memory loaded with 64 bit Windows 8. A 21" screen with 1600 × 900 resolution (pixel size 0.277 mm) was used. A UI-1550LE-C camera with daylight filter from IDS Imaging Systems ([en.ids-imaging.com](http://en.ids-imaging.com)) was placed just below the screen. The camera has a 1600×1200 sensor with pixel size 2.8 μm and has a native framerate of 18.3 fps at its maximum recording window of 1600×1200 pixels. The camera was fitted with a 10 mm lens from Lensation (<http://www.lensation.de/>). Two infrared illuminators were placed on either side of the camera at a distance of 220 mm from the camera. Although the camera was more sophisticated than a web camera, the entire system, including computer, screen and camera, can be acquired for less than USD 1,000. Figure 3 shows the physical arrangement of camera and participant in front of the screen.

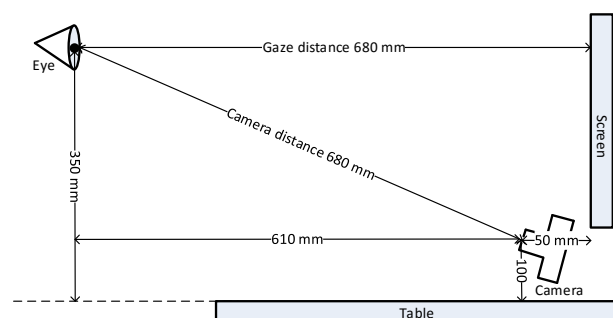


Figure 3: Hardware configuration

### Software

The software was developed using C# with .Net 4.5 along with the camera manufacturer’s software development kit (SDK) to control the camera settings. Figure 4 shows the **Settings** screen of this system.

The system used a fixed-size dynamic recording window of 600×150 px and included a functionality to allow the participant to move the head freely within a head box of about 400 mm sideways and 100 mm up and down. The position of the eyes in the recording window was used to adjust the recording window within the head box as the head moved around. The eyes were only lost if the recording window could not fit into the sensor area (1600×1200) (cf. Figure 4) or if the participant jerked the head to one side.

The infrared illuminators caused two glints in the eyes (Figure 2) but only the outer glint was used to mark the pupil-glint vector. The two illuminators provided enough light for a short exposure time to suffice and a framerate in excess of 200 Hz could be achieved with the recording window being much smaller than the sensor size. The system was optimized so that the computer’s duo-core CPU could be used optimally to ensure that all processing for a single frame could be achieved within the allotted 5 ms.

The **Settings** screen (Figure 4) also included an inspection panel with a live eye video and a panel indicating the position of the recorded window inside the camera’s sensor area. This panel moved as the participant moved the head around in the head box. The experimenter could also drag the panel with the mouse to cater for different participant heights or seating positions.

### Calibration procedure

When the user clicked on the **Calibrate** button in Figure 4, the calibration procedure was started. Participants were presented with five sets of 45 dots in a 9×5 grid as in Figure 5. Twenty-three of the dots (those encircled in Figure 5) were used as calibration targets, while the complete list of dots were used to select the best possible regression polynomial and to validate the accuracy of the regression.

The dots appeared in random order so as to prevent participants to pre-empt the position of the next dot and take the eyes away from a dot before the gaze was registered. Every dot was preceded with a shrinking number (Arial font starting at 48 pt reduced with 4 pt every 100 ms until 16 pt) to (i) attract attention to it and (ii) to provide feedback about the number of dots left. The numbers also contributed to moderate the tedious nature of the procedure.

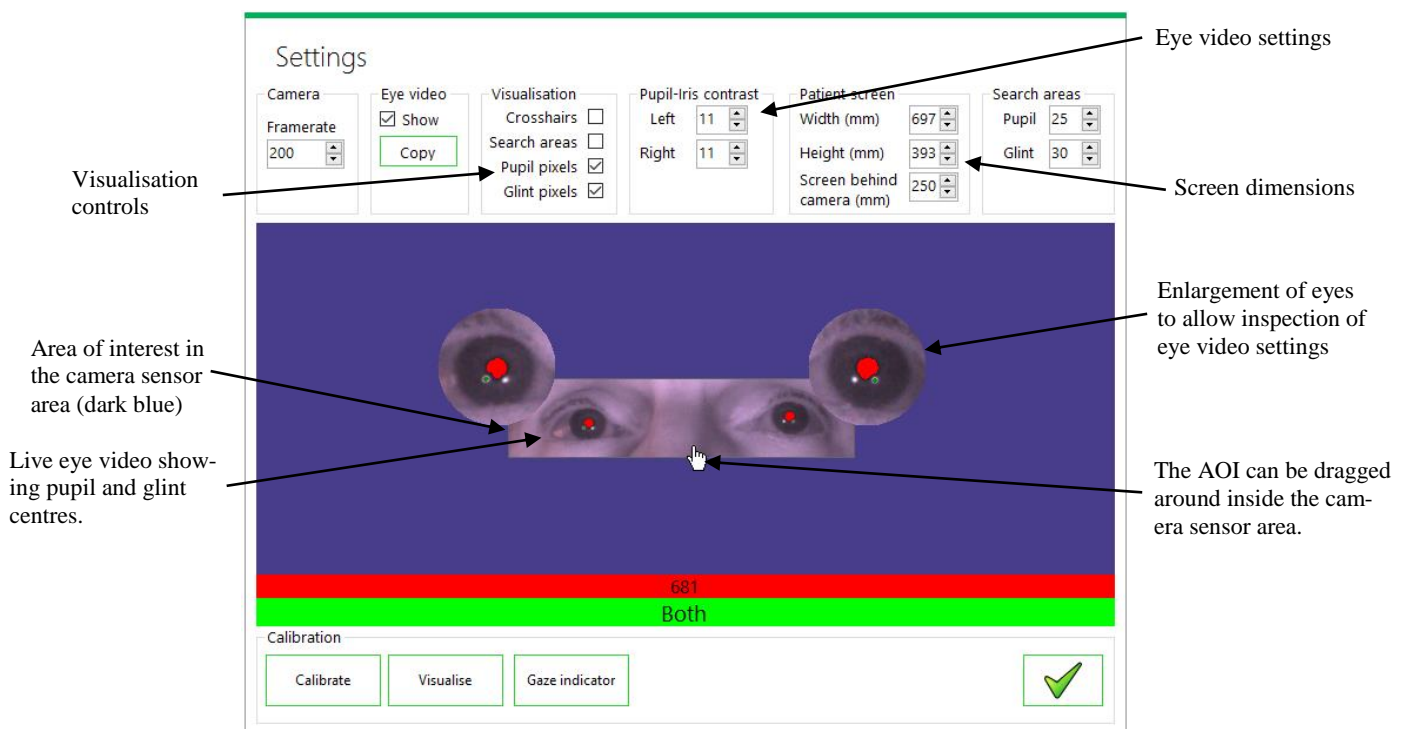


Figure 4: Experimenter’s live inspection screen

For every dot, the system waited for the eyes to stabilise and a window of 250 ms of stabilised gaze data (centre positions of the pupil and outer glint in each eye) was saved to an underlying SQLite database. (During live recording (next section), the calibration data was copied to internal memory to speed up processing time). For each dot, the median X and Y pupil and glint positions of the centre half of the window of samples were used in subsequent analyses. Gaze stability was measured in terms of precision on the camera sensor. (Note that this is not the same as the precision of mapped gaze coordinates on a stimulus plane.)

