The influence of retinal image motion on the perceptual grouping of temporally asynchronous stimuli

Adela S. Y. Park　　　　Department of Optometry & Vision Sciences,
The University of Melbourne, Melbourne, Australia

Andrew B. Metha　　　　Department of Optometry & Vision Sciences,
The University of Melbourne, Melbourne, Australia　✉

Phillip A. Bedggood　　　Department of Optometry & Vision Sciences,
The University of Melbourne, Melbourne, Australia　✉

Andrew J. Anderson　　　Department of Optometry & Vision Sciences,
The University of Melbourne, Melbourne, Australia　✉

Briefly presented stimuli can reveal the lower limit of retinal-based perceptual stabilization mechanisms. This is demonstrated in perceptual grouping of temporally asynchronous stimuli, in which alternate row or column elements of a regular grid are presented over two successive display frames with an imperceptible temporal offset. The grouping phenomenon results from a subtle shift between alternate grid elements due to incomplete compensation of small, fixational eye movements occurring between the two presentation frames. This suggests that larger retinal shifts should amplify the introduced shifts between alternate grid elements and improve grouping performance. However, large shifts are necessarily absent in small eye movements. Furthermore, shifts follow a random walk, making the relationship between shift magnitude and performance difficult to explore systematically. Here, we established a systematic relationship between retinal image motion and perceptual grouping by presenting alternate grid elements (untracked) during smooth pursuit of known velocities. Our results show grouping performance to improve in direct proportion to pursuit velocity. Any potential compensation by extraretinal signals (e.g., efference copy) does not seem to occur.

Introduction

The perceptual grouping of elements within grids has been studied extensively by Wertheimer (1923), who described rules governing how the percept of rows and columns could be biased by introducing spatial irregularities to the arrangement of grid elements. Purely spatial principles of grouping such as *proximity* and *continuation* can explain the resulting percept of columns (or rows) when elements are arranged such that they are closer in proximity vertically (horizontally) or if the linearity of row (column) elements are disrupted.

Perceptual grouping has been studied in the context of small eye movements by Wallis (2006). The grouping stimulus consisted of a regularly arranged grid of circular elements that were presented in physical alignment. On a given trial, however, alternate row (or column) elements were presented on two successive display frames. Despite having the appearance of a single grid flashed in its entirety, observers were able to reliably group the grid into rows or columns in accordance with the temporally asynchronous stimuli presentations (Usher & Donnelly, 1998; Dakin & Bex, 2002; Wallis, 2005; Wallis, 2006). Wallis (2006) proposed a purely spatial explanation for these results, suggesting that small, fixational eye movements—specifically, the small-amplitude and high-frequency tremor eye movements—occurring between the asynchronous presentation of grid elements results in a spatial shift of grid elements on the retina.

That these retinal shifts influence grouping performance suggests that they are either not perceptually compensated for, or that compensation is incomplete. While efference copy can potentially be used to perceptually compensate for the effects of larger, voluntary eye movements (Sommer & Wurtz, 2008; Sun & Goldberg, 2016), some types of small, fixational eye movements lack specific command signals. As such,

alternatives to the efference copy model have been suggested for the compensation of small (fixational) eye movements. In models using retinal-based mechanisms, retinal motion is estimated from the retinal image itself (Murakami & Cavanagh, 1998), and there is also strong evidence that the visual system primarily encodes differential motion to achieve perceptual stability (Tulunay-Keesey & VerHoeve, 1987; Murakami, 2003; Murakami, 2004; Poletti, Listorti, & Rucci, 2010). Arathorn, Stevenson, Yang, Tiruveedhula, and Roorda (2013) have shown that the compensation of fixational drifts is tuned for direction but not for speed, as similarly demonstrated for smooth pursuit eye movements (Festinger, Sedgwick, & Holtzman, 1976). Additionally, various Bayesian models have highlighted the role of extraretinal signals in the compensation of fixational eye movements (Pitkow, Sompolinsky, & Meister, 2007; Burak, Rokni, Meister, & Sompolinsky, 2010; Freeman, Champion, & Warren, 2010). Wallis provides a convincing argument that perceptual grouping of briefly presented stimuli arises from the mechanisms having insufficient time to estimate retinal motion. Therefore, he proposed that briefly presented temporally asynchronous stimuli could reveal the lower limit of the integration period in which global motion is calculated by a purely retinal-based mechanism (Wallis, 2006).

