



The temporal and spatial limits of compensation for fixational eye movements

Guy Wallis *

*Perception and Motor Systems Laboratory, Connell Building, University of Queensland, Qld 4072, Australia
Max Planck Institute for Biological Cybernetics, Spemannstraße 38, 72076 Tübingen, Germany*

Received 31 May 2005; received in revised form 13 December 2005

Abstract

High-fidelity eye tracking is combined with a perceptual grouping task to provide insight into the likely mechanisms underlying the compensation of retinal image motion caused by movement of the eyes. The experiments describe the covert detection of minute temporal and spatial offsets incorporated into a test stimulus. Analysis of eye motion on individual trials indicates that the temporal offset sensitivity is actually due to motion of the eye inducing artificial spatial offsets in the briefly presented stimuli. The results have strong implications for two popular models of compensation for fixational eye movements, namely efference copy and image-based models. If an efference copy model is assumed, the results place constraints on the spatial accuracy and source of compensation. If an image-based model is assumed then limitations are placed on the integration time window over which motion estimates are calculated.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Fixation; Eye-movements; Motion compensation

1. Introduction

Our eyes are constantly in motion. Even during periods of fixation our eyes produce a range of characteristic, oscillatory movements. This provides our visual system with a significant problem: It must somehow dissociate eye-based from real-world motion signals. One means of doing this involves subtracting a copy (efference copy) of the muscular control signals directed to the eye from the incoming retinal image, an idea which was first formally proposed in the 1950s (Sperry, 1950; von Holst & Mittelstaedt, 1950). In the case of saccadic eye movements (of which there are numerous identifiable types) there is certainly good evidence that such a signal is available to the visual system, even if it is not always utilized (Deubel, Schneider, & Bridgeman, 2002).

Although the efference copy model can account for the compensation of some forms of eye movement, it is unlikely that it could be used to counteract them all, especially those associated with periods of fixation. There are at least two reasons for thinking this: First, some eye movements may be due to spurious discharge in the ocular muscles, rather than being driven by a specific command signal. Second, motion of the retina may not be due to rotational eye movements, but rather to translational ones caused by motion of the head. An alternative to the efference copy model is that we estimate retinal motion by performing an optic-flow analysis on the retinal image itself. This has the advantage of integrating all forms of global motion in the image, irrespective of their source. However, it brings with it the disadvantage that it may, under certain circumstances, make mistakes. This potential for making mistakes has actually been offered as an explanation for certain types of visual motion illusions such as the jitter after-effect (Murakami & Cavanagh, 1998) and Leviant's Enigma (Mon-Williams & Wann, 1996).

* Fax: +61 7 3365 6877.

E-mail address: gwallis@hms.uq.edu.au.

URL: <http://www.hms.uq.edu.au/vislab>.

One image-based compensation model to have received considerable interest in recent years was put forward by Murakami and Cavanagh (1998). This model proposes that retinal motion is estimated on the basis of motion vectors drawn from numerous regions within the retinal image. A recent study of retinal cells in several vertebrate species has identified how such a subtraction process might be implemented in the eye (Ölveczky, Baccus, & Meister, 2003). Despite these theoretical and experimental advances, what has been lacking up until now is direct evidence for such a mechanism at work in humans. Instead, the direct evidence that does exist actually speaks against such a model. Studies conducted in the 60s and 70s looking at the compensation of fixational eye movements (e.g., Findlay, 1974; Matin, Matin, & Pearce, 1970), found no evidence for the correction of slower movements, and although Findlay (1974) did find some evidence for correction for microsaccadic movements, he attributed this to an efference copy model. The main element lacking from these earlier studies was a more detailed study of other forms of fixational eye movement. It has been known for many years that small amplitude, high-frequency movements of the eye take place during fixation (Bolger, Bojanic, Sheahan, Coakley, & Malone, 1999; Carpenter, 1988; Ratliff & Riggs, 1950), but they have often been thought of as being too small to affect perception. It is only relatively recently that debate on the topic has been reopened (Martinez-Conde, Macknik, & Hubel, 2004). This paper focuses on these low amplitude, high-frequency movements, and through a pair of experiments aims to establish spatial and temporal constraints on the two models of image motion compensation.

2. Experiment I

2.1. Introduction

The first experiment focuses on the temporal characteristics of retinal image motion compensation. The temporal characteristics are important because of a fundamental limitation of image-based motion compensation, namely its integration period. An image-based mechanism requires that compensation takes place over a narrow, but finite time-window, during which global retinal shift is estimated. If the compensation mechanism does contain an integration period of this type, it should be possible to identify the lower limit for the duration of the integration period using very briefly presented visual stimuli.

The stimulus used to search for this effect consisted of a grid of circular elements—see Fig. 1A. The perceptual grouping of elements within such grids was first studied by the early Gestalt psychologists. One of the most comprehensive studies of this effect was made by Wertheimer (1923), who measured how grouping is affected by introducing minor irregularities to the arrangement of grid elements. In particular, Wertheimer described how subjects tend to report seeing a grid as containing rows of elements

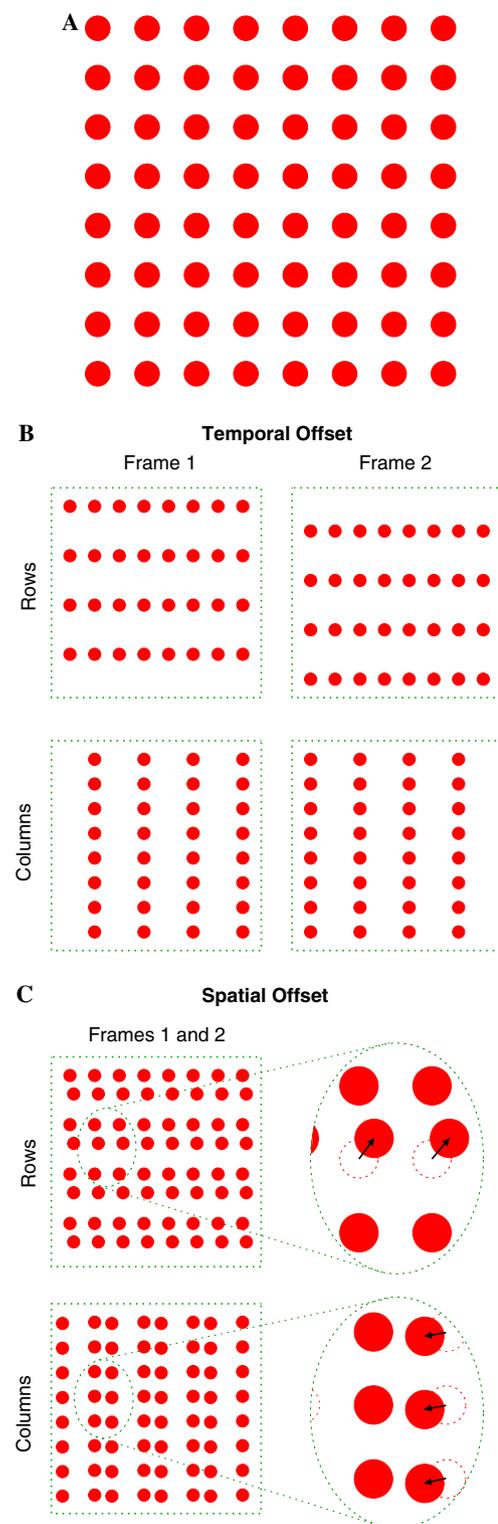


Fig. 1. Summary of the two presentation regimes used in the experiments. (A) The regular stimulus grid. (B) The asynchronous presentation paradigm used in Experiment I. Alternate rows (or columns) were shown in alternate frames on the CRT screen. (C) The synchronous condition used in Experiment II. All 64 grid discs appeared simultaneously, but with alternate rows (or columns) displayed with a small, randomly oriented spatial offset relative to the reference frame of the grid. The magnified sections on the right indicate the direction and magnitude of the displacement in each example case. The pale, dotted frames appear for illustrative purposes only and did not form part of the actual stimulus.