At a frame rate of 200 Hz (5 ms per interval), 50 samples were recorded for every dot. The dot disappeared immediately after the last sample was recorded and the next dot, preceded by its number, appeared. Participants with good reflexes and who could shoot to the next target and stabilize immediately, could visit the 45 dots in a reasonably short time. Recording of a single set of 45 dots mostly took less than 30 seconds. Recordings of participants with mascara took somewhat longer as these participants had to concentrate on opening their eyes wide to obtain the necessary gaze stability.

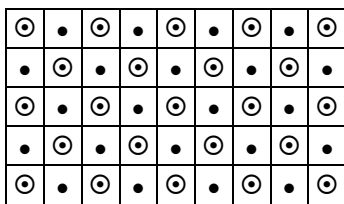


Figure 5: 9x5 grid of dots. All dots were displayed as • to participants. The ○ around the dots only serve to indicate the dots that were used for the regression.

### Recording procedure

The system, of which the **Settings** screen is shown in Figure 4, also makes provision for live recording of gaze data during on-screen reading. Gaze events are handled in real-time to map the pupil-glint vectors in the eye video to screen coordinates. This mapping is based on the recorded calibration data as explained above.

This paper proposes a three-step process to improve the accuracy of mapping. The procedure is outlined in Listing 1. For purposes of the listing, a three-dimensional array is used to define the polynomials in terms of the powers of  $x$  and  $y$  ( $i$  and  $j$  in Equation 1) for  $X$  and  $Y$  respectively. The Zhu and Ji (2005) set above (Equation 5) can, for example, be defined as  $(( (0,1),(1,0),(1,1) ), ( (0,1),(0,2),(1,0) ) )$ .

Firstly, instead of deciding on a specific polynomial and using it for all participants, it is proposed that a number of polynomials are evaluated and the best one used for mapping pupil-glint vectors to screen coordinates. See Step 1 in Listing 1 below.

Secondly (step 2.1 in Listing 1), it is proposed that during live recording, regression coefficients are based on the pupil-glint vectors of nearby calibration points only (LC). This will limit the effect of eye curvature or calibration points at gaze angles that are far away from the current point of regard. The subset of calibration points are based on the on-camera distance between the pupil-glint vector of the current sample and the pupil-glint vectors of the saved calibration points. The regression coefficients are used to interpolate an initial estimate of the point of regard.

One could argue that the regression process in Step 2.1 is time consuming and should be done once only before tracking is started. It should be noted, however, that since the subset of nearest points may change with every new camera frame, the regression coefficients must be calculated in real-time. When the LC flag is false, the algorithm indeed does the regression before recording starts.

Thirdly (step 2.3 in Listing 1), a real-time correction (RTC) is done according to the procedure described in Blignaut, et al. (2014) and referred to in Section 3 above. This entails that the regression coefficients from Step 2.1 are used to calculate screen coordinates for each of the calibration points in the list of 45. The nearest 4 points to the initial estimate of Step 2.1 are identified based on the on-screen distance from each point to the initial estimate. With a 9x5 grid on a 21" screen at a gaze distance of 680 mm, there will always be 4 calibrations points within a radius of 2° from the estimated gaze coordinates. The offsets of these points to the corresponding calibration targets are used in a regression to improve the calculated point of regard.

It should now be clear why a large number of calibration targets are needed during the calibration procedure. A set of points can be fitted with a polynomial that passes through or near all points but with large variations at in-between positions. To determine the average error, it would thus not suffice to use the same points that were used as calibration targets. In the proposed procedure, a set of 23 calibration points is used for the initial regression (Step 2.1) and then validated with a larger set of 45 points (Step 2.3). The larger set is also used to step through all possible polynomial sets and determine the best set for mapping gaze data in real-time.



```

int[,] currentPoly //Array of powers for x and y for X and Y

void StartRecording()
{
    currentPoly = GetPolynomial ()
    if (!LC) //Flag to indicate if calibration points are limited
        Using set of 23 calibration points, determine coefficients a0 ... an and b0 ... bn for currentPoly through regression
    Start eye tracking() //GazeEvent called on every camera sample
}

//Step 1: Define polynomial to use for interpolation
int[,] GetPolynomial ()
{
    if (findBestPoly)
    {
        //Get best polynomial by stepping through all polynomials and averaging the error over all points in the list of 45.
        minError = large value
        foreach (poly in set of polynomials)
        {
            sumError = 0
            foreach (point in list of 45)
                sumError += Distance(Actual, MapToScreen(poly, point.pg))
            if (sumError/45 < minError)
            {
                minError = sumError/45
                bestPoly = poly
            }
        }
        return bestPoly
    }
    else
        return SAICSIT or other fixed polynomial
}

//Step 2: Map to screen coordinates at every sample produced by the eye camera, based on the pupil-glint (pg) vector in the eye image
void GazeEvent(pg) //Every 5 ms while tracking
{
    MapToScreen(currentPoly, pg)
}

//Map pupil-glint (pg) vector to screen coordinates, using a given polynomial
void MapToScreen(poly, pg)
{
    //Step 2.1 Change
    if (LC) //Flag to indicate if calibration points are limited
    {
        Limit calibration set to nearest 11 points of set of 23 (based on Distance(pg, pt[i].pg))
        Determine coefficients a0 ... an and b0 ... bn for currentPoly through regression
    }

    //Step 2.2 Initial estimate
    Interpolate (pg) to get initial estimate of gaze coordinates.

    //Step 2.3 Real-time correction
    if (RTC) //Flag to determine if real-time regression is applied
    {
        - Interpolate (pg) to find calculated position of each of the 45 points.
          (There will be an offset (error) from each calculated position to the actual gaze position as defined by the calibration target)
        - Use the 4 nearest calculated points to the initial estimate (based on Distance(estimate, pt[i]) ) and do regression to correct for the offsets.
        - Interpolate (estimate) to find improved gaze position of initial estimate
    }
}

```

Listing 1. Recording procedure

## Methodology

### Participants

Forty-two participants were recruited through convenience sampling of passers-by. Participants wearing glasses were requested to remove them if they could see on the screen without them. Five participants were tested with their glasses after carefully adjusting the glasses such that

reflections off the brim or from the lenses of the glasses did not interfere with the corneal reflections (Figure 6).