Fixational eye movements follow a self-avoiding random walk (Engbert, Mergenthaler, Sinn, & Pikovsky, 2011; Herrmann, Metzler, & Engbert, 2017), making it difficult to systematically investigate the precise relationship between retinal slip and perceptual grouping. Wallis has argued that the grouping effect is likely due to tremor eye movements because of their speed (Wallis, 2005; Wallis, 2006). Despite drift eye movements resulting in large amplitude eye movements, their speeds are relatively low and are unlikely to result in sufficiently large stimulus shifts from one frame to the next (Wallis, 2005; Wallis, 2006). It may also be that retinal shifts larger than those seen with fixational eye movements, such as for untracked stimuli when the eye is making a smooth pursuit, can further improve grouping performance. We therefore used smooth pursuit eye movements in our current study to precisely control retinal motion, thereby allowing us to systematically investigate the relationship between the magnitude of retinal shift and perceptual grouping performance.

To use smooth pursuit to introduce retinal shifts, key differences between how perceptual stabilization is achieved for small, involuntary eye movements and for smooth pursuit eye movements must be considered. In a smooth pursuit, unknown eye velocities due to fixational instability are typically small compared with the pursuit velocity, which is known on a moment-to-moment basis from efference copy if the eye is traveling at a nominally fixed velocity. During fixation, however, most of the eye's velocity is not known through efference copy, and so all moment-to-moment velocities must be calculated from retinal signals. Furthermore, the visual system has the potential to use retinal and extraretinal information to stabilize retinal images during pursuits, although it has been suggested that extraretinal information may not be necessary for compensation of slow drifts of $\leq 1°$/s (Royden, Banks, & Crowell, 1992).

Perceptual stability has been proposed to occur through the comparison of efference copy signals and those signals corresponding to retinal image slip (Bridgeman, 1995; Mon-Williams & Tresilian, 1998; van Beers, Wolpert, & Haggard, 2001; Freeman et al., 2010). A nonfixated static object is subject to mislocalization when briefly flashed during a smooth pursuit eye movement (Ward, 1976; van Beers et al., 2001). The mislocalization has been suggested to arise due to a temporal mismatch between retinal and extraretinal signals. As the perceptual grouping task involves brief presentations of alternate grid elements, it might be anticipated that grid elements would be subject to this mislocalization if presented during the course of a pursuit eye movement, and that this mislocalization may be larger for higher-velocity eye movements. However, such errors should affect both presentations of the alternate grid elements equally, and thus be self-canceling.

