Methods for objective measurement of fixation disparity.

Corresponding author:
Wolfgang Jaschinski
Leibniz Research Centre for Working Environment and Human Factors
Ardeystr. 67
D 44139 Dortmund
Germany
Tel + 49 231 1084 264
Fax + 49 231 1084 401
e-mail: jaschinski@ifado.de

In the present study, the EyeLink II video eyetracker (SR Research Ltd., Osgoode, ON, Canada) was used which, in common with other video systems, has several advantages: (1) it is not invasive compared to, for example, the search coil system, (2) it is mechanically easy to install, and (3) binocular recordings are possible without the need to double the system. These features allow video systems to be used not only in sophisticated research laboratories, but potentially also in applied optometry, which is one aim of the present research. However, video eyetrackers might have limitations when investigating fixation disparity which is typically smaller than 1 degree of visual angle. Further, it is a general concern that without a bite bar no stable high resolution eye movement measures can be obtained. We did not use a bite bar however since we wished to have a convenient test procedure in which many subjects would be willing to participate: we aim to test large numbers of subjects in order to investigate individual differences.

In view of these concerns, we made additional control experiments and data analyses in order to describe the resolution, validity, and stability of data obtained with the EyeLink II system and our test procedures. Further, we describe the dynamic vergence processes in the early phase of fixations, and discuss the limitation that video eyetrackers cannot differentiate between ocular rotations and displacements.

1. Resolution

The theoretical noise limited resolution of the EyeLink II equipment is specified as 0.6 min arc for the dark pupil system that we used. But, the practical question remains about the smallest change in eye position that can realistically be measured under the actual experimental conditions. Therefore, we performed a control experiment (Appendix 1 in Jainta et al., 2009), where subjects performed saccades between targets separated by 3.0 deg plus an additional offset separation of 0.0, 3.2, 6.5, 9.7, or 13.0 min arc. It was shown that changes in saccade amplitude of 4 - 6 min arc could be recorded as the average across 29 – 41 single
saccades. This finding suggests that in the binocular case - vergence changes of the order of 8 – 12 min arc could be resolved.

2. Validity

The present study refers to fixation disparity, which is defined as the difference in vergence between the binocular observation of a target and a monocular reference observation (during calibration in the present study); i.e. fixation disparity is a relative measurement. Thus, to specify the validity with respect to fixation disparity, it is required to test the system in response to small vergence changes.

Validity is conventionally tested by using a gold-standard recording system for comparison. Since such an objective eyetracker is not available in our laboratory, we tested whether the EyeLink II system is able to provide measures of vergence changes that are geometrically expected if we change the stimulus for vergence by small amounts. For this purpose, the haploscope was modified to have half-silvered mirrors so that the subject was able to view straight ahead onto a flat screen where the fusion target was presented. The viewing distance was changed by defined small amounts, so that the stimulus for vergence and accommodation was varied. The following series of test conditions was applied. As a starting point, we measured the fixation disparity at an initial baseline viewing distance of 600 mm which corresponds to a baseline vergence angle of 6.0 deg. i.e., 2 arc tan (PD/2D), with an average pupillary distance PD=63 mm and the viewing distance D from the target to the plane of the centre of rotation of the eyes. For the subsequent tests, we shifted the target by -28, -14, 13.5, 26 mm relative to the initial viewing distance of 600 mm to vary the vergence stimulus by small amounts of ± 8 min arc and ± 16 min arc; minus (or plus) signs refer to a shift to more distant (or closer) positions. To allow for such small range adjustments, the flat monitor was mounted on a mechanical stand that could be shifted back and forth in a purpose-made slide. The following 6 experimental conditions were presented in two different orders to reduce possible effects of sequential testing: (1) 0, -16, +16, -8, +8, 0 min arc and (2) 0, +16, -16, +8, -8, 0 min arc. We made six of these runs (three of each order) in a single session and four sessions for each subject.