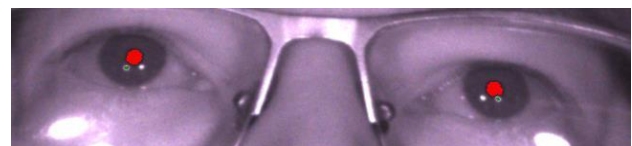


Figure 6: Adjustment of glasses so that extra reflections do not interfere with the glints



Six participants were wearing mascara and it could clearly be seen in the eye videos (Figure 7) that this adversely affected the image processing – especially if the eyes were narrow and the dark eye lids touched the pupils in the eye video. Sometimes it helped a bit if participants opened their eyes wide, causing the pupil and eyelids to move apart.



Figure 7: Mascara mistakenly being regarded as (part of) the pupils

### Data capturing

The experimenter used the eye video (Figure 4) to inspect participants’ position and requested them to adjust the chair so that they were seated at about 680 mm from the camera. The eventual average camera distance over all recordings was 668.0 mm (SD = 42.7 mm).

No headrest was used and participants were merely requested to keep their heads reasonably still. They were

allowed to move their heads sideways within bounds of the head box (see discussion below) and turn their heads towards the edges of the screen if it was necessary.

### Polynomials

Draper and Smith (1981) indicate that, apart from the brute-force method (i.e. testing all the possible equations), there is no systematic procedure that can provide the most suitable mapping equation. To determine an exhaustive list of all sets of polynomials to the third degree  $X = a_0 + a_1x + a_2y + a_3xy + \dots a_nx^3y^3$  is a matter of finding all possible combinations of "1,2,3,10,11,12,13,20,21,22,23,30,31,32,33" such that each number appear 0 or 1 times. This is  $2^{15} = 32,768$ . If all combinations of the polynomial for X are to be combined with all possibilities for Y, it means that  $2^{30} = 1,073,741,824$  iterations must be done in the outer loop of Step 1 of Listing 1. This can be done on a super computer but was not feasible in this study. For purposes of this study, polynomials were identified from literature (cf. Section 2.2 above) and potentially promising terms from one polynomial set were added to others to create an intuitive set of most probable good polynomials.

Table 1 lists the polynomials that were analysed for X and Y. For every participant, every polynomial for X was tested in combination with every polynomial for Y (29 x 27 = 783 polynomial combinations in total). A notation is used where  $X = a_0 + a_1x + a_2y + a_3xy$  is written as (1, x, y, xy). References are provided where applicable.

Table 1.1  
Polynomials for the X coordinates with references where applicable

1, x		1, x, x <sup>3</sup> , y	
1, x, y		1, x, x <sup>3</sup> , y <sup>2</sup>	
1, x, y, xy	Hennessey, 2008; Zhu & Ji, 2005	1, x, x <sup>3</sup> , xy	
1, x, xy		1, x, x <sup>3</sup> , y, y <sup>2</sup>	
1, x, y, xy, x <sup>2</sup> y <sup>2</sup>	Sesma-Sanchez, et al., 2016	1, x, x <sup>3</sup> , y, xy	
1, x, x <sup>2</sup>		1, x, x <sup>3</sup> , y <sup>2</sup> , xy	Blignaut & Wium, 2013
1, x, x <sup>3</sup>		1, x, x <sup>3</sup> , y, xy, x <sup>2</sup> y <sup>2</sup>	
1, x, x <sup>2</sup> , x <sup>3</sup>		1, x, x <sup>3</sup> , y <sup>2</sup> , xy, x <sup>2</sup> y <sup>2</sup>	
1, x, x <sup>2</sup> , y	Cerrolaza et al., 2008	1, x, x <sup>3</sup> , xy, x <sup>2</sup> y <sup>2</sup>	
1, x, x <sup>2</sup> , xy		1, x, x <sup>3</sup> , y, y <sup>2</sup> , xy	
1, x, x <sup>2</sup> , y, y <sup>2</sup>		1, x, x <sup>3</sup> , y, y <sup>2</sup> , xy, x <sup>2</sup> y <sup>2</sup>	
1, x, x <sup>2</sup> , y, xy		1, x, x <sup>2</sup> , x <sup>3</sup> , y, y <sup>2</sup> , xy, x <sup>2</sup> y <sup>2</sup>	
1, x, x <sup>2</sup> , xy, x <sup>2</sup> y <sup>2</sup>		1, x, x <sup>2</sup> , x <sup>3</sup> , y, xy, x <sup>2</sup> y, x <sup>3</sup> y	Blignaut, 2013 (SAICSIT)
1, x, x <sup>2</sup> , y, y <sup>2</sup> , xy	Mitsugami et al., 2003	1, x, x <sup>2</sup> , x <sup>3</sup> , y, y <sup>2</sup> , y <sup>3</sup> , xy, x <sup>2</sup> y, x <sup>3</sup> y	Blignaut, 2014
1, x, x <sup>2</sup> , y, y <sup>2</sup> , xy, x <sup>2</sup> y <sup>2</sup>	Mitsugami et al., 2003		

Table 1.2  
Polynomials for the Y coordinates with references where applicable

1, y		1, x, x <sup>2</sup> , y, y <sup>2</sup> , xy	Mitsugami et al., 2003
1, x, y		1, x, x <sup>2</sup> , y, y <sup>2</sup> , xy, x <sup>2</sup> y <sup>2</sup>	Mitsugami et al., 2003
1, x, y, xy	Hennessey, 2008	1, x, x <sup>2</sup> , y, y <sup>2</sup> , x <sup>2</sup> y	
1, x, y, y <sup>2</sup>	Zhu & Ji, 2005	1, x, x <sup>2</sup> , y, y <sup>2</sup> , xy <sup>2</sup>	
1, y, xy		1, x, x <sup>2</sup> , y, y <sup>2</sup> , xy, x <sup>2</sup> y	Blignaut, 2013 (SAICSIT)
1, x, y, xy, x <sup>2</sup> y <sup>2</sup>	Sesma-Sanchez, 2016	1, y, y <sup>2</sup> , xy, x <sup>2</sup> y <sup>2</sup>	
1, x, y, x <sup>2</sup> y <sup>2</sup>		1, x, y, y <sup>2</sup> , x <sup>2</sup> y	
1, x, y <sup>2</sup>		1, x, y <sup>2</sup> , x <sup>2</sup> y	
1, x <sup>2</sup> , y		1, x, y <sup>2</sup> , xy, x <sup>2</sup> y	
1, x, y <sup>2</sup> , xy <sup>2</sup>		1, x, y <sup>2</sup> , x <sup>2</sup> y <sup>2</sup>	
1, x <sup>2</sup> , y, xy		1, x, y <sup>2</sup> , xy, x <sup>2</sup> y <sup>2</sup>	
1, x <sup>2</sup> , y, x <sup>2</sup> y		1, x <sup>2</sup> , y, y <sup>2</sup> , xy, x <sup>2</sup> y <sup>2</sup>	
1, x <sup>2</sup> , y, xy, x <sup>2</sup> y	Cerrolaza et al., 2008	1, x, x <sup>2</sup> , y, xy, x <sup>2</sup> y	Blignaut, 2014
1, x <sup>2</sup> , y, xy, x <sup>2</sup> y <sup>2</sup>			

### Data selection

Eye tracking error can stem from any of the following three sources, namely the hardware, the software and the participant. When measuring the accuracy of an eye tracker, one should try to eliminate human error as far as possible. The system cannot be blamed if participants could not maintain concentration over 45 dots or some or other circumstantial event caused them to look away or interfered with data capturing.