rather than columns, if the elements are arranged more closely horizontally (proximity rule), or if the linearity of the columns is disrupted (good continuation). Conversely, columns of elements are perceived in preference to rows if the proximity and/or alignment of the elements is reversed (see also Ben-Av & Sagi, 1995; Fahle & Koch, 1995). Sensitivity to these types of grid misalignment forms the basis of the investigations described in this paper.

The experiment combines the idea of an integration time window with the perceptual grouping of grid elements by introducing a temporal delay between the appearance of either alternate rows or alternate columns. By keeping presentation delay times very low it is possible to create the impression of a single, rapidly flashed stimulus (see discussion of stimulus). However, despite the apparent unified nature of the stimulus, it was hypothesized that a mechanism attempting to estimate retinal motion from the incoming image might fail. Failure would arise if the mechanism did not fully compensate for the spatial offset between the two visual frames caused by the retina being in different locations during the two presentation periods. Any image motion not compensated for should result in an apparent spatial shift in the location of the grid elements as they appear on the retina. Any such shift should, in turn, result in the disruption of the perceived row and column arrangement of the elements, and hence affect observer responses.

2.2. Methods

2.2.1. Stimuli

Subjects viewed a square grid of 64 regularly spaced circular elements—see Fig. 1A. Subjects viewed the stimuli at a distance of 1570 mm at which point the grid subtended an angle of approximately $5^\circ \times 5^\circ$ and a display pixel subtended $0.01^\circ \times 0.01^\circ$. The stimuli were presented as filled red circles on a dull grey background in a darkened room. Red was chosen because persistence of our monitor's P22 red channel is minimal, with most of the stored energy dissipating within 2 ms. This is not true of the green and blue channels however, which despite decaying quickly initially, produce a visible afterglow for a much longer period (Kuhn, 2002; Sherr, 1993). Images were presented on a Sony Trinitron CRT monitor using a 1.6 GHz Dell PC. Mean screen luminance was 11 cd/m^2 with a stimulus Michelson contrast of 20%.

In this first experiment, stimuli were presented in perfect spatial alignment, but only half of the elements (e.g., odd numbered rows) were displayed for a single frame ($\approx 12 \text{ ms}$), followed by the other half of the elements (e.g., even numbered rows) during the second frame—see Fig. 1B. As no trial specific stimulus offset was required in this experiment, the four images required were rendered off-line using an SGI ONYX 350 (see Section 3). Despite the temporal offset, and in accord with earlier investigations (Usher & Donnelly, 1998), the percept described by the five naive subjects was of a single, stationary stimulus.

Their subjective impression is important because the perception of temporally offset stimuli is a large and complex issue. Westheimer and McKee (1977a) have reported that temporal offsets of as little as a few milliseconds can be detected reliably, possibly due to their triggering motion detection mechanisms. Aware of such a possibility, Usher and Donnelly (1998) investigated whether apparent motion was responsible for the sensitivity to the temporal offsets which they described. They found no compelling evidence to suggest that it did.

The use of a cathode ray tube display device brings with it the potential for artifacts due to the temporal offset between the appearance of early versus later horizontal scan lines. Taking into account the screen refresh period (11.6 ms—see below), the vertical blanking period ($\approx 0.5 \text{ ms}$), and the visual angle subtended by the stimulus (5°) relative to that of the screen ($\approx 10^\circ$), the first row of dots would have appeared approximately 5 ms before the last row. The timing for column appearance is clearly different and it is possible that the two might lead to interactions with subjects' perceptual judgments. One simple method for testing this possibility is to rotate the screen through 90° since this inverts the relationship between monitor scan direction and rows versus columns. Pilot studies, using four of the subjects who took part in the two main experiments, revealed consistent levels of sensitivity and bias irrespective of the orientation of the monitor. If u indicates scores for the upright monitor and r those for the rotated monitor, individual changes in sensitivity were: $[d'_u - d'_r = 0.233]t(7) = 0.675$, $p = 0.521$ and bias: $[c_u - c_r = 0.121]t(7) = 0.717$, $p = 0.497$.

The image generation PC communicated with a second PC via a hard-wired Ethernet interface. The second PC sampled data from an SMI Eye-Link I eye tracker, recording eye position data at a rate of 250 Hz. The second PC runs under DOS and allows for the time stamping of events happening on the display generator at a rate of 1000 Hz. The DOS PC clock formed the basis against which all other timing was measured. Screen refresh time for valid trials was estimated on the basis of this clock and was found to be 11.6 ms with a SD of 0.48 ms. Trials in which the presentation time of a single frame deviated by more than 2 ms from the mean were excluded from the analysis (average 2.79% of trials per subject). Also excluded were trials in which saccades were detected. Saccades were distinguished from other eye movements on the basis of current eye velocity and acceleration. A saccade was deemed to be in progress if eye-movement velocity exceeded $30^\circ/\text{s}$ and accelerations were in excess of $8000^\circ/\text{s}^2$. A saccade was said to start if these criteria were met for over more than two sampling periods, and continued as long as the criteria were met again within the next 20 ms. A period of 25 ms was then added after the end of the saccade before fixation was deemed to have been achieved. Using these criteria an average of 1.5% of trials were rejected per subject.

The SMI eye tracker is a video-based system, which measures eye movement on the basis of pupil position

within the video image. The image is recorded by two lightweight high-frequency video cameras, one placed just below each eye, and supported by a head-mounted brace. The system has been compared with a scleral coil system and been found to be remarkably accurate and precise, even for tracing saccadic eye movements with speeds of up to 300°/s—far beyond the maximum velocities associated with fixational eye movements (van der Geest & Frens, 2002). Although drift over several seconds or minutes can be an issue with a head-mounted system, the extremely brief measuring periods used in this study mean that the measurements are actually noise limited. Measurements made by the manufacturer with an artificial pupil place the level of this noise at around 0.005° RMS. A spectral analysis of noise measured in the setup described in this paper appears in the results section of this experiment.

