

The effect of calibration errors on the accuracy of eye movement recordings

Jörg Hoormann
Institut für Arbeitsphysiologie

Stephanie Jainta
Institut für Arbeitsphysiologie

Wolfgang Jaschinski
Institut für Arbeitsphysiologie

For calibrating eye movement recordings, a regression between spatially defined calibration points and corresponding measured raw data is performed. Based on this regression, a confidence interval (CI) of the actually measured eye position can be calculated in order to quantify the measurement error introduced by inaccurate calibration coefficients. For calculating this CI, a standard deviation (SD) - depending on the calibration quality and the design of the calibration procedure - is needed.

Examples of binocular recordings with separate monocular calibrations illustrate that the SD is almost independent of the number and spatial separation between the calibration points – even though the later was expected from theoretical simulation. Our simulations and recordings demonstrate that the SD depends critically on residuals at certain calibration points, thus robust regressions are suggested.

Keywords: eye-movement, calibration, confidence interval, robust regression,

Introduction

Eye movements can be measured with different technical systems (for a review see Collewijn, 1999) that all need a calibration procedure to provide the angular position of the eyes. Only the search-coil technique can be calibrated objectively (i. e. physically), all other techniques (e. g. limbus tracking, Purkinje image tracking, video systems) require a subjective calibration, i. e., the recording during steady fixation of single targets at known angular positions. Unusually, a linear regression is calculated between spatially defined calibration points x_i (deg) and corresponding raw data y_i (arbitrary units), measured during fixation of calibration points. Least square (LS) fits determine the coefficients b_0 and b_1 , i. e. the y-intercept and the slope, respectively:

$$y_i = b_0 + b_1 * x_i \quad (1)$$

For any measured raw data Y_m (arbitrary units) within the calibration range, the corresponding eye position X_m (deg) can be calculated by:

$$X_m = \frac{Y_m - b_0}{b_1} \quad (2)$$

Figure 1 shows examples of typical calibration curves for the two eyes (relative to the x-axis, the position of the two eyes is indicated at the bottom). Both curves have been recorded separately for each eye with 7 calibration targets that have been presented monocularly.

Usually, the measured data points do not lie on a perfect line. Consequently, the measured eye position is subject to an uncertainty that can be described by a standard deviation (SD) given by the following equation (Fogt & Jones, 1998a; Fogt & Jones, 1998b; Neter, Wasserman & Kutner, 1990):

$$SD^2(x_m) = \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{(n-2) * b_1^2} * \left[1 + \frac{1}{n} + (x_m - \bar{x})^2 / \sum_{i=1}^n (x_i - \bar{x})^2 \right] \quad (3)$$

with y_i : representing the measurement value
 \hat{y} : representing the calculated value
 n : number of calibration points

Considering the mathematical characteristics, the standard deviation SD depends on the following 4 aspects (see equation 3):

- (1) The actual angular position of the eye (relative to the calibration centre) is important: the more the eye position (x_m) deviates from central fixation (\bar{x}), the larger the SD.
- (2) Increasing the number of calibration points n decreases the SD - at least at points far from the central fixation.
- (3) The separation between the calibration points ($x_i - \bar{x}$) is contributing to the SD, depending on the eccentricity.
- (4) Generally, outliers contribute to the SD, because of the squared influence of the residuals on the mean square error.

For accurate eye movement recordings, one wishes to indicate a confidence interval (CI) for all angular positions that occur in a particular experiment. In general, it is desirable to calibrate the eye movements in a way that a small CI (or small SD as described so far) may result, reflecting minor uncertainties attributed to the calibration process. Obviously, the SD will be small if the mean square error (MSE; see first part of equation 3) during the calibration is small. But, the SD also depends on design parameters of the calibration procedure itself, i.e. the number of calibration points and their separation – as mentioned above.

Considering the literature of one-dimensional, horizontal eye movements, sometimes only 2 calibration points were used (see for example, Semmlow & Yuan, 2002; Semmlow, Chen, Pedrone & Alvarez, 2008); these authors argue that a straight line can be determined with two points assuming linearity of the recording system. This requires a strongly reliable measure of the two points for a small SD. Other strategies include more calibration points to reach a good approximation of the calibration function. This procedure should result in small SD, but is time consuming. Thus, the question arises, how the calibration procedure should be designed to achieve small SD within an appropriate period of time.

We investigated the effect of the number and the angular separation of the calibration points on the SD in two ways: 1. We performed simulations according to equation (3) and 2. we compared the simulations with empirical data measured under experimental variations of the calibration procedure. This study was made to show that the calculation of SD may be a useful procedure to specify the quality eye movement recordings concerning the calibration; this is still uncommon, despite the previous contributions of Fogt and Jones (1998 a and b).