The fixation stimulus was a central cross of 18 min size, surrounded by 12 crosses of 30 min arc size in a circular area of 8 deg diameter to assist fusion. The following time scheme was used. The fusion stimulus was presented for 3 s. After this period, the stimulus was removed from the screen for 1 s, during which the experimenter shifted the screen to the next position. The subjects were instructed to blink during this period, in order to stimulate a fusional response to the new target position. Before and after the presentation of these 6 target positions, a monocular calibration was performed. The monocular calibration targets were presented in a haploscope which allowed the calibration targets to be viewed monocularly on two monitors left and right of the mirrors; the fusion target at eye level was observed straight ahead through the mirrors. Although a bite bar was not used, the subject’s head was stabilized in a chin and forehead rest; additionally, translation of the head was minimized by applying firmly adjusted pads for the cheeks and by fixing the head with a flexible band around the back of the head.

The results in Figure S1 show on the x-axis the shift in vergence stimulus and on the y-axis the vergence response, both relative to the geometrically expected baseline vergence angle of 6.0 deg corresponding to the baseline viewing distance of 600 cm. Thus, a line with a slope of one with zero y-intercept (heavy line) indicates the expected result if a subject had a zero fixation disparity and a change in vergence response as expected from geometry. This is approximately the case in subject WP. The y-intercept of the regression lines (dotted line) represents the fixation disparity, which was negative indicating an exo (uncrossed) fixation disparity in three of these four subjects, up to
Figure S1. The change in vergence response (y-axis; min arc) as a function of change in the stimulus vergence (x-axis; min arc) for 4 different observers: WK, WP, JW, and TB. The dots represent trimmed means (by 20%; ± SD). The heavy line has a slope of 1.0 and zero y-intercept which represents the geometrically expected result for a subject with zero fixation disparity. The dashed lines are regression lines to the data points.

an amount of 40 min arc in subject TB. A statistical mixed-effects model estimated a standard deviation of the fixation disparity (y-intercept) of 6.6 minutes of arc, averaged across these four observers. Most important, the slope is close to one in all four subjects, which indicates that the present instrumentation and test procedure are able to detect and appropriately quantify small vergence changes in the range of ± 16 min arc.

3. Methods for time-averaging the objective fixation disparity from raw data
For averaging the continuous vergence signal v(t) over the period of testing between the pre- and post-calibration, we primarily used the “trimmed mean procedure”, as described in our paper. This appeared to be a simple and efficient procedure. However, for comparison, we also applied two alternative methods of analysis, i.e. the “weighting function procedure”, and the “single fixation procedure”. Here, we provide a comparison of these three procedures since particularly the weighting function procedure has advantages for some experimental conditions.

The weighting function procedure comprised the following steps of analysis. In the first step, we applied to the vergence signal v(t) a combined numerical differentiation and low pass filtering with a 6 dB cutoff frequency of 2 Hz; this filter used 401 sampling points (2 ms separation) of the first derivative of a normally distributed density function. The positive
part of the resulting velocity signal was squared; it attributed a high weight to the moments in time of rightward saccades, it is referred to as $w_{Rsac}(t)$. Similarly, the square of the negative part of the velocity signal gave a high weight to the leftward saccades, i.e. $w_{Lsac}(t)$. These two weighting functions $w_{Lsac}(t)$ and $w_{Rsac}(t)$ were used to formulate the following weighting function for fixation periods: i.e. $w_{FIX}(t) = 1 / \{1 + w_{Lsac}(t) + w_{Rsac}(t)\}$. This function adopts high values close to 1.0 whenever the saccadic weighting functions adopt values close to zero, i.e. in moments of time where no saccadic eye movement is performed, thus a fixation occurs. The weighting function $w_{FIX}(t)$ is multiplied by the vergence signal $v(t)$, the result is averaged over time and normalized by the integral of the weighting function to find a mean vergence state across the period of testing.