One of the advantages of the video-based system over scleral coils is that the comfort level of subjects is considerably improved. Recent papers comparing eye movements using the Eyelink I eyetracker to a scleral coil system have found systematic changes in observer behavior in terms of duration, amplitude and frequency of saccades when wearing scleral coils (Frens & van der Geest, 2002; Smeets & Hooge, 2003). Since fixational eye movements are, to some extent, subject to voluntary and task specific control (Steinman, Cunitz, Timberlake, & Herman, 1967; Winterson & Collewyn, 1976), it is possible that the overall nature of fixational eye movements is influenced by using coils. A video-based system helps avoid some of these issues.

2.2.2. Subjects and testing procedure

Six subjects took part in the experiment. The grid was viewed monocularly with head placement regulated by means of a bite bar. Both eyes were tested independently for 200 trials. Each trial was initiated by the observer via a button press, after which a fixation cross appeared in the center of the screen for 1.5 s, followed by the stimulus.

During the experiment fixation was monitored via the eye tracker. A trial was only allowed to proceed if the calculated gaze location was within one degree of the fixation cross. Otherwise an audible tone was sounded and the trial restarted.

2.3. Results

Performance in the row/column decision task was calculated using signal detection techniques (Green & Swets, 1974; Macmillan & Creelman, 1991). The observer d' values were calculated on the basis of the sensitivity of participants to the type of temporal offset being employed on a particular trial. If z computes the inverse of the normal cumulative density function, C_{ROW} represents the number of times alternate rows were temporally displaced and the subject correctly responded 'ROW', and I_{ROW} the number of times alternate rows were temporally displaced and the subject incorrectly responded 'column' etc., then the formula used for calculating d' was as follows:

$$d' = z \left[\frac{C_{\text{ROW}}}{C_{\text{ROW}} + I_{\text{ROW}}} \right] - z \left[\frac{I_{\text{COL}}}{C_{\text{COL}} + I_{\text{COL}}} \right].$$

The upper half of Fig. 2A presents eye movement recording data from a single trial and the relevant section used to estimate eye motion is shown alongside in Fig. 2B. Fig. 3A presents the number of times a difference in average eye location from frame 1 to frame 2 occurred during the experiment, averaged across subjects and collapsed into bins of 0.01°. The majority of trials (74.8%) recorded an eye movement of between 0.01° and 0.06°. Fig. 3B displays the corresponding responses as a function of retinal image displacement from frame 1 to frame 2. Sensitivity for movements in the range 0.01–0.06° was always well above that for eye movement in the range 0.00–0.01°. For the 15% of trials in which movement amplitude was above 0.06°, sensitivity fell dramatically and no longer differed signifi-

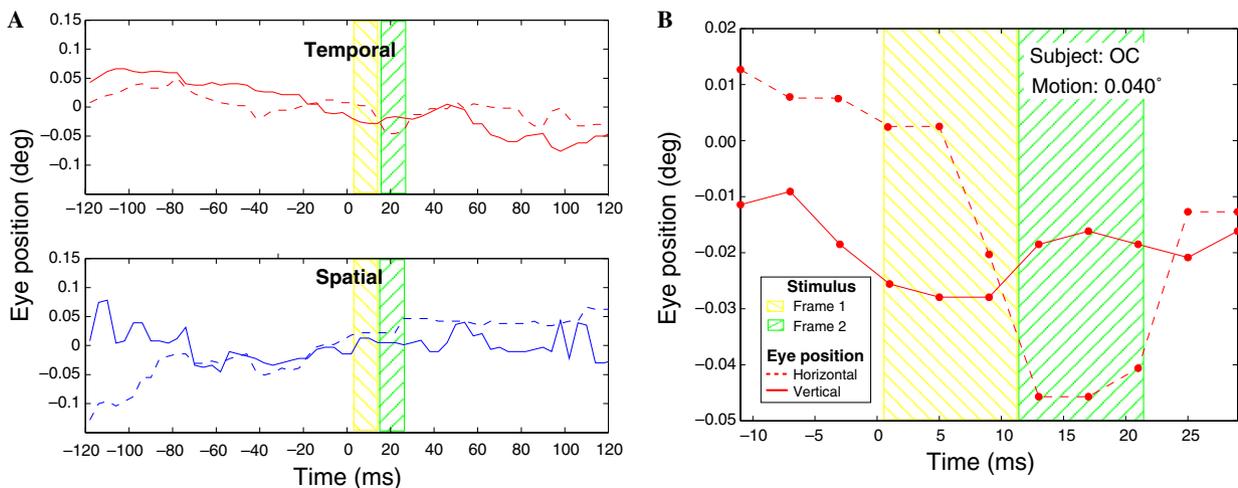


Fig. 2. Eye-movement data from both experiments. (A) Traces of eye movement measured along the horizontal and vertical axes. The hatched area indicates the period during which the two stimulus frames were presented. (B) Expanded view of the data for Experiment I.

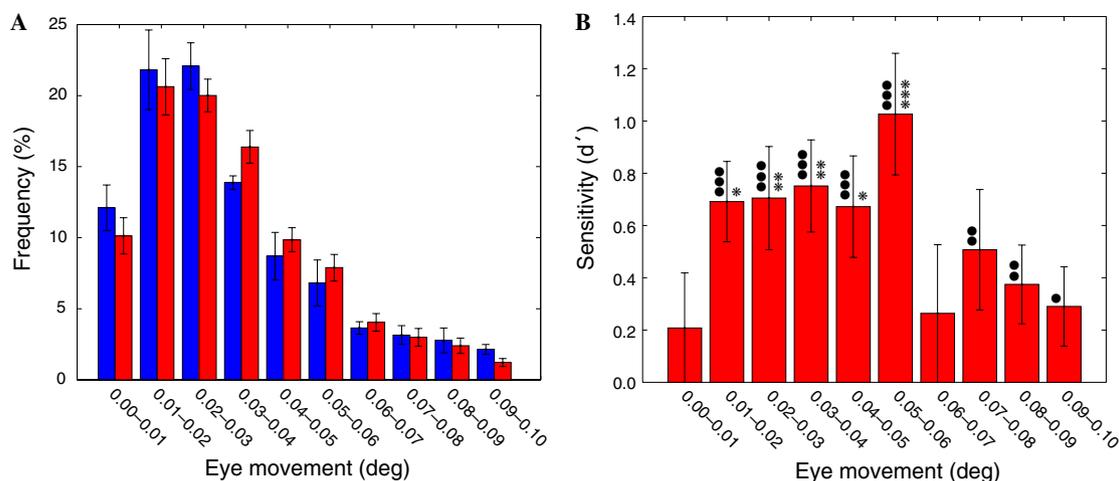


Fig. 3. Summary of eye movement amplitudes during stimulus presentation. (A) Percentage of trials in which eye movements of a particular amplitude range were recorded. Note that during majority of trials (70%), the eye moved from 0.01° to 0.05° between presentation frames. Bin sizes from the two experiments are essentially the same, discounting the possibility that participants adopted abnormal eye movement behavior to help solve the temporal offset version of the experiment. (B) Sensitivity in the row/column task with trials collected into 0.01° bins for Experiment I. Error bars indicate standard error of the mean. Asterisks indicate significant differences: (i) $*p < 0.1$, $**p < 0.05$, $***p < 0.01$ for the difference between results for the $0.00\text{--}0.01^\circ$ range. (ii) $*p < 0.1$, $**p < 0.05$, $***p < 0.01$ for differences from chance level ($d' = 0$).

cantly from that for motion in the range $0.00\text{--}0.01^\circ$, although it differed significantly from purely random selection ($d' = 0$). Movements of this size presumably indicate the occurrence of more rapid, microsaccadic or saccadic eye movements, during which normal image processing may well be disrupted. The relatively small number of such trials insured that overall performance was well above chance $t(11) = 5.251$, $p < 0.001$ (average $d' = 0.75$) revealing that the subjects' percept was significantly influenced by the manipulation over the entire set of trials. This is consistent with earlier studies of grid-based stimuli (Fahle, 1993; Farid, 2002; Usher & Donnelly, 1998; Wallis, 2005).