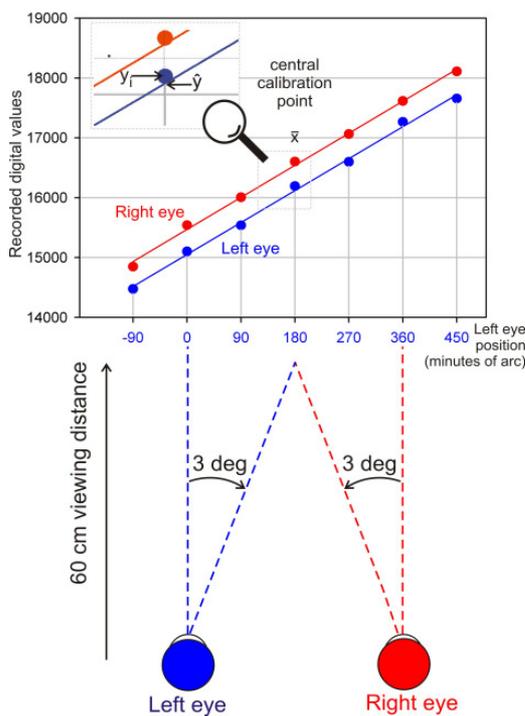


Figure 1:

Example of typical calibration curves for the two eyes. Relative to the x-axis, the position of the two eyes is indicated at the bottom.

Simulation of the number and separation of calibration points

The standard deviation SD attributed to each measured eye position is given by equation (3). Since this equation is rather complex and does not allow for an immediate overview of the specific effects, we performed the following simulations to illustrate the quantitative influence of the number of calibration points and their angular separation

Method

We used calibration curves, which had a priori constant $R^2 = 0.975$, and calculated the SD for a range of 720 min arc (which is equivalent to 12 deg) as a function of the number of calibration points (see equation 3). The central calibration point was at 180 min arc, thus 3 deg to the right of the straight ahead position of the left eye (see Figure 1). In a first run, the calibration points had fixed separations of 90 min arc and we varied the number of points from 3, 5, 7 to 9. In a second run, we varied the separation amounting to 90, 180, 270, and 360 min arc in a constant set of 3 calibration points.

Results

Considering a constant separation, Figure 2 shows the typical hyperbola curves for the SD for all 4 number variations. For a small region of x_m near the central calibration point (around 180 min arc), the simulation including 3 calibration points yield a flat SD curve. Thus, for small fixational eye movements close to central fixation, three calibration points are sufficient to reach small SD. In this case, however, SD strongly increases with eccentricity. For a wider range of x_m , i.e. across the whole range of 12 deg, five or seven calibration points result in a flat curve. In this case, for eye movements in a larger angular range, more calibration points across the measurement range keep SD relatively small and constant across the whole range.

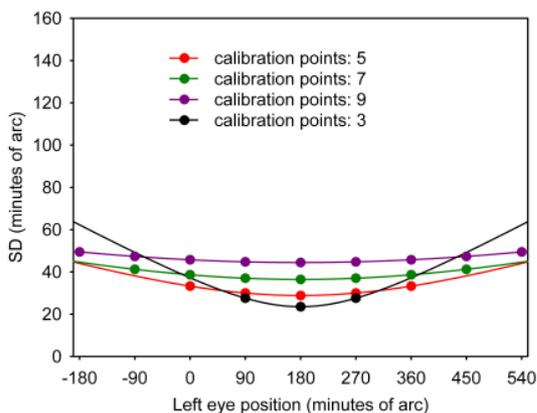


Figure 2:

Typical, simulated hyperbola curves for the SD containing 3, 5, 7 or 9 calibration points with constant separation of 90 min arc.

For a variable separation between 3 calibration points, Figure 3 shows the typical hyperbola curves for the SD. As expected, increasing the separation flattens the curves; for our simulation context, the smallest separation produces the smallest SD with greatest differences being allocated near the central fixation. With bigger separation, the overall SD increases and if only three calibration points are distributed across a wide range, the SD becomes large and increases with eccentricity.