Although the visual system has access to both retinal and extraretinal signals during smooth pursuit (Bridgeman, 1995; Mon-Williams & Tresilian, 1998), compensation of eye position change resulting from smooth pursuit eye movements has been proposed to occur primarily based on retinal image information (Festinger et al., 1976). If this were the case, then it would be subject to the same temporal constraints as a purely retinal-based mechanism involved in the compensation of involuntary eye movements, as compensation must occur over a narrow, but finite time window. Note, however, that this reliance on local compensation is likely to be limited to conditions involving briefly flashed stimuli; use of both retinal and extraretinal signals is more likely to reflect what happens under more typical, extended viewing conditions, such as in the Bayesian model proposed by Freeman et al. (2010). For our study, we hypothesize that for brief, untracked, asynchronous grid stimuli presented while engaged in smooth pursuit, retinal slip compensation will fail, resulting in improved performance on the grouping task with increasing magnitudes of retinal slip. However, if full compensation were instead achieved using extraretinal signals, knowledge of eye position could be used to compensate for the retinal slip arising from smooth pursuit. In this situation, we would anticipate that full compen-
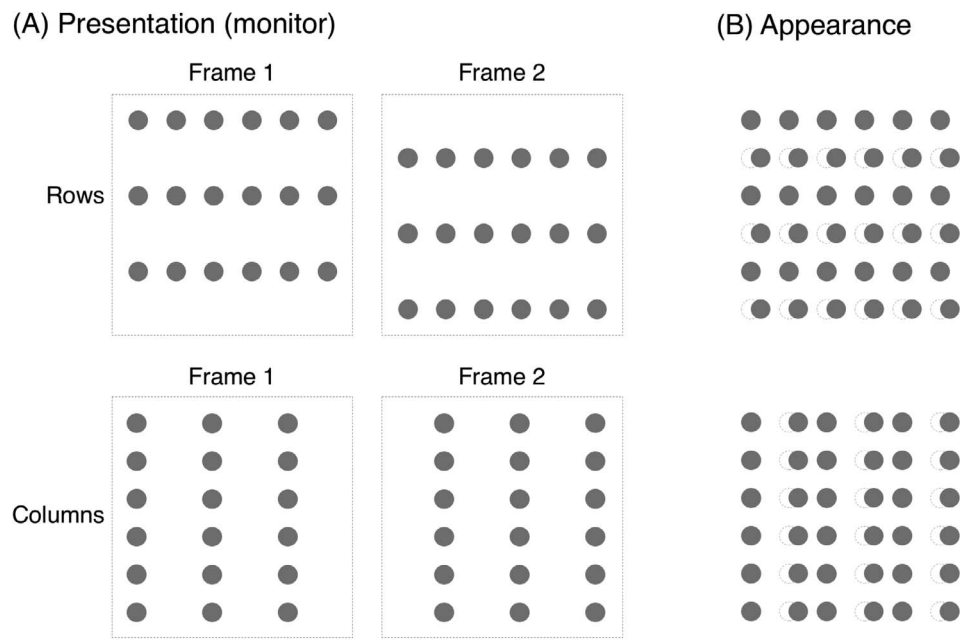
Figure 1. Schematic of the temporally asynchronous grouping stimulus. (A) Alternate rows or columns are presented on two successive display frames in physical alignment. (B) Apparent percept of the stimulus is of a single briefly flashed grid grouped into rows (upper) or columns (lower), resulting from horizontal pursuit eye movements occurring between the two presentation frames. The dotted frames are for illustrative purposes only and did not form a part of the actual stimulus. Note that a 6 × 6 grid has been used to illustrate the presentation and appearance of the grouping stimulus, whereas the experimental stimuli consisted of an 8 × 8 grid.

sation of smooth pursuit by extraretinal signals should result in grouping performance to drop to levels equivalent to that of fixation and remain unaffected by increasing pursuit velocities. Therefore, this study aims to directly and systematically measure the influence of retinal image shifts from pursuit eye movements on perceptual grouping of temporally asynchronous stimuli.

General methods

Twelve healthy observers participated in the experiments. All observers had corrected-to-normal vision better than logMAR 0.10 and viewed stimuli using their habitual spectacle correction and natural pupils. The study followed the Declaration of Helsinki guidelines and was approved by our institutional ethics committee. All observers gave informed consent prior to participation.

Stimulus and experimental setup

The stimuli were presented on a calibrated computer monitor system (ViSaGe graphics card, Cambridge Research Systems, UK; Mitsubishi Diamond Pro 2070sb CRT monitor, resolution 1,024 × 768, frame rate 120 Hz, subtending 41° × 31° at 55 cm) in a darkened room.

The grouping stimulus consisted of 64 regularly arranged circular elements (8 × 8) that were 23.8 min arc in diameter (Figure 1), with horizontal and vertical spacings of 23.8 min arc; these spacings remained fixed for all trials. The entire grid subtended an angle of approximately 6° × 6°. Grid elements were filled gray circles (16 cd/m^2) on a uniform gray background (10 cd/m^2). To check the persistence of the grouping stimulus, we measured the energy dissipation of the gray luminance used for our grid elements with a photodiode, which showed the stored energy to dissipate down to 5% within 1 ms.

On a given trial, alternate row (or column) elements were presented on single, successive display frames while observers tracked a horizontally moving pursuit target (see Figure 2).