Given the very high precision (as opposed to accuracy) required by the eye tracking system to faithfully record displacement of the eye, it is important to assess the level actually achieved using the set up employed in these experiments. This was done by comparing the spectral content of the eye-location signal with that of measurement noise. Noise was estimated by running the entire experiment with the eye tracker monitoring the location of a fixed, artificial pupil. Results appear in Fig. 4, and suggest that noise levels were well below the content of the signal even at high frequencies. Note that the noise levels shown here offer a worse case figure since at higher frequencies, power would be reduced by the frame-by-frame location averaging procedure.

One further concern about these studies is that subjects may be adopting eye movement strategies to enhance spatial distortions derived from the temporal offset. As described above, to minimize the possible role of any such effects the analysis specifically excluded trials in which saccades were detected, and required a period of measured fixation before the trial proceeded. Nonetheless, there is good evidence that observers are capable of affecting the preva-

lence of microsaccades (Steinman et al., 1967; Winterson & Collewijn, 1976) and so it is informative to see if rapid, small amplitude eye movements coincided with stimulus presentation. To this end, average eye-movement velocities were estimated for all valid trials in which an individual responded correctly. Measurements were made over a 40 ms period for each subject before, during, and after stimulus presentation. T tests were used to establish whether the difference between average velocity 'before versus during' or 'after versus during' differed significantly from zero. The data are displayed in graphical form in Fig. 5. Of the 24 comparisons, three were significant. All three revealed a tendency for the velocity to be slightly lower during stimulus presentation than after. Indeed, the figure suggests a slight trend across subjects for this to happen. The difference was small $\bar{x} = -0.2875^\circ/\text{s}$, but significant $F(1,11) = 7.362$, $MSE = 0.1348$, $p = 0.02$. In fact, there was also a small trend for motion before the stimulus presentation to be higher than during presentation as well $\bar{x} = -0.2797^\circ/\text{s}$, $F(1,11) = 4.417$, $MSE = 0.2126$, $p = 0.059$. Overall the results confirm that subjects were not adopting a strategy of increasing eye motion during stimulus presentation to enhance the amount of stimulus displacement caused. If there was any tendency at all, it was to decrease eye movement velocity during presentation.

2.4. Conclusion

The major outcome of the first experiment is that eye movement amplitude is an excellent predictor of whether a subject's response will be affected by the temporal delay or not. It seems, therefore, that although the elements on the screen were displayed in perfect alignment, motion of

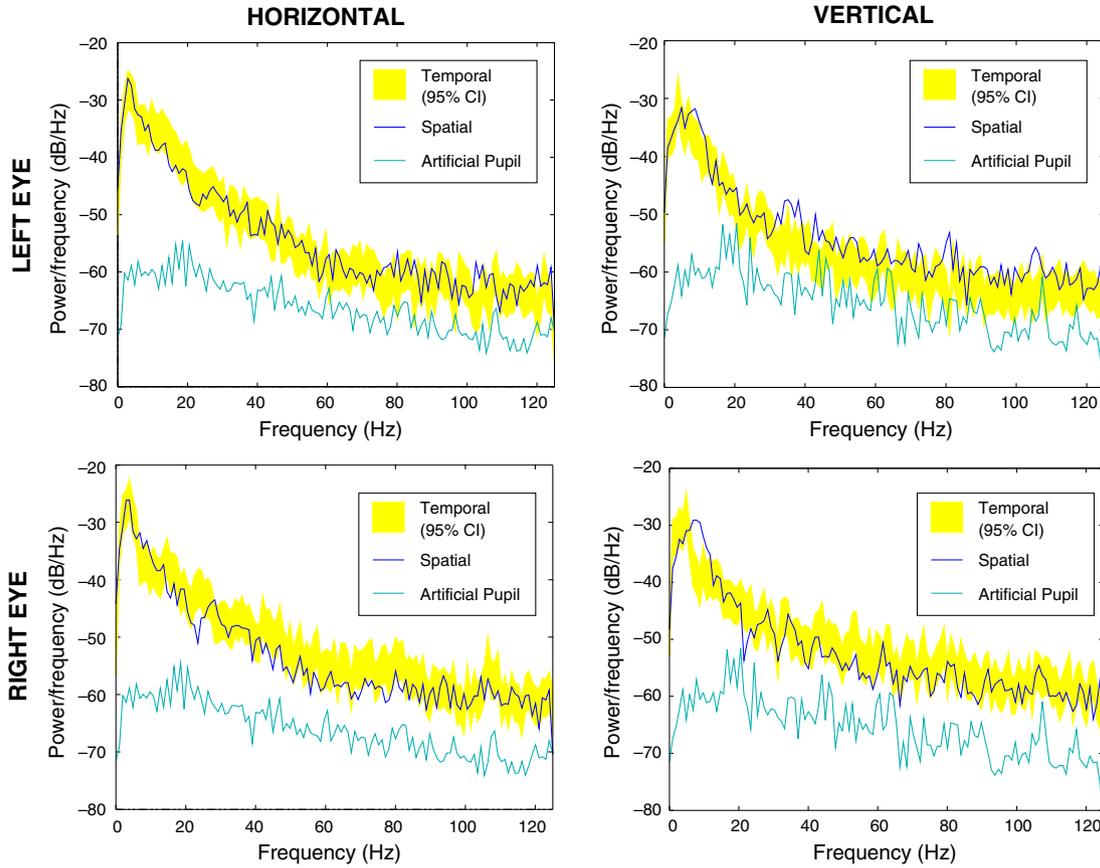


Fig. 4. Power spectral density plots for movement of each eye for one subject, split into horizontal and vertical components. The light-grey band represents the 95% confidence interval of the spectrum for movements recorded during Experiment I. The dark line is the spectrum calculated during Experiment II. These are seen to largely overlap, suggesting that the two spectra generated during the two experiments are essentially the same. The light-grey line presents system noise estimated by recording data from an artificial, stationary pupil.