The single fixation procedure was used for the data of the short fixation periods of about 1.5 s between saccadic gaze shifts: we determined the offset of the saccade to be 30 ms after the moment of maximal velocity and averaged vergence across a period 200 – 300 ms after saccade offset. In this way, we found a vergence measure for each of the 240 objectively measured short fixation periods.

Box-plots for each subject are shown in Figure S2 to illustrate individual differences in objective fixation disparity (for short fixation periods) and to compare the different procedures for analysing the vergence signal. In Figure S2a, we plotted the results of all 240 short fixation periods as calculated by the “single observation procedure” described above. The scatter even within individuals is rather large and amounts to a range of about 2 deg including the outliers; it remains unclear to what extent this scatter may represent physiological variability due to some unknown factors or random measurement error. However, the boxes illustrating the median ± one quartile suggest that some subjects show a clear deviation of the inter-quartile range from zero fixation disparity, mostly in the exo direction and fewer in the eso direction. Figure S2b shows results of the “weighting function procedure” (see above) which gives

Figure S2. Objective fixation disparity for short fixation periods described by individual box-plots for each of the 17 subjects ordered by their objectively measured heterophoria (from left to right: exo to eso). The three methods of signal analysis are compared. (a) shows the results of the “single observation analysis” with $n = 240$ data points per subject (from 8 runs each with 30 short fixation periods). (b) shows the results of the “weighting function procedure” based on 8 runs and (c) the results of the “trimmed mean procedure” based on 8 runs.
an average value across an experimental run, comprising 10 observations of each of the three targets, i.e. 30 observations in total; objective recordings are available from 8 of these runs. This averaging procedure reduces the variance within each subject as compared to the single observation procedure (Figure S2a). Figure S2c shows results of the “trimmed mean procedure”, which represents an average across the same data. The resulting individual boxplots of the trimmed mean procedure resemble those of the weighting function procedure. The individual mean values of objective fixation disparity were highly correlated between the three methods of signal analysis ($r > 0.95$).

This comparison of the different procedures shows that the objective fixation disparity data obtained are robust in the sense that they are not dependent upon a particular method of signal analysis. For the data presentation in our paper, we used the results of the trimmed mean procedure, since it is appropriate for the condition of long fixation periods, where no stimulus-triggered saccades were present.

4. Stability

In the present experiment, the subjective measures of fixation disparity require a considerable period of time until sufficient subjective responses to the nonius lines are collected. Particularly, the objective recordings of the 10 short fixation periods of each of the three targets took a period of 45 s (before and after which a calibration was performed). Thus, the question arises whether the head position remains sufficiently fixed to allow for stable eye movement recordings over 45 s. To test whether the resulting data might have been contaminated by drifts due to residual head movements, we calculated a comparative data set based on only 10 s directly after calibration. As shown in Figure S3, we found a high correlation ($r = 0.93$, $p < 0.0001$, $n = 17$) between objective fixation disparity from the reduced 10 s period and the full 45 s period (both calculated with the trimmed mean procedure). This suggests that residual head movements did not have a detrimental effect on the dataset as a whole, at least with regard to correlations between different optometric measures which is the main aim of the present study. Detrimental effects of movements of the head are reduced in our procedure by analysing the difference between the signals of the left and right eye; the resulting full vergence angle is unaffected by small translational head movements (head position is much more critical if monocular components of fixation disparity are to be measured).