Observers viewed the stimulus display binocularly, with their heads stabilized by a chin rest. We tracked horizontal eye position using infrared reflection oculometry (Ober Consulting, Poznan, Poland; sampling rate 1,000 Hz). Pursuit eye movements were recorded during all trials, to assess the accuracy of pursuit movements, especially during the brief presentation of the grid stimuli. The analog signal output from the oculometer was converted to a digital signal
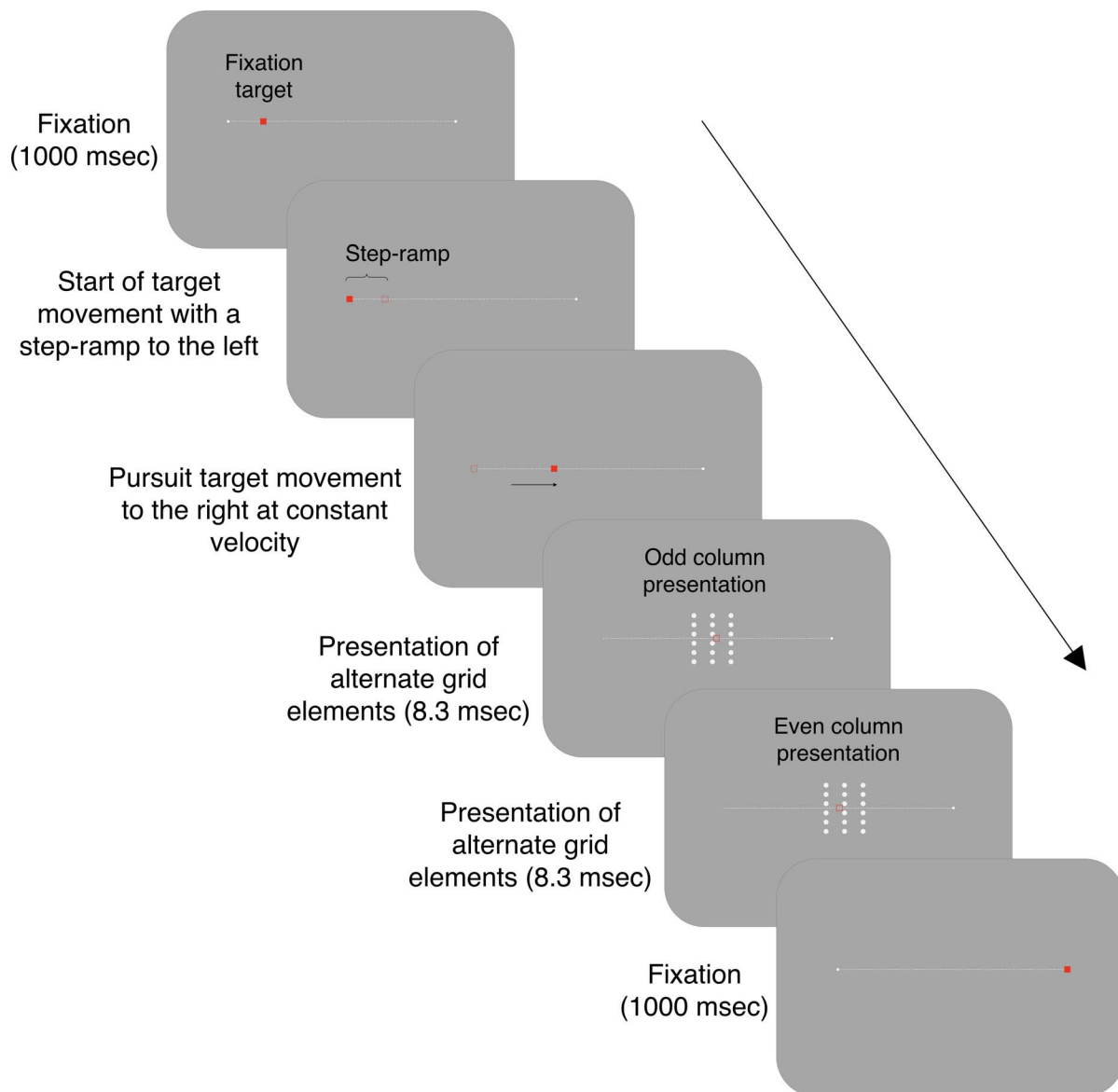
Figure 2. Presentation sequence. To account for the delay in smooth pursuit eye movement initiation, the pursuit target (filled red square) began slightly displaced to the right relative to the beginning of the pursuit path, then after 1,000 ms, jumped to the beginning of the pursuit path (previous target position shown as an unfilled, dotted red square) and immediately began moving to the right at a constant velocity. When the target reached the midpoint of the pursuit path (±10-frame temporal jitter), alternate grid elements (rows or columns) were presented over two successive display frames. The pursuit target briefly disappeared during these two display frames (shown as unfilled, dotted red squares). Note that a 6 × 6 grid has been used to illustrate the presentation of the grouping stimulus, whereas the experimental stimuli consisted of an 8 × 8 grid.

by the ViSaGe ADC system and stored for each trial for offline processing. For simplicity, all motion of the pursuit target was from left to right.