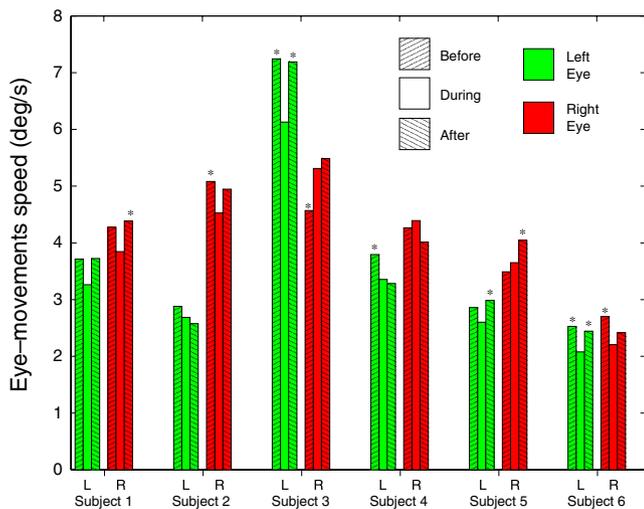


Fig. 5. Mean eye-movement velocities arranged by subject and eye. The three bars represent the average velocity measured over a 40ms period before, during, and after stimulus presentation. Asterisks denote a significant difference ($p < 0.05$) between the velocity either before or after stimulus presentation relative to that measured during presentation. Overall a slight tendency for velocities to increase after presentation is apparent although the pattern is not always consistent across subjects. This difference dispels the suggestion that subjects adopted a strategy to deliberately increase eye movement velocities during stimulus presentation to increase the effective spatial displacement caused.

the eye caused them to appear misaligned, directly affecting response choice. Only in the case that the eye movement between frames fell below 0.01° did selection return to purely random levels $t(11) = 0.987, p = 0.345$.

The apparent drop in sensitivity with eye movement amplitude beyond 0.06° may well, as mentioned above, be due to the fact that these trials represent the small portion of occasions in which microsaccades were underway during presentation. Certainly the size of the displacement is consistent with a high velocity, microsaccadic movement. From previous studies of microsaccades, we know that they occur at a rate of around 2 per second and last around 25 ms (Coakley, 1983; Martinez-Conde et al., 2004). Given that each trial lasted 25 ms, one might predict that in approximately 10% of trials, stimulus presentation would overlap with a microsaccadic movement. This figure accords remarkably well with the number of trials in which offsets of over 0.06° were recorded (10.7%).

So why might microsaccades reduce sensitivity to the temporal offset? Findlay (1974) has argued that microsaccades are corrected for on the basis of actual eye movement, which would explain why subjects were unaffected by the temporal offset in these trials. A less interesting, but nonetheless real possibility is that microsaccades blur

the image, reducing the capacity of subjects to resolve the spatial offsets induced by the eye movement. As it stands the experimental paradigm used cannot distinguish between perfect, veridical perception and poor acuity, since both will result in perceptual grouping returning to random, chance levels ($d' = 0$).

Irrespective of the explanation for what happens in the ten percent of trials in which larger displacements occur, the important discovery is that for the vast majority of trials, movement amplitude is a strong predictor of the degree to which subjects' selections were influenced by the temporal offset. This could only happen if retinal motion compensation is, at least in part, lacking in this case.

3. Experiment II

3.1. Introduction

On the basis of the first experiment it appears that little or no compensation of fixational eye movement is taking place for very small amplitude eye movements. The hypothesis being put forward is that eye movements are turning temporal offsets into discernible spatial offsets. This second experiment focuses on this issue by investigating observer performance on a purely spatial version of the perceptual grouping task. If the spatial offset explanation holds, subjects should be sensitive to real spatial offsets with amplitudes similar to those measured in Experiment I.

It has been known for many years that humans are capable of extremely fine spatial discriminations. Reports place the limit of visual acuity at levels well below the width of a single photoreceptor (Fahle, 2002, chap. 11; Westheimer & McKee, 1977b; Wulffing, 1892). The classical test stimulus in studies of visual acuity involves a pair of vertically oriented lines but many other stimuli have revealed similar effects including the alignment of dot stimuli (Ludvig, 1953; Westheimer & McKee, 1977b). All of these studies suggest that the spatial offsets seen in Experiment I should be well within the capability of the human visual system to resolve. On the other hand, eye movements with a temporal frequency sufficient to underlie the effects described here would seem to be approaching the flicker fusion rate. Indeed, in the past it has often been assumed that such eye movements were of too high a frequency to affect perception (Gerrits & Vendrik, 1970; Martinez-Conde et al., 2004).

The results from Experiment I appear to dispute those assumptions but it would be informative to compare sensitivity to purely spatial offsets to that for the temporally induced offsets of Experiment I. If the levels of performance are comparable, it will add further support to the proposal that the temporal offsets are being converted into spatial offsets in the grid and that no compensation is taking place.

A further question to arise from the first experiment is whether the eye movements produced by the subjects are

typical, or whether the subjects may have adopted a strategy of deliberately increasing the instability of their gaze. This new version of the experiment provides a situation in which eye movements can only serve to hinder performance rather than enhance it, and hence a benchmark against which eye movement behavior in the previous experiment can be compared.

3.2. Methods

3.2.1. Stimulus

Fig. 1C summarizes the pattern of stimulus presentation adopted in the experiment. A single image was displayed for two presentation cycles (frame time ≈ 12 ms) with alternate rows (or columns) spatially offset from perfect alignment—SYNC presentation. Misalignment varied from 0.00° to 0.04° in 0.01° steps. The direction of displacement was chosen at random on each trial. The 8×8 grid was displayed using 8×8 pixel, on-line antialiasing, providing fine grain sub-pixel position accuracy. Stimuli were presented on the same monitor used in Experiment I but the image was now displayed directly by the SGI Onyx350 which had been used to generate the image displayed by the PC in Experiment I.

3.2.2. Subjects and testing procedure

The same six subjects involved in Experiment I took part in Experiment II. Procedures were identical to those employed in Experiment I.

3.3. Results

Fig. 2 displays a record of eye movement data for a typical trial. Fig. 3 presents a summary of eye movement amplitude binned into 0.01° intervals. The average number of eye movements falling within each range is seen to be comparable to that obtained during the temporal presentation version of the experiment, serving to discount the possibility that participants were adopting an abnormal eye movement strategy to solve the temporal offset version of the experiment. Further evidence for this conclusion is provided by the power spectral density plot in Fig. 4. The spectrum for the spatial offset experiment is seen to fall within the 95% confidence interval of the spectrum measured during the temporal offset version of the experiment. In general the spectra can be described as exhibiting a low frequency peak, followed by a rapid drop, and then a flattening between 50 and 100 Hz, a pattern broadly consistent with earlier measurements (Eizenman, Hallett, & Fecker, 1985; Spauschus, Marsden, Halliday, Rosenberg, & Brown, 1999). The characteristic, second drop off beyond 120 Hz is beyond the capacity of the current eye tracking system to determine.