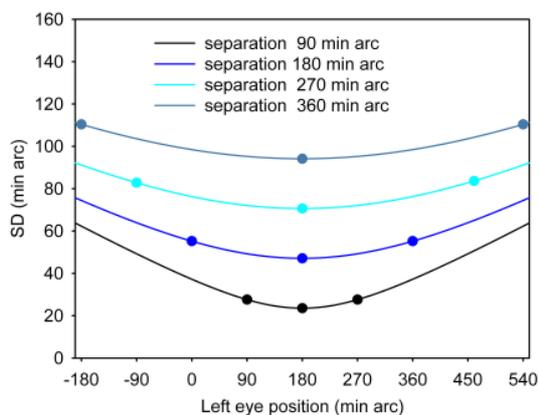


Figure 3:

Typical, simulated hyperbola curves for the SD for a variable separation between 3 calibration points.

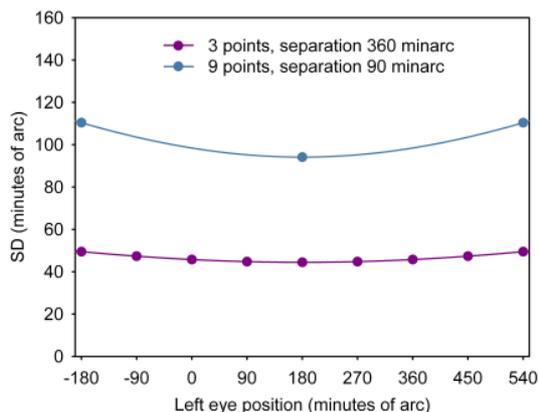


Figure 4:

Two, theoretically extreme outcomes of the combined variations of point numbers and separation: at the same calibration range of 12 deg, more calibration points at smaller separations result in a reduction of SD following the simulation.

If one is interested in the combined effect of number and separation of calibration points, Figure 4 shows the two extreme simulated outcomes: at a constant measurement range (12 deg in our example) more calibration points (9 versus 3) at smaller separations (90 min arc versus 360 min arc) result in a considerable reduction of SD.

In sum, our simulations suggest (1) that for eye movements in a small angular range, three calibration points seem sufficient to reach small standard deviations and (2) that with large eye movements, more calibration points at smaller separations are required to reduce the standard deviation.

Experimental variation of the number and separation of calibration points

For comparison with the simulations reported above, we investigated empirically the effect on SDs based on measured calibration where the number and separation of the calibration points was varied.

Method

We used a mirror stereoscope (Howard, 2002) with two mirrors at right angle and two VDU screens (CRT Sony F500 T9). In order to minimize head movements, we used a chin and forehead rest including a narrow temporal rest, which was adjusted to the size of subject's head. The eye movements were recorded with the video-based EyeLink II[®], which tracks the centre of the pupil by an algorithm similar to a centroid calculation. The EyeLink II system has a linear horizontal tracking range of $\pm 30^\circ$ and a spatial resolution of 0.6 min arc (more details provided by SR Research Ltd, Osgoode ON, Canada). The EyeLink cameras were attached to the head rest. We did neither use the head tracking system, nor the calibration procedure of the original EyeLink II system, rather we recorded the raw data with a sampling rate of 500 Hz and used the following calibration procedure.

Subjects were requested to carefully fixate calibration targets that appeared (for 1400 ms) randomly at different screen positions with 100 ms temporal gaps; monocular

presentations to the right and left eye were randomly interleaved. Two of these calibration series were repeated directly one after the other and results were averaged. In order to draw attention to the calibration points and to facilitate exact fixation, the diameter of the calibration spot initially subtended 1 deg and shrank immediately during 1000 ms to a remaining cross of 8.1 x 8.1 min arc (stroke width: 2.7 min arc); the remaining cross was visible for additional 400 ms during which calibration data were stored. The whole calibration range subtended 720 min arc (12 deg) at 60 cm viewing distance.

This procedure was chosen since it represents a fixation task that is not difficult to perform for the subject: it includes a very small final target of only 8 min arc which requires central foveal fixation and thus stimulated an eye position corresponding to a very precise spatial location as required for calibration. But this small target was only presented for a short 400 ms interval; for comparison, fixation durations of about 220 ms are typical during reading. Longer periods of steady fixation would be rather unnatural and give rise to drifts and micro-saccades. The saccades from one calibration point to the next were stimulated by targets that initially had a large diameter of 1 deg in order to be easily perceived in peripheral vision and to draw attention to the next calibration point; the latter feature resembles the one used in Tobii[®] eye movement recording systems.

Generally, eye movement recordings and calibrations are more accurate and stable, if a bite-bar is used. However, even though a bite-bar has not been used for convenience of the subjects in the present study, the resulting standard deviations were in the same order of magnitude as in the studies of Fogt & Jones (1998 a and b) using a bite-bar and a search-coil recording system. Probably, our short recording period of less than 45 seconds had reduced the risk of possible artifacts due to small head movements.