The pursuit target was a red square (9 × 9 arc min) that moved to the right at a constant velocity (see Figure 2). The pursuit path was horizontal and started and ended 1.6° from the left and right edges of the screen, respectively (spanning a total horizontal extent of 38° located at the vertical midline of the screen) unless stated otherwise. For every trial, the pursuit target remained stationary for 1,000 ms at the beginning and end of the pursuit path. Observers were required to fixate on the stable pursuit target with these endpoints used as a reference for trial-by-trial calibrations. The presentation of untracked alternate grid elements occurred when the target reached the midpoint of the pursuit trajectory, with a temporal jitter of ±83.3 ms (10 video frames). This jitter was applied to minimize possible expectation effects of the observer from having knowledge of the precise moment of stimulus onset. As velocity varied from trial to trial, this temporal jitter resulted in the grid

stimulus being presented in slightly different positions, although this variation was only a maximum of 0.30° and 1.79° for the slowest and fastest pursuit velocities, respectively. However, the grid elements were positioned and displayed on the monitor such that the center of the grid area always coincided with the pursuit target and the horizontal path of the pursuit target bisected the grid stimulus.

A step-ramp paradigm was used to account for the typical 150-ms delay associated with the initiation of smooth pursuit eye movements (Rashbass, 1961). At the beginning of the trial, the pursuit target appeared at a fixed distance inward from the pursuit path. This fixed distance was determined by the amount of pursuit target movement expected to occur over the 150-ms delay period and thus differed for the range of pursuit velocities used. After the fixation period of 1,000 ms at the beginning of the trial (see above), the pursuit target would jump to the beginning of the pursuit path (to the left), after which it immediately started moving at a fixed velocity to the right. By the time the smooth pursuit eye movement was initiated, the pursuit target had traveled toward the eye to ideally coincide with where the eyes were pointing.

Procedure

Pursuit target velocities ranged from 0 to 21.6°/s (0.0°/s, 3.6°/s, 7.2°/s, 10.8°/s, 14.4°/s, 18.0°/s, 21.6°/s), where each pursuit velocity was selected to produce retinal shifts in increments of 0.03° steps between successive presentations of alternate grid elements in the grouping task. The range of pursuit velocities tested included a fixation condition (0.0°/s), three velocities that introduced retinal shifts within the range empirically determined for fixational eye movements (3.6°/s, 7.2°/s, 10.8°/s; Wallis, 2006), and three velocities greater than fixational eye movements (14.4°/s, 18.0°/s, and 21.6°/s). This range of pursuit velocities is well within the maximal velocities of about 80°/s to 160°/s for accurate smooth pursuit in humans and where smooth pursuit gain (the ratio between eye velocity and target velocity) has been shown to be near 1.0 (Rashbass, 1961).

As slower velocities would take longer to traverse the entire pursuit path, we shortened the length of the pursuit path for the two slowest velocities (3.6°/s and 7.2°/s). The shortened path was centered about the horizontal midline of the screen, and its extent was determined by the distance the target needed to travel within an equivalent period to the 10.8°/s condition (5.2 s), being the next highest velocity investigated. These adjustments ensured trial durations for the slower velocities were not excessively long relative to the faster velocity trials. For the fixation condition, the fixation

target appeared in the center of the screen and followed the same timings (stayed on for 5.2 s) as above. All other timings for the faster velocity conditions remained unchanged, and the pursuit path was as described in the previous section.

There were 360 trials for each target velocity, and target velocities were randomized from trial to trial. The number of trials ensured sufficient resolution to detect a 5% change on an individual level, based on binomial 95% confidence limits (Clopper & Pearson, 1934). Observers were instructed to follow the target as accurately as possible on every trial. A two-alternative forced choice (2-AFC) method of constant stimuli was used, in which observers were required to indicate by button press whether the grid was grouped into rows or columns. Auditory feedback was given for correct and incorrect responses for every trial.