Fig. 6 displays the subject sensitivity to spatial offsets of the grid elements. As one would expect, at zero degrees offset subjects are responding at chance level (corresponding to a d' of zero). However, for spatial offsets above zero

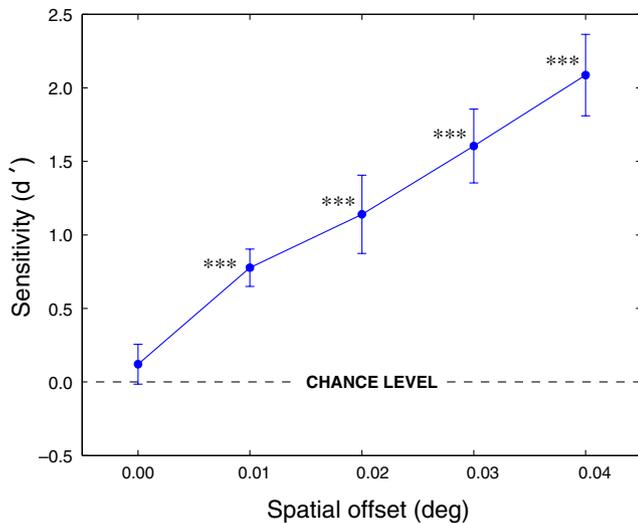


Fig. 6. Results from experiment II. The graph displays the tendency of subjects to choose rows (columns) if adjacent rows (columns) are displaced by spatial offsets of 0.01° and above. At zero offset subjects are understandably at chance level ($d' = 0$) but even at 0.01° they show a significant tendency to choose in accordance with the spatial offset. Asterisks indicate level of difference from chance level. Error bars indicate standard error of the mean. Significance level: *** $p < 0.01$.

degrees, observers demonstrated a consistent and significant sensitivity to that offset. In fact, sensitivity to offsets of as little as 0.01° was significantly above chance ($t(11) = 6.149$, $p < 0.01$). There was also a significant main effect of spatial offset amplitude: $F(1, 11) = 12.635$, $MSE = 0.5415$, $p < 0.001$.

3.4. Conclusion

On the basis of the second experiment we can conclude that subjects are indeed sensitive to remarkably small spatial offsets within a grid stimulus. This is despite the fact that only one of the six subjects actually reported noticing any grid misalignment during the experiment. Although remarkable, sensitivity to such small offsets is consistent with what one might have predicted based on results from other tasks focusing on visual acuity (Fahle, 2002, chap. 11; Westheimer & McKee, 1977b; Wulffing, 1892), and more specifically the (mis)alignment of circular elements (Ludvig, 1953). Nonetheless the experiment has served to confirm that high levels of acuity are maintained despite the unusually short presentation time and novel form of the acuity task.

What this experiment can also tell us is the degree to which the temporal offset data corresponds to the spatial offset data. To assess this, the average level of sensitivity recorded for displacements of less than 0.06° was calculated for each eye of each subject for the temporally offset stimulus. Then the expected values of image displacement, based on individual bin frequency data, were calculated. This provided a reference value of displacement from which to gauge performance in Experiment II for each eye of each subject. Average performance in the spatial

version of the experiment was generally higher than predicted by the estimate of the eye-movement induced offset (1.29 vs 0.80). Although relatively large, the difference in d' values narrowly failed to achieve statistical significance $t(11) = -2.090$, $p = 0.06$.

If, as seems likely, this small discrepancy in sensitivity is due to more than random fluctuation, it may be attributable to one of a number of effects such as: (i) The simplistic averaging model used to estimate the induced spatial offset is inaccurate (i.e., effective offset may have been smaller than the simple vector average calculated). (ii) Image blurring associated with the failure of normal motion compensation may have reduced image quality in Experiment I.

Whatever the reason for the discrepancy, the most important conclusion is that sensitivity to spatial offsets is more than sufficient to account for the temporal offset effects described in Experiment I. It does not prove that temporal offsets are being turned into spatial ones in the first experiment, but it is certainly consistent with the suggestion that they are.

4. Discussion

The purpose of this paper has been to investigate the mechanisms underlying compensation for retinal image motion caused by both lateral and rotational movements of the eye. The two experiments have provided evidence that for eye movements up to around 0.06° , little or no compensation takes place when the eye is subjected to very brief image presentations. This result has important implications for the roles played by two popular models of motion compensation.

To further tease these roles apart it is helpful to consider the different categories of fixational eye movement established in the literature. Researchers generally identify three categories: drift, ocular motor tremor, and microsaccades, broadly distinguished on the basis of their speed, amplitude and frequency (Adler & Fliegelman, 1934; Carpenter, 1988; Martinez-Conde et al., 2004; Ratliff & Riggs, 1950). Eye drift produces relatively large amplitude movements, but its low speed (Ratliff & Riggs, 1950; Martinez-Conde et al., 2004) would be insufficient to result in a large stimulus shift on the time scales considered in this paper. Like drift, microsaccades are relatively large fixational eye movements, but they are also fast, making them a more likely cause of the effect described here. However, their relative in frequency (approximately 1 every 500 ms (Coakley, 1983; Martinez-Conde et al., 2004)) would limit their influence over many trials. There is also some evidence that they are, at least in part, corrected for via an internally generated (e.g., efference copy) motion compensation mechanism (Findlay, 1974).

The only remaining form of fixational eye movement, ocular microtremor (OMT), has a relatively small amplitude but is both fast and continuous (Bolger et al., 1999; Carpenter, 1988; Ratliff & Riggs, 1950), making it a much more likely candidate for producing the eye-movement

related sensitivity measured in these studies. What is more, the extremely small amplitude of OMT accords well with the range of movement amplitudes over which sensitivity was greatest, namely $0.01\text{--}0.06^\circ$. Although OMT is a likely contributor to the effects reported here, any source of small amplitude but relatively high frequency image motion would also contribute. For example, head movements can double the amplitude of fixational jitter during free viewing (Ferman, Collewijn, Jansen, & Vandenberg, 1987). On the other hand, fixational head movements have stronger low frequency components than OMT and so may well be too slow to have played more than a minor role in these studies. Nonetheless, there may well be other sources of movement between the stimulus and eyes. Indeed, it is the multiple sources of internally and externally generated motion noise which would seem to make some form of image-based correction essential.

One question which remains unresolved is the precise reason for the lack of compensation in the $0.01\text{--}0.06^\circ$ image motion range reported here. It may simply indicate the spatial resolution of both efference copy and image-based compensation mechanisms. There is some evidence to support this from earlier studies of microsaccades. Findlay recorded visual acuity in subjects using a 250 ms delay between presentation of a pair of vertical bars. In the event that a microsaccade occurred during the delay interval, subjects were able to resolve the relative location of the bars at around 60% accuracy for offsets of around 0.03° , somewhat less accurately than the baseline of 73% measured in the absence of a microsaccade. The fact that the errors were often in the direction of the microsaccade lead Findlay to conclude that subjects were compensating for the microsaccades, but that they actually tended to over-compensate. The subjects' relatively poor performance at this level of visual angle suggests that the microsaccadic compensation Findlay reported was operating close to the limit of its spatial resolution.

To track microsaccadic eye movements to the resolution reported by Findlay, any compensation must presumably have excellent temporal resolution, since the angular velocities attained by the eye can be very high up to 100 times that of OMT (Martinez-Conde et al., 2004). This would appear to rule out the possibility that the mechanism employed to compensate for microsaccades is ineffectual in the studies reported here because of its temporal resolution. Although the spatial resolution issue seems plausible there is another alternative. If, as Findlay's results suggest, microsaccadic movement compensation is based on an efference copy model, it is conceivable that the eye movement command signal is copied upstream of the source of tremor. This is interesting because, contrary to earlier findings, recent work has found that tremor is partially coherent in the two eyes (Spauschus et al., 1999). This would seem to suggest that tremor contains an internally driven component whose origin is fairly far upstream. This, in turn, would constrain the origin of the efference copy signal (Spauschus et al., 1999).