The system provided an option for visualisation of the reported gaze coordinates in relation to the actual position of the 45 dots. These visualisations were inspected and the recording was removed from the data set if one or more dots was obviously not in line with the trend for the specific recording. See Figure 8 for two examples. Many recordings of participants with mascara or participants wearing glasses had to be removed.

Eventually, 134 recordings from 168 (79.8%) were retained for an average of 3.19 recordings (SD = 1.25) per participant. It should be noted that in a study with a different focus, one would possibly retain more recordings. However, in an accuracy study it is important that the results are not contaminated if there is any doubt regarding the validity of a recording. This is not viewed as superficially enhancing the accuracy, but rather seen as the cleansing of data where the source of the error is probably participant related.

The availability of live eye videos and eye images proved to be invaluable to do troubleshooting of difficult cases and clean the data. If all commercial manufacturers can provide these, it could go a long way towards improved data quality and reporting of it.

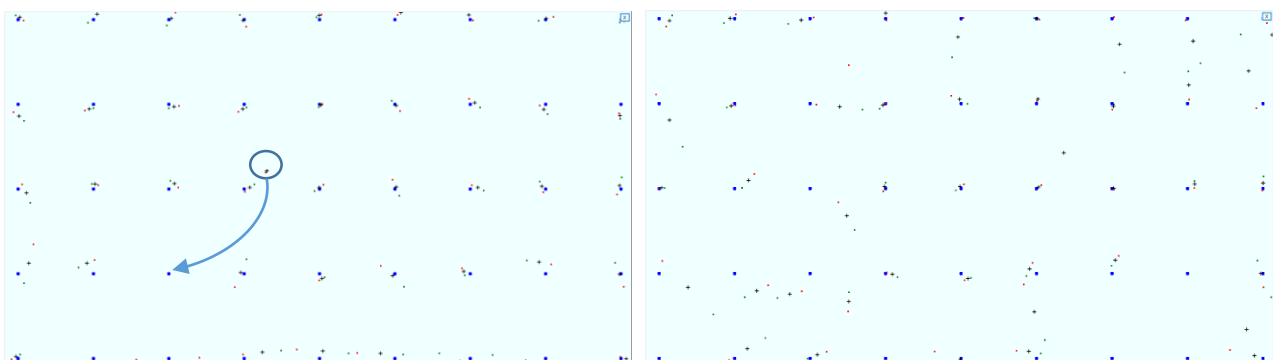


Figure 8. Visualisation of mapped gaze coordinates in relation to 45 calibration targets (indicated in blue). The green dots represent the gaze positions of the left eye and the red dots that of the right eye. The + indicates the average gaze position of the two eyes.

Left: Illustration of the occurrence of an obvious outlier.

Right: Illustration of the effect of mascara. The mapping is mostly very accurate but there are specific areas where the mascara causes large offsets.

## Analysis

The goals of this study can be summarised as follows:

- (i) Does the identification and application of different polynomials for different participants have a significant impact on the accuracy of gaze mapping?
- (ii) Does the limitation of calibration points to a regional area around the current gaze position have a significant impact on the accuracy of gaze mapping?
- (iii) Does a real-time correction (RTC) approach based on offsets between actual and calculated positions of calibration targets have a significant impact on the accuracy of gaze mapping?

To accommodate these goals, Listing 1 makes provision for two factors in a 2×4 design:

- Polynomial set
  - Fixed polynomial. The SAICSIT polynomial set (Blignaut, 2013) is used to determine the average error over all participants for each one of the correction techniques (findBestPoly = false).
  - A set of 783 potential polynomials is traversed to find the best polynomial per participant and correction technique (BPPP) (findBestPoly = true). Please note that this large number of polynomials are used in this study only. The best candidates are identified later and those are the only ones that should be utilised in future experiments.
- Real-time correction (RTC)
  - The full set of 23 calibration points is used without real-time correction (LC = RTC = false). This is the standard way of gaze mapping and will serve as benchmark against which the following three alternatives will be compared.
  - The full set of 23 calibration points is used with real-time correction (LC = false, RTC = true).
  - Limit calibration points to those nearest to the reported pupil-glint vector without correction (LC = true, RTC = false).
  - Calibration points are limited to those nearest to the reported pupil-glint vector with correction (LC = RTC = true).

For each recording, the various combinations of polynomial sets and correction technique are used to find the distance, in degrees, between the position of the 45 known target points and the reported point of regard. This error is averaged over all target points and participant recordings.

Although better accuracies were reported in the past, one needs to compare correction techniques and polynomial sets for a specific participant sample, hardware configuration and experimental circumstances. The fact that the results may be different than on a previous occasion does not impact on the outcome of the three goals stated above.

## Magnitude of Errors of the Proposed Corrections

### Magnitude of error per polynomial set and correction technique

The procedure described in Step 1 of Listing 1 was used to calculate the magnitude of mapping errors for various combinations of polynomial set and correction technique. The analysis below will use two fixed polynomials, namely the SAICSIT polynomial as published in the SAICSIT proceedings of 2013 (Blignaut, 2013), as well as the polynomial that provided the minimum overall error (averaged over all participants) for each correction technique. The following key is used to refer to the respective polynomial sets and correction techniques:

- The SAICSIT group of correction techniques refers to the polynomial set published in the SAICSIT proceedings of 2013 (Blignaut, 2013):

$$\begin{aligned} X &= a_0 + a_1x + a_2x^2 + a_3x^3 + a_4y + a_5xy + a_6x^2y + a_7x^3y \\ Y &= b_0 + b_1x + b_2x^2 + b_3y + b_3y^2 + b_4xy + b_5x^2y \end{aligned}$$

- BP refers to the polynomial that provided the minimum overall error (averaged over all participants) for each correction technique (cf Table 4):

With no correction:

$$\begin{aligned} X &= a_0 + a_1x + a_2x^2 + a_3x^3 + a_4y + a_5y^2 + a_6y^3 + a_7xy + a_8x^2y + a_9x^3y \\ Y &= b_0 + b_1x + b_2x^2 + b_3y + b_3y^2 + b_4xy + b_5x^2y \end{aligned}$$

RTC (With real-time correction):

$$\begin{aligned} X &= a_0 + a_1x + a_2y + a_3xy \\ Y &= b_0 + b_1x + b_2y^2 \end{aligned}$$

LC (Limit calibration points):

$$\begin{aligned} X &= a_0 + a_1x + a_2x^2 + a_3x^3 + a_4y + a_5y^2 + a_6xy + a_7x^2y^2 \\ Y &= b_0 + b_1x + b_2x^2 + b_3y + b_4y^2 + b_5xy + b_6x^2y^2 \end{aligned}$$

LC+RTC (Limit calibration points with real-time correction):

$$\begin{aligned} X &= a_0 + a_1x + a_2x^2 + a_3y + a_4xy \\ Y &= b_0 + b_1x + b_2y + b_3y^2 + b_4x^2y \end{aligned}$$

- BPPP (best polynomial per participant) refers to the polynomial that provided the minimum error for a specific participant recording for each correction technique.