Pursuit eye movement analysis

The digital signals recorded for pursuit eye movements were processed offline for blinks and saccades and eye velocity calculated, which determined whether the target was faithfully tracked and hence whether the trial should be included for analysis of grouping performance. Blinks appeared as negative deflections, although magnitudes differed across observers. To identify blinks, the raw position digital signal was first smoothed by a 28-point moving average to minimize noise (Souman, Hooge, & Wertheim, 2006), the derivative then taken, and blinks identified using the *isoutlier* MATLAB function that identifies samples that are greater than three scaled median absolute deviations (MADs) from the median. The MAD is defined as the median of the absolute deviations from the data's median (Coren, Bradley, Hoenig, & Girgus, 1975). The scaled MAD is defined and is given by the formula: $MAD = c \times median(|A - median(A)|)$, where A denotes the data set (in this case, the velocity data trace for a given trial) and c is a constant (=1.4826). Any samples in the trace that fell outside ±3MAD were flagged as outliers in the data (see red solid line in Figure 3).

However, this method of identifying blink episodes did not distinguish between blinks or saccades, as both could result in large deviations in velocity. As such, outlier samples identified could be from either blinks or saccades. The voltage values for these samples were removed from the raw trace, while the timing of samples was preserved.

Next, a least-squares trilinear model was fit to the blink-processed data using a least-squares fitting procedure, with the initial and final slopes set to zero (blue dashed line in Figure 3). As the physical extent of
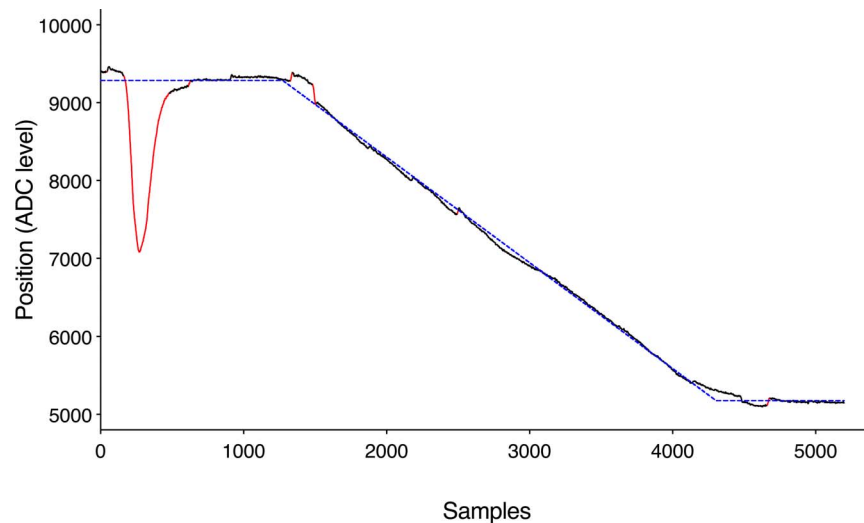
Figure 3. Example of position trace (ADC level) for one participant for the 10.8°/s pursuit velocity condition. The raw position trace is shown (black solid line) after blink and saccade processing. Blinks and saccades were identified as samples falling outside of ±3MAD and were removed (red solid line). A trilinear model was fit to the processed data using a least-squares fitting procedure (blue dashed line).

the pursuit path was known for each trial, and this extent corresponded to the difference between the first and final elements in the trilinear fit, traces could be calibrated in degrees on a trial-by-trial basis.

Lastly, the velocity of eye movement at the time of the grid presentations needed to be assessed to determine the accuracy of the pursuit. However, while the oculometer and the stimulus generator were precise in their timings individually, the two systems were not synchronized with each other. This resulted in a small but variable discrepancy between the timings of the two systems from trial to trial, which was less than 200 ms across all trials. Therefore, a 400-sample window centered about the putative grid presentation was extracted for further processing as the presentation of the grid was guaranteed to occur within this time window. Trials in which saccades were made within this window were identified using a velocity threshold of 60°/s (Souman et al., 2006) and excluded from further analyses. A least-squares linear regression model was fit to the data in this 400-sample window, the slope of which gave the velocity of the pursuit eye movement around the time of the grid presentation for each trial. We also reran our analysis with a lower velocity threshold for saccades of 30°/s to check whether our results may have been affected by smaller saccades being missed. This reanalysis did not alter general findings but did result in a significantly greater number of excluded trials, owing to this velocity criterion being quite close to our highest pursuit velocity condition of 21.6°/s.

Results

Figure 4 shows a monotonic improvement in grouping performance with increasing target velocity. A linear regression of the data showed that the average grouping performance could be well predicted from pursuit target velocity (performance $= 0.02 \times$ target velocity $+ 0.63$; $r^2 = 0.98$, $p < 0.0001$).