5. Statistical properties of objective fixation disparity

From the data reported in the result section, we can deduce some statistical properties of the objective fixation disparity measures. The mixed-effects model in Table 1 (main article) estimated an average intra-individual standard deviation $SD_{\text{intra}}$ due to the $n$ repeated measurements (each of these represent the average resulting from the trimmed-mean procedure). These were $SD_{\text{intra}} = 14.4$ min arc, $n = 8$ for short fixation periods and $SD_{\text{intra}} = 10.6$ min arc, $n = 4$ for long fixation periods. From these
standard deviations, we calculate a confidence interval \( \text{CI} = z(0.05) \times SD_{\text{intra}} / \sqrt{n} \), where \( z(0.05) = 1.96 \) is the critical value corresponding to the area of 0.95 below the normal probability function for a two-tailed significance level of 0.05. For short and long fixation periods, we find figures of \( \text{CI} = 9.9 \text{ min arc} \) and \( \text{CI} = 10.4 \text{ min arc} \), respectively. Thus, generally for our data set, we can conclude that an objective fixation disparity with an absolute amount of about 10 min arc (or larger) is significantly different from 0.0. If two subjects are compared, their average objective fixation disparity must differ by at least \( 10 \times \sqrt{2} = 14 \text{ min arc} \) to be significantly different. Such interindividual differences occurred in a number of cases of the present sample which resulted in the significant correlations reported.

6. Dynamics of fixation disparity after saccades

The stimulus-triggered saccades in short fixation periods allow the time course of vergence adjustment during the early phase of fixation to be tested. For this purpose, we calculated the fixation disparity separately for three intervals after the saccade: \( \text{FD}_{100} \) refers to the interval 0 – 100 ms after saccade offset (which was assumed to be 30 ms after maximal saccade velocity), \( \text{FD}_{200} \) refers to 100 – 200 ms and \( \text{FD}_{300} \) to 200 – 300 ms. \( \text{FD}_{300} \) corresponds to the data reported in the result section. In Figure S4, we describe changes in fixation disparity after the saccades calculated by \( \text{FD}_{100} - \text{FD}_{300} \) (open circles) and \( \text{FD}_{200} - \text{FD}_{300} \) (closed circles). These two differences are plotted for each subject; the zero level in this graph represents the fixation disparity of the third interval (\( \text{FD}_{300} \)) for each subject, while any offset from zero shows the deviation in fixation disparity from \( \text{FD}_{300} \) during the earlier intervals; arrows denote the change from \( \text{FD}_{100} \) to \( \text{FD}_{200} \). In this presentation, a large arrow starting at a negative level means that after the saccade a large transient divergent state occurs that is reduced during the early phase of fixation; such effects are less pronounced or absent in other subjects represented by smaller arrows. Figure S4 shows that in 15 of 17 subjects, the arrow is directed towards zero indicating that the fixation disparity is reduced and approaches the final level \( \text{FD}_{300} \); in most cases, this a vergence movement in the (positive) convergent direction starting at a more divergent level immediately after the saccade. The \( \text{FD}_{200} - \text{FD}_{300} \) data are mostly scattered around zero in a range of \( \pm 3 \text{ min arc} \) so that the final vergence state during fixation appears to be reached in the period 100 – 300 ms after saccade offset.

**Figure S4.** For illustrating the dynamic process of vergence adjustment in the early phase of the short fixation periods, we calculated the fixation disparity separately for three intervals after the saccade offset: \( \text{FD}_{100} \) refers to the interval 0 – 100 ms after saccade offset, \( \text{FD}_{200} \) refers to 100 – 200 ms and \( \text{FD}_{300} \) to 200 – 300 ms. The y-axis shows differences in fixation disparity relative to \( \text{FD}_{300} \) (i.e., the final fixation disparity measure used in the “single fixation analysis”); two measures of changes in fixation disparity are shown: \( \text{FD}_{100} - \text{FD}_{300} \) (open circles) and \( \text{FD}_{200} - \text{FD}_{300} \) (closed circles), the arrows indicate the vergence change from \( \text{FD}_{100} \) to \( \text{FD}_{200} \). For each subject shown on the x-axis, the data of Session 2 are shown, while the order of the presentation of the subject was made according to the difference \( \text{FD}_{100} - \text{FD}_{300} \) from Session 1.