The magnitudes of errors of selected polynomials (averaged over all participant recordings) in combination with the four correction techniques are shown in Table 2 and visualised in Figure 9.

Table 2  
Magnitude of mapping error per polynomial set and correction technique

Polynomial	Correction technique	Error (degrees)				
		Min	Max	Avg	SD	SEM
SAICSIT	None	0.40	1.75	0.69	0.25	0.021
	RTC	0.22	1.30	0.44	0.18	0.016
	LC	0.25	1.45	0.50	0.21	0.018
	LC+RTC	0.24	2.03	1.07	0.37	0.032
BP	None	0.39	1.75	0.69	0.25	0.021
	RTC	0.23	0.90	0.41	0.14	0.012
	LC	0.25	0.55	0.48	0.21	0.018
	LC+RTC	0.21	0.99	0.37	0.14	0.012
BPPP	None	0.39	1.74	0.67	0.23	0.020
	RTC	0.22	0.83	0.37	0.11	0.010
	LC	0.24	1.44	0.45	0.17	0.015
	LC+RTC	0.16	0.76	0.32	0.10	0.009

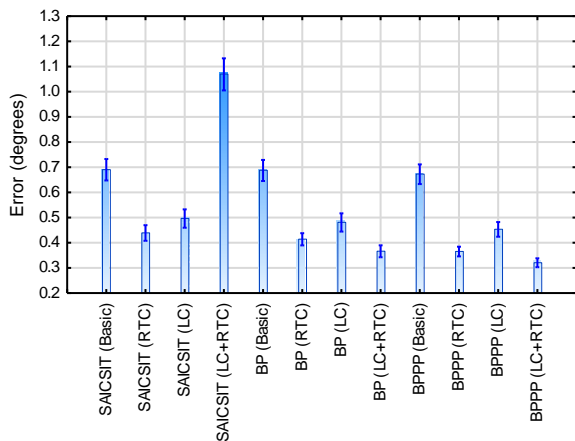


Figure 9: Magnitude of mapping error per polynomial set and correction technique. The spreads indicate the 95% conf intervals of the means.

Table 3  
Number of polynomials providing an average error over all participants below/above specific thresholds.

Correction technique	Number of polynomials below/above specified threshold													
	≤ 0.40	≤ 0.45	≤ 0.50	≤ 0.55	≤ 0.60	≤ 0.65	≤ 0.70	≤ 0.75	≤ 0.80	≤ 0.85	≤ 0.90	≤ 0.95	≤ 1.00	> 1
None							5	16	18	24	27	27	54	729
							0.6%	2.0%	2.3%	3.1%	3.4%	3.4%	6.9%	93.1%
RTC	682	783												
	87.1%	100%												
LC		8	52	118	193	240	286	372	507	623	673	717	66	
		1.0%	6.6%	15.1%	24.6%	30.7%	36.5%	47.5%	64.8%	79.6%	86.0%	91.6%	8.4%	
LC+RTC	404	667	675	675	675	675	675	675	685	689	694	711	722	61
	51.6%	85.2%	86.2%	86.2%	86.2%	86.2%	86.2%	86.2%	87.5%	88.0%	88.6%	90.8%	92.2%	7.8%

### General performance of polynomials per correction technique

Table 3 shows the number of polynomial sets (out of 783) that provide an average accuracy over all participant recordings below specific thresholds. If no correction is applied, only 0.6% of the polynomial sets provide an error of less than 0.7° and only 6.9% of polynomial sets provide an error of less than 1°. This means that the choice of polynomial is critically important to get an accuracy that is just acceptable.

By applying real-time correction (RTC) as explained above, not only does the average accuracy of any given polynomial improves, but the choice of polynomial also becomes basically irrelevant as all polynomial sets return an error of less than 0.5°.

By limiting the calibration points to within a certain distance of the observed pupil-glint vector (LC), 1.0% of polynomials returned an average error of less than 0.5°. On its own, therefore, there is no justification to use this technique instead of the basic procedure followed by real-time correction.

When real-time correction is added to LC (LC+RTC), 51.6% of polynomial sets provided an error of less than 0.40°. This is excellent, but it is important to be able to identify the polynomial sets that are able to achieve this high accuracy. The SAICSIT set that was previously confirmed to provide very good results with the basic procedure, does not perform well with this technique and returns a huge average error of 1.07°.

### Identification of good polynomials

Although the procedure to select the best polynomial set per participant (Step 1 in Listing 1) is done only once for every participant recording prior to live gaze recording, it can still be time consuming if a large number of polynomials must be evaluated. It is therefore advisable to limit the set of polynomials that must be evaluated to those with high probability of delivering good accuracy. Table 4 shows the best performing polynomial sets per correction technique as well as the average accuracy achieved when the best polynomial is applied per participant. The significance of the effect of polynomial set on mapping error is discussed in the following section.

With no real-time correction, the SAICSIT polynomial set (boldfaced in Table 4) is confirmed to provide good results. It is outperformed slightly if higher order  $y^2$  and  $y^3$  terms are added for the X-coordinate. With real-time correction, the SAICSIT polynomial set performs much worse – both with regard to its ranking in the list of best polynomials and with regard to the number of participants for

which it provides an accuracy within  $0.1^\circ$  of the best possible accuracy that can be attained for that participant.

Although the best performing polynomials can be included in the group from which the best polynomial set per participant and correction technique (BPPP) must be selected, it is always possible that a polynomial set not included in the group would be preferable for a specific individual. It is thus important to ensure that the polynomials that are included will provide good accuracy for every participant. Table 4 includes a column for the number of recordings for which a particular polynomial provided the smallest error. The table also includes columns for the number of recordings (out of 134) for which the difference in error provided by the particular polynomial and the smallest error provided by any polynomial, are less than  $0.1^\circ/0.2^\circ/0.5^\circ$ . The top few polynomials for each correction technique performed well with most of the participants. For limiting calibration points with real-time correction, about 75% of participant recordings could be accommodated within  $0.1^\circ$  of error from the best performing polynomial and about 97% within  $0.5^\circ$ .