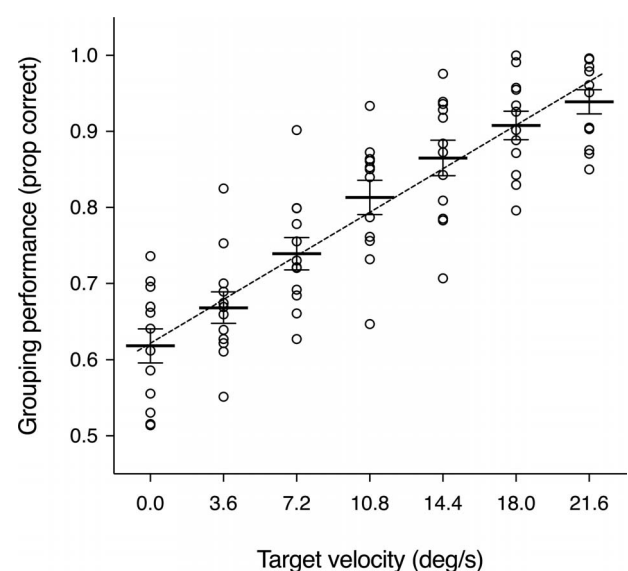Retinal image shifts were extracted from calculated eye (rather than target) velocities for each trial and

Figure 4. Grouping performance for individual observers (circle symbols) as a function of pursuit target velocity. Bars indicate group means ±1 *SEM*.
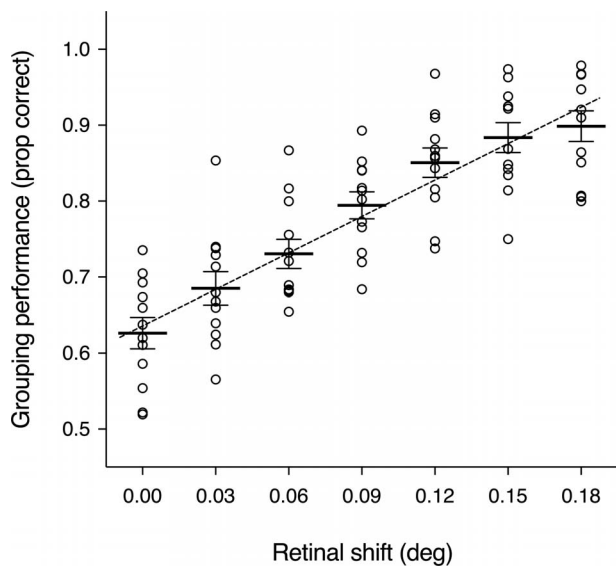
Figure 5. Grouping performance for individual observers (circle symbols) as a function of retinal shift magnitude occurring between presentations of alternate grid elements. Bars indicate group means ±1 *SEM*.

triaged into bins of 0.03°. The results are plotted in Figure 5. Grouping performance improves with increasing retinal image shifts between presentations of alternate grid elements and is described by: performance $= 1.59 \times$ retinal shift $+ 0.64$ ($r^2 = 0.98$, $p < 0.0001$).

To assess whether observers were adopting eye movement strategies (e.g., moving their eyes substantially faster than the target) to enhance the shift in the arrangement of alternate grid elements derived from
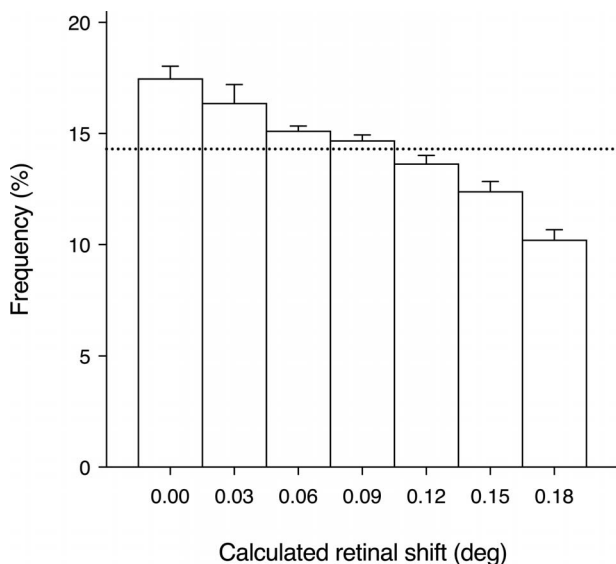