In the case of image-based compensation, the results may once again indicate that the limit of spatial resolution has been reached. However, unlike the efference copy model a temporal limitation may also be responsible if the integration time window for motion estimation exceeds 15 ms. One recent technological advance which may provide some insight into whether it is the spatial or temporal limit which is decisive in this case, is the scanning laser ophthalmoscope. Using such devices it has become possible to track the motion of the retina to the level of individual photoreceptors (Roorda et al., 2002). Early indications of work involving motion discrimination in the presence of a visual reference frame, suggest that motion is compensated for and that the frame's motion calibrates perceived location rather than the exact receptors being stimulated (Stevenson, Raghunandan, Frazier, Poonja, & Roorda, 2004). This is consistent with the idea that a spatially highly accurate, image-based mechanism is at work which would suggest that it is the temporal limitations of compensation which are affecting performance in the experiments described here. Incidentally, the work also suggests that in the absence of a visual reference frame, eye motion is still factored into motion estimates, although it is considerably less reliable (Stevenson et al., 2004), consistent with the functioning of a spatially less accurate, efference copy model.

To further corroborate the interpretation being proposed here it would be interesting to investigate the effects of altering the frame duration. If the effects described here are due to eye-movement induced spatial offsets, then changing the presentation time should alter the size of this shift, and hence observer sensitivity. Usher and Donnelly (1998) investigated the role of presentation duration, but their measurements used varying numbers of multiple frame presentations, making it difficult to dissociate the effect of frame duration from that due to varying the number of repetitions of each frame pair. If a study did restrict presentations to a single pair of stimulus frames, it seems likely that the current temporal offset is close to optimal in its ability to exploit the effects of tremor. Shorter frame durations, with frequencies above the fundamental frequency of tremor (≈ 50 to 100 Hz), will tend to produce ever smaller spatial shifts from frame 1 to frame 2. Longer frame durations too, will tend to induce smaller spatial offsets because of the sinusoidal nature of tremor. These issues are dealt with in more detail in Appendix A.

By placing the visual system in an unusual situation involving extremely rapidly presented stimuli in combination with an alignment task (for which we have a remarkable degree of sensitivity), it has been possible for fixational eye movements (including ocular motor tremor) to have measurable effects on perception. Beyond the rather limited range of stimuli involving grids of elements, it is unlikely that perceptual grouping tasks of almost any other stimuli would be affected by the minuscule spatial offsets generated by fixational image motion. Nonetheless, in the face of certain visual patterns, the motion estimation system can fail

quite dramatically, producing a compelling but illusory impression of motion in static stimuli (Hine, Cook, & Rogers, 1995; Mon-Williams & Wann, 1996; Murakami & Cavanagh, 1998). This paper serves to bridge the gap between theory and behavior by providing evidence for a link between fixational eye movements and perception, as well as placing a lower limit on the spatiotemporal resolution of an imaged-based, motion estimation mechanism.

Acknowledgments

I am grateful to Susana Martinez-Conde, David Burr, Mike Land and Ulrike Siebeck for comments on the manuscript, and to Heiner Deubel, Trevor Hine, James Tresilian, and Daniel Berger for helpful comments and discussion. This work was supported by the Australian Research Council, Australia and Max Planck Society, Germany.

Appendix A. Estimating the optimal ratio of frame rate to tremor frequency

The degree to which motion of the eye leads to spatial displacement of a pair of temporally offset images, depends on the period of the eye oscillation T_E relative to the duration of the two stimulus frames T_S . This effect will be further scaled by the amplitude of the eye movement A , and the phase of the eye movement at which the stimulus first appears. If the first image frame appears τ seconds after the start of positive displacement of the eye from its mean position, the offset of the two image frames Ψ is given by:

$$\begin{aligned} \Psi &= \frac{2}{T_S} \int_{t_1}^{t_2} A \sin \left[\frac{2\pi t}{T_E} \right] dt - \frac{2}{T_S} \int_{t_2}^{t_3} A \sin \left[\frac{2\pi t}{T_E} \right] dt \\ &= \frac{AT_E}{\pi T_S} \left(\cos \left[\frac{2\pi t_1}{T_E} \right] - 2 \cos \left[\frac{2\pi t_2}{T_E} \right] + \cos \left[\frac{2\pi t_3}{T_E} \right] \right), \end{aligned}$$

where $t_1 = \tau$, $t_2 = \tau + T_S/2$, $t_3 = \tau + T_S$ (see Wallis, 2005).

Over multiple trials the value of τ varies randomly over the range $0-T_E$. An estimate of the expected mean squared offset $\overline{\Psi^2}$ can be obtained as follows:

$$\overline{\Psi^2} = \frac{1}{T_E} \int_0^{T_E} (\Psi(\tau))^2 d\tau = 2 \left(\frac{T_E A}{\pi T_S} \right)^2 \left(1 - \cos \left[\frac{\pi T_S}{T_E} \right] \right)^2.$$

The form of this relationship appears plotted in Fig. 7. For any given value of A and T_E , variations in T_S produce a smooth oscillatory change in the estimate of squared offset. The maxima and minima of this function can be identified by equating the following partial differential to zero:

$$\frac{\partial \overline{\Psi^2}}{\partial T_S} = 4 \left(\frac{T_E A}{\pi T_S} \right)^2 \left(\frac{\pi}{T_E} \sin \left[\frac{\pi T_S}{T_E} \right] - \frac{1}{T_S} \left(1 - \cos \left[\frac{\pi T_S}{T_E} \right] \right) \right) \left(1 - \cos \left[\frac{\pi T_S}{T_E} \right] \right).$$

From this equation $\overline{\Psi^2}$ has minima for $T_S = nT_E$ for any even integer n , but also has a series of maxima which decrease in amplitude as the ratio of T_S to T_E increases—see Fig. 7. The first maximum occurs for $T_S = 0.7419T_E$

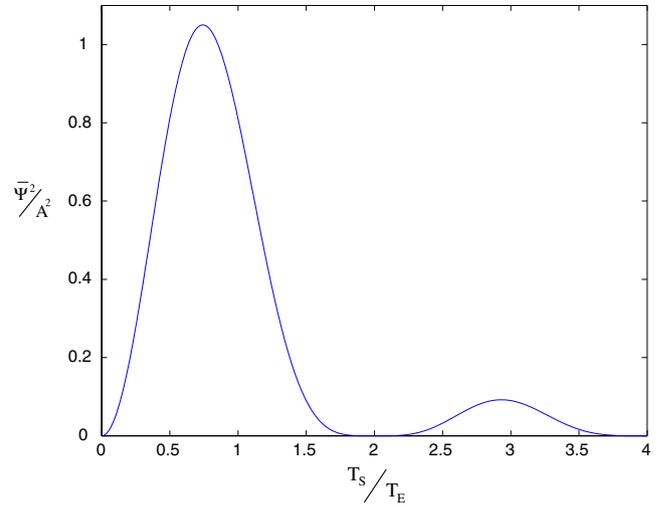


Fig. 7. Variation in the estimated squared image displacement $\overline{\Psi^2}$ (normalized by the square of eye movement amplitude A^2), as a function of the ratio of the oscillation period of tremor T_E , and stimulus display period T_S .

at which point $\overline{\Psi^2} = 1.05A^2$. Since each stimulus presentation contains two image frames, the optimal presentation duration for a frame is $0.37 T_E$. Current estimates for T_E lie in the range 10–25 ms (Bolger et al., 1999; Martinez-Conde et al., 2004), which corresponds to a monitor refresh rate of 100–270 Hz.