Table 4

Best five performing polynomial sets, based on average error over all participants, per correction technique. Records are sorted first on average and then on standard deviation. The accuracies of the best and worst performing recordings are also listed.

Notes:

1. Polynomials are defined in terms of the powers of  $x$  and  $y$  ( $i$  and  $j$  in Eq. 1) with first digit in every pair representing the exponent of  $x$  and the second digit the exponent of  $y$ . The sequence of terms is such that the coefficients are numerically ordered and therefore the  $y$ -terms are listed before the  $x$ . E.g.  $X = a_0 + a_1x + a_2y + a_3xy$ ;  $Y = b_0 + b_1x + b_2y + b_3y^2$  is written as  $((01,02,11), (01,02,10))$ .
2.  $n$  Best: Number of recordings (out of 134) for which this polynomial is the best
3.  $d$ : Number of recordings for which the accuracy of the specified polynomial falls within  $0.1^\circ/0.2^\circ/0.5^\circ$  of the accuracy of the best polynomial for the participant.

Table 4.1 No correction

Rank	CoefsX	CoefsY	Best	Worst	Avg	SD	n Best	$d \leq 0.1^\circ$	$d \leq 0.2^\circ$	$d \leq 0.5^\circ$
	Best per participant	Best per participant	0.39	1.74	0.672	0.229				
1	01,02,03,10,11,20,21,30,31	01,02,10,11,20,21	0.39	1.75	0.687	0.245	23	95	120	128
<b>2</b>	<b>01,10,11,20,21,30,31</b>	<b>01,02,10,11,20,21</b>	<b>0.40</b>	<b>1.75</b>	<b>0.690</b>	<b>0.249</b>	<b>14</b>	<b>93</b>	<b>120</b>	<b>128</b>
3	01,02,03,10,11,20,21,30,31	01,02,10,11,20,22	0.40	1.78	0.691	0.243	8	92	119	128
4	01,10,11,20,21,30,31	01,02,10,11,20,22	0.40	1.78	0.694	0.248	9	89	119	128
5	01,02,03,10,11,20,21,30,31	01,02,10,20,21	0.41	1.74	0.699	0.244	9	85	120	128

Table 4.2 Real-time correction

Rank	CoefsX	CoefsY	Best	Worst	Avg	SD	n Best	$d \leq 0.1^\circ$	$d \leq 0.2^\circ$	$d \leq 0.5^\circ$
	Best per participant	Best per participant	0.22	0.83	0.365	0.112				
1	01,10,11	02,10	0.23	0.90	0.414	0.142	5	94	120	131
2	01,10,11	02,10,12	0.24	0.89	0.415	0.141	3	92	119	131
3	01,10,11,30	02,10	0.24	0.90	0.415	0.142	4	88	120	132
4	10,11	02,10	0.23	0.92	0.416	0.141	6	88	119	132
5	02,10,11,30	02,10	0.24	0.92	0.416	0.142	4	86	121	132
6	01,10,11,22	02,10	0.23	0.98	0.416	0.143	4	88	119	131
...										
<b>471</b>	<b>01,10,11,20,21,30,31</b>	<b>01,02,10,11,20,21</b>	<b>0.22</b>	<b>1.30</b>	<b>0.439</b>	<b>0.181</b>	<b>2</b>	<b>77</b>	<b>114</b>	<b>130</b>

Table 4.3 Subset of calibration points

Rank	CoefsX	CoefsY	Best	Worst	Avg	SD	n Best	$d \leq 0.1^\circ$	$d \leq 0.2^\circ$	$d \leq 0.5^\circ$
	Best per participant	Best per participant	0.24	1.44	0.453	0.170				
1	01,02,10,11,20,22,30	01,02,10,11,20,22	0.25	0.55	0.481	0.212	15	101	122	128
2	01,02,10,11,20,22,30	01,02,10,11,20,21	0.25	0.52	0.486	0.210	7	99	121	129
3	01,10,11,20,21,30,31	01,02,10,11,20,22	0.25	0.48	0.492	0.216	6	94	120	128
4	01,02,10,11,20,22,30	01,02,11,20,22	0.28	1.50	0.493	0.202	2	90	119	129
5	01,02,10,11,20,22,30	01,02,10,11,20	0.28	1.60	0.495	0.217	3	90	120	128
<b>6</b>	<b>01,10,11,20,21,30,31</b>	<b>01,02,10,11,20,21</b>	<b>0.25</b>	<b>1.45</b>	<b>0.496</b>	<b>0.213</b>	<b>6</b>	<b>92</b>	<b>120</b>	<b>128</b>

Table 4.4 Subset of calibration points with real-time correction

Rank	CoefsX	CoefsY	Best	Worst	Avg	SD	n Best	$d \leq 0.1^\circ$	$d \leq 0.2^\circ$	$d \leq 0.5^\circ$
	Best per participant	Best per participant	0.16	0.76	0.320	0.104				
1	01,10,11,20	01,02,10,21	0.21	0.99	0.366	0.139	2	98	120	133
2	01,10,11,30	01,02,10,21	0.20	1.01	0.366	0.141	0	101	120	132
3	01,10,11,22,30	01,02,10,21	0.19	1.08	0.367	0.148	0	101	119	132
4	01,10,11,30	01,02,10,22	0.20	0.93	0.369	0.143	0	101	119	131
5	01,10,11,22,30	01,02,10,22	0.19	1.10	0.369	0.153	2	99	117	130
...										
<b>742</b>	<b>01,10,11,20,21,30,31</b>	<b>01,02,10,11,20,21</b>	<b>0.24</b>	<b>2.03</b>	<b>1.069</b>	<b>0.373</b>	<b>0</b>	<b>2</b>	<b>4</b>	<b>31</b>

One should observe that the average of the SAICSIT polynomials without correction ( $0.69^\circ$ , cf Table 3.1) is worse than the reported accuracy for the same polynomial set in Blignaut (2013) (avg =  $0.46^\circ$ , SD =  $0.17^\circ$ ). It needs to be noted, however, that a different sample of participants and a different grid ( $6 \times 4$ ) were used in Blignaut (2013) – although this is unlikely to have made a big difference as the number of points (23 in this experiment) are about the same.

The most likely cause for the worse results in this study is that two infrared illuminators were used as opposed to a single illuminator in the 2013 study. The higher illumination allowed for an increased framerate (200 Hz vs about 60 Hz) that will allow researchers to do studies that would otherwise not be possible. However, one might argue that the higher framerate comes at a price of less accuracy.

This leads one to conclude that the identification of a polynomial set can be dependent on the hardware configuration and it once again underlines that the accuracy of an eye tracker cannot be absolute and needs to be reported for every study.

Having said this, one should acknowledge that although  $0.46^\circ$  is less than the magical figure of  $0.5^\circ$ ,  $0.69^\circ$  is not bad either for a low cost eye tracker. The result above should rather serve as confirmation that the SAICSIT polynomial set is working well under different conditions, rather than emphasising the somewhat weaker results in this study.