The data shown originate from Session 2, while the order of the presentation of the subject was arranged according to the difference \( \text{FD}_{100} - \text{FD}_{300} \) from Session 1, to illustrate a possible stability of these effects: subjects with larger negative values of \( \text{FD}_{100} - \text{FD}_{300} \) (open circles) are predominately found in the left half of the ordering scheme suggesting that they have reliably large
transient divergence in Session 1 and in Session 2. The latter observation is confirmed by a significant correlation \((r = 0.88, p < 0.0001, n = 17)\) for \(FD_{100} - FD_{300}\) between the two sessions. This result suggests that larger initial deviations in the early phase of fixation appear to be reliable in some subjects.

The analyses of the early phase of fixation showed the typically expected convergent movement fixation that was reported in many conditions, e.g. with single point light sources (Collewijn et al., 1988, 1997; Zee et al., 1992; Kapoula et al., 1987; Fioravanti et al., 1995; Yang and Kapoula, 2003) and during reading (Hendriks, 1996; Heller and Radach, 1998; Nuthmann and Kliegl, 2009; Vernet and Kapoula, 2009). Some of these studies reported that subjects differ in the amount of the convergent change during fixation, as confirmed in the present study.

7. Ocular rotation versus ocular displacement

Video eye trackers (or other techniques based on the image of the eye or the cornea reflex) do not differentiate whether the eye has rotated or a displacement has occurred. In particular viewing conditions, the eyes may translate in the nasal/temporal direction (lateral) and/or may shift forward or backward in the orbit. For example, temporal shifts of the eyes of 0 – 200 µm were precisely measured by Enright (1980) in a task where subjects changed monocular fixation between viewing distances of 4 m and 15 cm; although vision was monocular, a 20 deg change in accommodative vergence was expected. Effects of such ocular displacements on vergence measures have been investigated by Tani et al. (1956) using photographic recordings of the cornea reflex: the authors measured vergence, while the stimulus was first increased until the break point of fusion was reached and then reduced towards the recovery point and to fusion again. In these conditions of large amounts of forced vergence, Tani et al. (1956) interpreted their own results and earlier findings as indicating that the eyes did not rotate round a fixed point in the orbit, rather displacements of the eyes can occur: thus, the center of rotation may follow a curve when the eyes turn by large angles. However, a quantitative description of these patterns of shifts in eye position did not appear from these studies. Contradictory observations from different studies were also noted; e.g., Alpern (1957) found forced vergence fixation disparity curves measured with an electro-oculogram to be similar to those based on the corneal reflex method and concluded that eye displacements are extremely unlikely during vergence eye movements. Considering that Tani et al. (1956) noted that convergence of the eyes within 10 degrees usually does not induce ocular displacements, we see no evidence that any relevant displacements of the eyes occurred in the present experiment where we used a constant comfortable vergence state of 6 deg, i.e. without high vergence load or particularly near vision.

Conclusion

The control experiments and additional analyses made in this Appendix provide evidence that our objective fixation disparity measures are appropriate for the aim of the present study. Certainly, the instrumentation and procedures used in the present study have some inherent limitations (e.g., using no bite bar and only horizontal calibration) that will have introduced some measurement error. However, these sources of error are random, independent and unbiased and therefore are reduced by averaging the recorded data over time during each recording period and across a number of repeated measurements in two experimental sessions. Apparently, the measurement noise was small enough so that statistical properties of the data set suggest that – on the average in the sample – objective measures of fixation disparity of about 10 min arc or larger are significantly different from zero fixation disparity (which represents the monocular reference condition). Further, differences in objective fixation disparity between individuals of at
least 14 min arc are significant, as found in our data set. Due to these statistical properties, the resulting individual differences in objective fixation disparity are reflected in significant correlations between the two sessions (about $r = 0.84$) and between heterophoria and objective fixation disparity (about $r = 0.82$). Further, the main results of the study were found to be very similar in the two conditions of fixation (short and long fixation periods with intercorrelations of about $r = 0.85$); this replication shows the internal consistency of our findings. Therefore, the Eye Link II instrumentation, the experimental procedures and data analyses appear to be sufficient for the purposes of the present study.

References