To test the calibration procedure, we had calibration runs for each eye in a sample of 16 subjects: in a first run, we used separate calibrations with 3, 5 or 7 calibration points with constant inter-point separations of 90 min arc. Additionally, we had a second run containing 3 calibration points with inter-point separations of 60, 180, and 360 min arc.

Results

First of all, in our experiments we reached average standard deviations (SD) of less than 20 min arc. Varying the number of calibration points from 3 to 7 points resulted in mean SD of the two eyes as shown by the distributions in Figure 5. No significant difference between the average SD for the 3, 5 or 7 point calibration was observed. Nevertheless, as seen in Figure 5, using more calibration points reduces the appearance of large outliers.

In a similar way, the average SD was not significantly different when comparing the separations of 60, 180, and 360 min arc using 3 calibration points (not shown graphically).

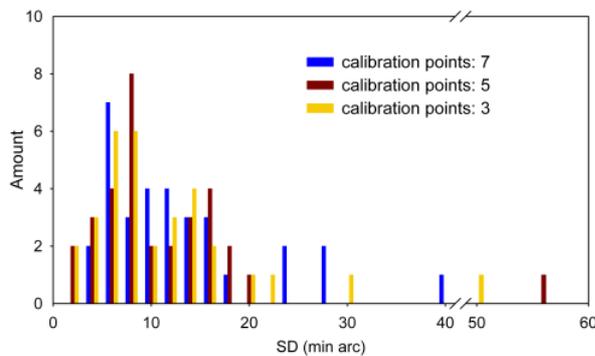


Figure 5:

SD distribution for empirical calibrations with 3, 5 and 7 calibration points at constant inter-point separations of 90 min arc.

Discussion

The accuracy of a measured eye position can be described by a standard deviation that depends on the quality of the measurement of calibration (i.e., the mean square error of the calibration regression) and the design of the calibration procedure, i.e. the number of calibration points and the separations between them. Our simulation of equation (3) suggested that the SD depends on the number and the separation in a way that we described in Figures 2, 3 and 4. Our experimental data, however, showed only small insignificant effects on SD, which e.g. was 11.8, 10.5, and 10.2 min arc with 3, 5, and 7

calibration points, respectively. This suggests that one should be careful to use equation (3) and the resulting simulation as a guideline to design the calibration procedure, since the systematic variation of SD could not be validated by our empirical data set. The most convincing reason for this discrepancy between simulation and experimental data is the following: for our simulations we kept the R^2 of the calibration regression per definition constant (at a value of 0.975). Such an assumption is necessary, in order to make the simulations comparable. However, for the empirical data the assumption was not true; we calculated the R^2 for our last sample of 32 calibrations (16 subjects x 2 eyes) and observed a decrease of R^2 with the reduction of the number of calibration points (see Figure 6).

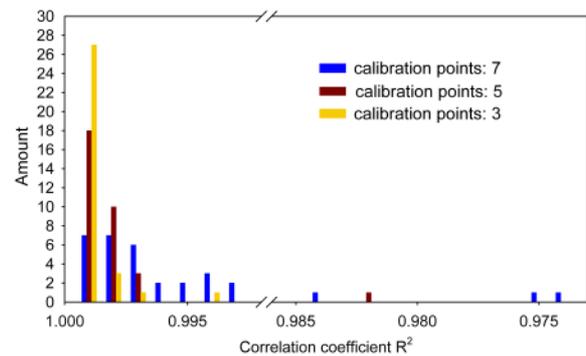


Figure 6:

R^2 distribution for empirical calibrations with 3, 5 and 7 calibration points at constant inter-point separations of 90 min arc.

The larger the number of calibration points the smaller will be the effect of single outliers on the standard deviation; more specifically, it can be seen from equation (3) that an increase of the number of calibration points n from 3 to 5 reduces the SD by a factor of three in spite of the squared influence of individual residues of single outliers.

In sum, even though the simulation shows dependencies of the SD on the design of the calibration procedure with constant R^2 , the empirical SDs are supposed to remain stable. Nevertheless, with large eye movements, a three point calibration results in large SD for eccentric eye positions and more calibration points are

required to reduce SDs. Thus, a 5-point calibration is still a good choice, since the regression is less effected by extreme outliers and it is possible to calculate a robust regression, which can reduce the mean square error (see Appendix).

Although the present study was made with horizontal calibration positions, the principle results can be transferred to the vertical direction. The next step of research could be to calculate the regression coefficients (horizontal and vertical) in a multivariate design and to estimate confidence ellipses instead of confidence intervals for eye positions.