### The effect of polynomial set on mapping error

Taking all the above into consideration, the basic procedure followed by real-time correction (RTC), is the safest approach as any polynomial set is guaranteed to give good results (Table 3). The LC+RTC technique can give better results – provided that care is taken in selecting the polynomial set.

The procedure described in Step 1 of Listing 1 allows the selection of the best polynomial set per participant and correction technique (BPPP). Table 4 above shows the accuracies for the overall best performing polynomial sets per correction technique (BP) along with the accuracy achieved by applying separate polynomial sets to different participants (BPPP). Table 5 shows the results of a repeated measures analysis of variance for the effect of polynomial set (SAICSIT, BP and BPPP) while controlling for correction technique. Tukey’s post-hoc test for the honest significant difference (HSD) between pairs of values is also indicated.

Although the variations in effect sizes (absolute differences between pairs of means) are small, the differences were significant in all cases ( $\alpha = .01$ ) because of the consistent direction of the improvements for all participants. Furthermore, the overall improvement, from what was previously available (SAICSIT with no corrections,  $0.69^\circ$ ) to what can be achieved with the proposed approach of finding the best polynomial per participant and applying correction techniques ( $0.32^\circ$ ), is substantial.

Table 5  
Significance of differences between polynomial sets while controlling for correction technique (\* $\alpha = .05$ ; \*\* $\alpha = .01$ ; \*\*\*  $\alpha = .001$ )

Correction	Polynomial set	Mapping error (degrees)					Significance		Tukey's HSD	
		N	Min	Max	Avg	SD	F(2,266)	p	BP	BPPP
None	SAICSIT	134	0.40	1.75	0.69	0.25	24.325	***	-	***
	BP	134	0.39	1.75	0.69	0.25			***	
	BPPP	134	0.39	1.74	0.67	0.23			***	
RTC	SAICSIT	134	0.22	1.30	0.44	0.18	55.956	***	**	***
	BP	134	0.23	0.90	0.41	0.14			***	
	BPPP	134	0.22	0.83	0.36	0.11			***	
LC	SAICSIT	134	0.25	1.45	0.50	0.21	36.668	***	**	***
	BP	134	0.25	0.55	0.48	0.21			***	
	BPPP	134	0.24	1.44	0.45	0.17			***	
LC+RTC	SAICSIT	134	0.24	2.03	1.07	0.37	696.71	***	***	***
	BP	134	0.21	0.99	0.37	0.14			-	
	BPPP	134	0.16	0.76	0.32	0.10			***	

*The effect of correction technique on mapping error*

Table 6 shows the results of repeated measures analysis of variance for the effect of correction technique on mapping error while controlling for the polynomial set. Tukey's

post-hoc test for the honest significant difference (HSD) between pairs of values is also indicated. All techniques proved to be significantly ( $\alpha = .001$ ) better than the basic case, except when limiting the calibration points are combined with real-time correction for the SAICSIT polynomials, which is significantly worse than the basic case.

Table 6  
Significance of differences in mapping error between correction techniques while controlling for polynomial set (\* $\alpha = .05$ ; \*\* $\alpha = .01$ ; \*\*\*  $\alpha = .001$ )

Polynomial set	Correction	Mapping error (degrees)					Significance		Tukey's HSD		
		N	Min	Max	Avg	SD	F(1,133)	p	RTC	LC	LC+RTC
SAICSIT	None	134	0.40	1.75	0.69	0.25	388.87	***	***	***	***
	RTC	134	0.22	1.30	0.44	0.18			*	***	
	LC	134	0.25	1.45	0.50	0.21			***		
	LC+RTC	134	0.24	2.03	1.07	0.37			***		
BP	None	134	0.39	1.75	0.69	0.25	271.85	***	***	***	***
	RTC	134	0.23	0.90	0.414	0.14			***		
	LC	134	0.25	0.55	0.48	0.21			***		
	LC+RTC	134	0.21	0.99	0.37	0.14			***		
BPPP	None	134	0.39	1.74	0.67	0.23	377.90	***	***	***	***
	RTC	134	0.22	0.83	0.36	0.11			***		
	LC	134	0.24	1.44	0.45	0.17			***		
	LC+RTC	134	0.16	0.76	0.32	0.10			***		

**Summary**

This study is aimed at the improvement of low-cost eye tracking to such an extent that it can be used in studies where high accuracy ( $\leq 0.5^\circ$ ) is required at mid to high framerates ( $\geq 200$  Hz). A camera was used that is capable of attaining 200 Hz provided that the IR illuminators provide enough light and that the recording window is small

enough. A technique was used where the recording window, is just big enough to capture both eyes at once and follows the user in a reasonably sized (400 mm  $\times$  100 mm) head box.

It was further reasoned that the polynomial expressions that are normally used for remote, video-based, low cost systems are not always ideal to accommodate individual differences in eye cleft, position of the eye in the socket,



corneal bulge, astigmatism, etc. This paper proposed a procedure to identify a set of polynomial expressions that will provide the best possible accuracy for a specific individual. It is also proposed that regression coefficients are recalculated in real-time, based on a subset of calibration points in the region of the current gaze. Finally, it was proposed that a real-time correction is applied based on the offsets from calibration targets that are close to the estimated point of regard.

This study follows on Blignaut (2013) and Blignaut (2014) respectively as far as the polynomial expressions and real-time corrections are concerned. A calibration procedure of 45 dots, appearing in quick succession in random order, is proposed. Twenty-three of the dots are used as calibration targets and the complete list of dots are used to select the best possible regression polynomial and to validate the accuracy of the regression.

A system was developed as part of a larger project to implement the proposed procedure and capture calibration and real-time gaze data. One hundred and thirty-four (134) recordings of 42 participants were used to examine the effects of polynomial set and the two proposed correction techniques. It was found that if no correction is applied, the choice of polynomial is critically important to get an accuracy that is just acceptable. Previously identified polynomial sets, e.g. the SAICSIT set identified by Blignaut (2013), were confirmed to provide good results in the absence of any correction procedure.

By applying real-time correction (RTC), the accuracy of any given polynomial improves while the choice of polynomial becomes less critical. This means that it is no longer necessary to use long polynomials with many higher order terms, which are dependent on a large number of calibration points. For the data captured in this study, the following very basic set of polynomials returned an average error (over calibration plane and participants) of  $0.42^\circ$  and an error of within  $0.2^\circ$  of the error that can be achieved with the best polynomial set for 119/134 (88.9%) participants:

$$\begin{aligned} X &= a_0 + a_1x + a_3xy \\ Y &= b_0 + b_1x + b_2y^2 \end{aligned} \quad (10)$$

It needs to be noted, though, that a large set of calibration points would still be needed to have four calibration points within  $2^\circ$  of any first estimate of point of regard.