References

- Adler, F., & Fliegelman, F. (1934). Influence of fixation on the visual acuity. *Archives of Ophthalmology*, *12*, 475–483.
- Ben-Av, M., & Sagi, D. (1995). Perceptual grouping by similarity and proximity: Experimental results can be predicted by intensity autocorrelations. *Vision Research*, *35*, 853–866.
- Bolger, C., Bojanic, S., Sheahan, N., Coakley, D., & Malone, J. (1999). Dominant frequency content of ocular microtremor from normal subjects. *Vision Research*, *39*(11), 1911–1915.
- Carpenter, R. (1988). *Movements of the eyes*. London: Pion.
- Coakley, D. (1983). *Minute eye movement and brain stem function*. Boca Raton, FL: CRC Press.
- Deubel, H., Schneider, W., & Bridgeman, B. (2002). Transsaccadic memory of position and form. *Progress in Brain Research*, *140*, 165–180.
- Eizenman, M., Hallett, P., & Fecker, R. (1985). Power spectra for ocular drift and tremor. *Vision Research*, *25*, 1635–1640.
- Fahle, M. (1993). Figure-ground discrimination from temporal information. *Proceedings of the Royal Society, London, B*, *254*, 199–203.
- Fahle, M. (2002). Learning to perceive features below the foveal photoreceptor spacing. In M. Fahle & T. Poggio (Eds.), *Perceptual Learning* (pp. 197–218). Cambridge, MA: MIT Press.
- Fahle, M., & Koch, C. (1995). Spatial displacement, but not temporal asynchrony, destroys figural binding. *Vision Research*, *35*, 491–494.
- Farid, H. (2002). Temporal synchrony in perceptual grouping: A critique. *Trends in Cognitive Sciences*, *6*, 284–288.
- Ferman, L., Collewijn, H., Jansen, T., & Vandenberg, A. (1987). Human gaze stability in the horizontal vertical and torsional direction during voluntary head movements, evaluated with a three-dimensional scleral induction coil technique. *Vision Research*, *27*(5), 811–828.
- Findlay, J. (1974). Direction perception and human fixation eye movements. *Vision Research*, *14*, 703–711.

- Frens, M., & van der Geest, J. (2002). Scleral search coils influence saccade dynamics. *Journal of Neurophysiology*, *88*, 692–698.
- Gerrits, H., & Vendrik, A. (1970). Artificial movements of a stabilized image. *Vision Research*, *10*, 1443–1456.
- Green, D., & Swets, J. (1974). *Signal detection theory and psychophysics*. Huntington, NY: R.E. Krieger.
- Hine, T., Cook, M., & Rogers, G. (1995). The ouchi illusion: An anomaly in the perception of rigid motion for limited spatial frequencies and angles. *Perception and Psychophysics*, *59*, 448–455.
- Kuhn, M. (2002). Optical time-domain eavesdropping risks of CRT displays. In *Proceedings of the IEEE symposium on security and privacy*. IEEE, New York, pp. 3–18.
- Ludvigh, E. (1953). Direction sense of the eye. *American Journal of Ophthalmology*, *36*, 139–142.
- Macmillan, N., & Creelman, C. (1991). *Detection theory: A user's guide*. Cambridge, UK: Cambridge University Press.
- Martinez-Conde, S., Macknik, S., & Hubel, D. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, *5*, 229–239.
- Matin, L., Matin, E., & Pearce, D. (1970). Eye movements in the dark during the attempt to maintain a prior fixation position. *Vision Research*, *10*, 837–857.
- Mon-Williams, M., & Wann, J. (1996). An illusion that avoids focus. *Proceedings of the Royal Society, London [B]*, *263*, 573–578.
- Murakami, I., & Cavanagh, P. (1998). A jitter after-effect reveals motion-based stabilization of vision. *Nature*, *395*, 798–801.
- Ölveczky, B., Baccus, S., & Meister, M. (2003). Segregation of object and background motion in the retina. *Nature*, *423*, 401–408.
- Ratliff, F., & Riggs, L. (1950). Involuntary motions of the eye during monocular fixation. *Journal of Experimental Psychology*, *40*, 687–701.
- Roorda, A., Romero-Borja, F., Donnelly, W., Hebert, T., Queener, H., & Campbell, M. (2002). Adaptive optics scanning laser ophthalmoscopy. *Optics Express*, *10*, 405–412.
- Sherr, S. (1993). *Electronic displays*. New York: Wiley.
- Smeets, J., & Hooge, I. (2003). Nature of variability in saccades. *Journal of Neurophysiology*, *90*, 12–20.
- Spauschus, A., Marsden, J., Halliday, D., Rosenberg, J., & Brown, P. (1999). The origin of ocular microtremor in man. *Experimental Brain Research*, *126*, 556–562.
- Sperry, R. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative Physiological Psychology*, *43*, 482–489.
- Steinman, R. M., Cunitz, R. J., Timberlake, G., & Herman, M. (1967). Voluntary control of microsaccades during maintained monocular fixation. *Science*, *155*, 1577–1579.
- Stevenson, S., Raghunandan, A., Frazier, J., Poonja, S., & Roorda, A. (2004). Fixation jitter, motion discrimination and retinal imaging. *Journal of Vision*, *4*, 85a.
- Usher, M., & Donnelly, N. (1998). Visual synchrony affects binding and segmentation in perception. *Nature*, *394*, 179–182.
- van der Geest, J., & Frens, M. (2002). Recording eye movements with video-oculography and scleral search coils: A direct comparison of two methods. *Journal of Neuroscience Methods*, *114*, 185–195.
- von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip: Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften*, *37*, 464–476.
- Wallis, G. (2005). A spatial explanation for synchrony biases in perceptual grouping. *Perception and Psychophysics*, *67*, 345–353.
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forshung*, *4*, 301–350.
- Westheimer, G., & McKee, S. (1977a). Perception of temporal order in adjacent visual stimuli. *Vision Research*, *17*, 887–892.
- Westheimer, G., & McKee, S. (1977b). Spatial configurations for visual hyperacuity. *Vision Research*, *17*, 941–947.
- Winterson, B., & Collewijn, H. (1976). Microsaccades during finely guided visuomotor tasks. *Vision Research*, *16*, 1387–1390.
- Wulffing, E. (1892). Über den kleinsten Gesichtswinkel. *Zeitschrift für Biologie*, *29*, 199–202.