In conclusion, for quantifying the uncertainty of the measured eye position due to calibration errors we suggest that equation (3) is a useful tool for the calculation of the standard deviation based on the actually recorded calibration and the chosen positions of calibration points.

The practical procedure for designing the calibration might be to define the calibration range to cover the angular dimensions of the eye movements to be recorded. The number of calibration points is equally spaced across the calibration range. Although the number of calibration points and their separation did not have much effect on the standard deviation, the effect of outliers can be reduced by increasing the number of calibration points (particularly if robust regression analysis is used).

Appendix

Simple regression versus robust regression: Our simulations and recordings demonstrate that SD depends critically on the amount of the residuals. Because of the strong influence of single large residuals, we suggest to perform a robust regression, if possible. Robust regression analyses have previously used in other eye movement studies (e.g., Ruetsche, Baumann, Jiang & Mojon, 2003; Jaschinski, Jainta & Schürer, 2006; Kloke, Jaschinski & Rinckenauer, 2007).

Figure A1 shows a typical calibration graph with a simple least square (LS) regression and a robust regression with a reweighed LS algorithm (Welsch 2.0) (Draper & Smith, 1998). Although the coefficients of the regression are similar, the mean square error (MSE) is reduced by nearly 30 % (from 339 to 245) by using the robust regression.

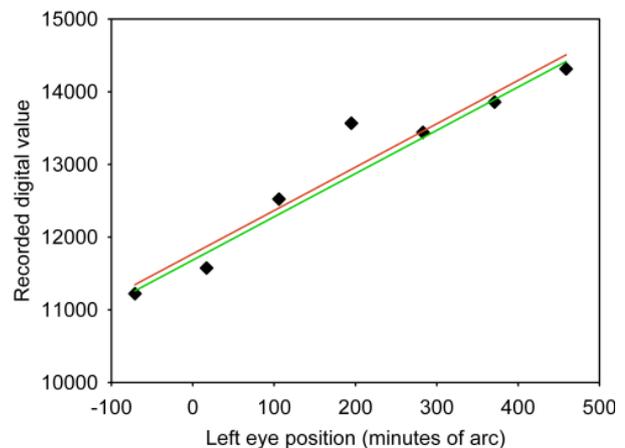


Figure A1:

Examples of calibration regressions: LS algorithm versus robust regression. The red regression line reflects the LS and the green line the robust regression.

References

- Collewijn H. (1999) Eye movement recordings. In: Carpenter, R.H.S. and Robson., J.G.: *Vision Research: An practical guide to laboratory methods*. Oxford University Press, Oxford.
- Draper, N.R., & Smith, H. (1998). *Applied Regression Analysis*, 3rd ED., Wiley: NY.
- Fogt, N., & Jones, R. (1998a). Comparison of fixation disparities obtained by objective and subjective methods. *Vision Research*, 38, 411-421.
- Fogt, N., & Jones, R. (1998b). The effect of forced vergence on retinal correspondence. *Vision Research*, 38, 2711-2719.
- Howard, I.P. (2002). *Seeing in Depth; Basic Mechanisms*. 1 (p. 392ff). Toronto, Canada: Porteous, I.
- Jaschinski, W., Jainta, S., & Schürer, M. (2006). Capture of visual direction in dynamic vergence is reduced with flashed monocular lines. *Vision Research*, 46, 2608-2614.

Kloke, W.B., Jaschinski, W., & Rinkeauer, G. (2007).
Beidäugige Koordination in Sehbedingungen von
Head-up-Displays in Kraftfahrzeugen. *Zeitschrift für
Arbeitswissenschaft* 61, 225-233.

Neter, J., Wasserman, W., & Kutner, M.H. (1990).
Qualitative independent variables. *In: Applied linear
statistical models. Burr Ridge: Irwin*, 349-385.

Ruetsche, A., Baumann, A., Jiang, X., & Mojon, D.S.
(2003). Automated analysis of eye tracking
movements. *Ophthalmologica*, 217, 320-324.

Semmlow, J.L., & Yuan, W. (2002). Adaptive
modification of disparity vergence components: an
independent component analysis study. *Investigative
Ophthalmology & Visual Science*, 43, 2189-2195.

Semmlow, J.L., Chen, Y., Pedrone, C., & Alvarez, T.
(2008). Saccadic Behavior during the Response to
Pure Vergence Stimuli I: General Properties. *Journal
of Eye Movement Research*, 1(2), 1-11.