Using a subset of calibration points (LC) resulted in better accuracy than when no correction was done, but it

was not better than the RTC technique. When LC is combined with RTC, very good accuracies ( $\leq 0.4^\circ$ ) can be attained – provided that the correct polynomial sets are identified. This critical aspect can be overcome with the procedure to identify the best polynomial set per participant and correction technique. Using this combination of BPPP and LC+RTC, an average error over 134 participant recordings of  $0.32^\circ$  (SD =  $0.10^\circ$ ) was recorded, which is significantly better than could be recorded by any specific polynomial or correction technique.

The results of this study can have important applications for the future of eye tracking. By applying the techniques of BPPP along with LC+RTC, systems that are accurate and fast enough to do reading research or other studies where high accuracy is expected at framerates in excess of 200 Hz, can be built for less than USD 1,000.

The large number of polynomials might be regarded as a drawback, but a procedure of 30 – 45 seconds should not stand in the way of better gaze data. Furthermore, if the results of the calibration process are visualised as in Figure 6, an experimenter would know when to recalibrate if it is observed that a participant did not concentrate for one or more targets. Using the eye video and eye images, the experimenter would also be able to do troubleshooting and recalibrate when mascara or glasses seems to be problematic.

## References

- Abe, K., Ohi, S., & Ohshima, M. (2007). An eye-gaze input system using information on eye movement history. In C. Stephanides (Ed.), *Universal access in HCI, Part II. HCI2007, LNCS 4555*, (pp 721-729). Berlin Heidelberg: Springer-Verlag.
- Blignaut, P.J. (2013). A new mapping function to improve the accuracy of a video-based eye tracker. *Proceedings of the South African Institute of Computer Scientists and Information Technologists (SAICSIT)*, 8-9 October 2013, East London, South Africa.
- Blignaut, P.J., & Wium, D.J. (2013). The effect of mapping function on the accuracy of a video-based eye tracker. *Proceedings of the Eye Tracking South Africa Conference (ETSA)*, 29-31 August 2013, Cape Town, South Africa.
- Blignaut, P.J. (2014). Mapping the Pupil-Glint vector to gaze coordinates in a simple video-based eye tracker. *Journal of Eye Movement Research*, 7(1):4, 1-11.

- Blignaut, P.J., Holmqvist, K., Nyström, M., Dewhurst, R. (2014). Improving the accuracy of video-based eye tracking in real-time through post-calibration regression. In Horsley, M., Eliot, M., Knight, B., Reilly, R. (Eds.), *Current Trends in Eye Tracking Research*, Springer.
- Borah, J. (1998). Technology and application of gaze based control. *RTO Lecture series on Alternative Control Technologies*, 7-8 October 1998, Brétigny, France and 14-15 October 1998, Ohio, USA.
- Brolly, X.L.C. & Mulligan, J.B. (2004). Implicit calibration of a remote gaze tracker. *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04)*, 8, 134-146.
- Buscher, G., Cuttrel, E. & Morris, M.R. (2009). What do you see when you're surfing? Using eye tracking to predict salient regions of web pages. *CHI2009*, 4-9 April 2009, Boston, Massachusetts.
- Cerrolaza, J., & Villanueva, A. (2008). Taxonomic study of polynomial regression applied to the calibration of video-oculographic systems. *Proceedings of the 2008 Symposium on Eye Tracking Research and Applications (ETRA)*, 26-28 March 2008, Savannah, Georgia.
- Cerrolaza, J.J., Villanueva, A., Villanueva, M., & Cabeza, R. (2012). Error characterization and compensation in eye tracking systems. *Proceedings of the 2012 Symposium on Eye Tracking Research and Applications (ETRA)*, 28-30 March 2012, Santa Barbara, California.
- Crane, H.D., & Steele, C.M. (1985). Generation-V dual-Purkinje-image eye tracker. *Applied Optics*, 24, (pp 527-537).
- Draper, N.R., & Smith, H. (1981). *Applied Regression Analysis*. John Wiley and Sons.
- Hansen, D.W., & Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3), (pp 478-500).
- Hennessey, C., Noureddin, B., & Lawrence, P. (2008). Fixation precision in high-speed noncontact eye-gaze tracking. *IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics*, 38(2), April 2008.
- Holmqvist, M., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. and Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.
- Hoorman, J. (2008). The effect of calibration errors on the accuracy of eye movement recordings. *Journal of Eye Movement Research*, 1(2):3, 1-7.
- Hornof, A.J. & Halverson, T. (2002). Cleaning up systematic error in eye tracking data by using required fixation locations. *Behavior Research Methods, Instruments & Computers*, 34(4), 592-604.
- Hyrskykari, A. (2006). Utilizing eye movements: Overcoming inaccuracy while tracking the focus of attention during reading. *Computers in Human Behavior*, 22(4), 657-671.
- Kliegl, R. & Olson, R.K. (1981). Reduction and calibration of eye monitor data. *Behavior Research Methods and Instrumentation*, 13, 107-111.
- McConkie, G.W. (1981). Evaluating and reporting data quality in eye movement research. *Behavior Research Methods & Instrumentation*, 13(2), 97-106.
- Mitsugami, I., Ukita, N., & Kidode, M. (2003). Estimation of 3D gaze position using view lines. *Proceedings of the 12th International Conference on Image Analysis and Processing (ICIAP'03)*, IEEE.
- Morimoto, C.H., Mimica, R.M. (2005). Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 98(1), 4-24.
- Rayner, K., Pollatsek, A., Drieghe, D., Slattery, T., & Reichle, E. (2007). Tracking the mind during reading via eye movements: Comments on Kliegl, Nuthmann, and Engbert (2006). *Journal of Experimental Psychology: General*, 136(3), 520-529.
- Sesma-Sanchez, L., Zhang, Y., Bulling, A., Gellersen, H. (2016). Gaussian processes as an alternative to polynomial gaze estimation functions. *Proceedings of the 2016 Symposium on Eye Tracking Research and Applications (ETRA)*, 14-17 March 2016, Charleston, SC.
- Sheena, D. & Borah, J. (1981). Compensation for some second order effects to improve eye position measurements. In D.F. Fisher, R.A. Monty & J.W. Senders (Eds.), *Eye Movements: Cognition and Visual Perception*, Hillsdale: Lawrence Erlbaum Associates.
- Tobii Technology. (2010). *Product description, Tobii T/X series eye trackers*, rev. 2.1, June 2010. Tobii Technology AB.
- Van Walree, P. 2015. *Distortion*. Available online <http://toothwalker.org/optics/distortion.html>. Accessed 21 March 2016.
- Zhang, Y. & Hornof, A. J. (2011). Mode of disparities error correction of eye tracking data. *Behavior Research Methods*, 43(3): 834-842.
- Zhu, Z., & Ji, Q. (2005). Eye gaze tracking under natural head movements. *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*