Figure 6. Histogram showing the distribution of mean retinal shifts introduced across trials included in the analysis of grouping performance for all observers. Error bars denote ±1 *SEM*.
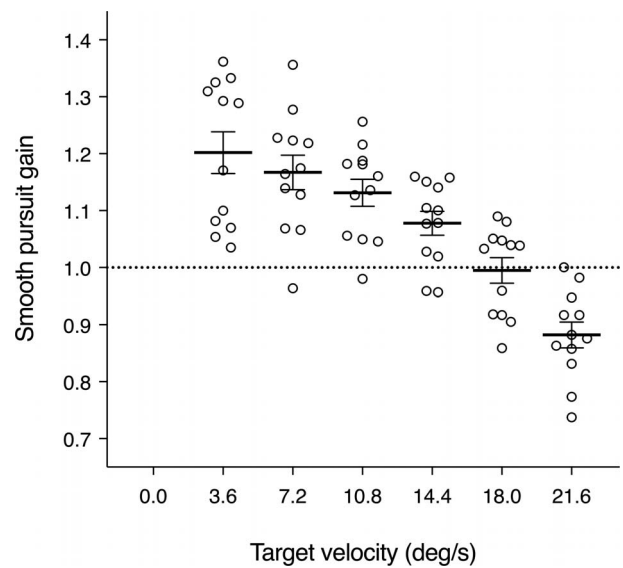

Figure 7. Smooth pursuit gain shown for individual observers (circle symbols) for all pursuit target velocities. Bars indicate group means ±1 *SEM*.

the temporal offset, the distribution of retinal shifts was examined. The distribution of retinal shifts for all trials included for analysis of grouping performance is shown in Figure 6. Target velocities were specifically chosen to introduce retinal shifts in increments of 0.03°, corresponding to seven bins, and as such, the expected frequency of trials for each bin would be 1/7 or about 14% (as shown by the horizontal dotted line). Across all observers, there were 18% or fewer trials excluded, all of these being trials in which observer eye movements exceeded 21.6°/s.

Another way of demonstrating that observers were faithfully pursuing the target at all velocities is to look at the gain of the smooth pursuit. Figure 7 shows that the gain was close to 1, and so the observers were faithfully tracking the pursuit target, although with slightly reduced gain for the highest velocity, which is consistent with the slight reduction in the largest calculated retinal shifts given in Figure 6.

Discussion

Our results show a systematic relationship between retinal image motion and temporal perceptual grouping and provide direct evidence for retinal slip perceptually influencing observer responses by introducing a subtle shift in the arrangement of grid elements. The results show eye motion magnitudes to be a good predictor of the degree to which observers' responses are influenced by the temporal offset in presentation and indicate that compensation of retinal slip is lacking for briefly presented stimuli during smooth pursuit.

Using smooth pursuit eye movements of various known velocities allowed the retinal shift introduced between presentations of alternate grid elements to be precisely controlled. Previous work on perceptual grouping of grids has demonstrated the visual system's high sensitivity to changes in the ratio between horizontal and vertical interelement spacings (Ben-Av & Sagi, 1995), and our results are in good agreement: increasing horizontal spacings will increase the proportion of observer column responses. There is also some general quantitative agreement between our results and that of previous work. Ben-Av and Sagi (1995) showed that the percentage of vertical grouping responses increased with increasing spacing ratios (d_h/d_v): for a ratio of 1.2, the performance was 80% and rose to 95% for a ratio of 1.4. Although we measure percentage correct in a 2-AFC task, rather than percentage of a particular grouping response, we show a similar rise in performance when our results are converted into spacing ratios: 69% correct for a ratio of 1.15 and 88% for a ratio of 1.45. In particular, horizontal shifts between alternate column elements would induce changes in the spacing ratio, in which principles of grouping by proximity are likely to govern grouping (Kubovy, Holcombe, & Wagemans, 1998). However, for *row* presentations, the situation is somewhat different. When alternate row elements are horizontally offset by horizontal eye movements, grouping of the resulting grid may be governed by principles of good continuation, in which elements are continuous horizontally but are discontinuous vertically due to the horizontal shifts introduced between alternate rows.

The challenge with the Gestalt principles that govern perceptual organization is that they often lack quantification, making the statements above somewhat speculative (see Wagemans et al., 2012, for a comprehensive review of more recent quantitative work on perceptual grouping). However, of the studies that have attempted to quantify the effect of grouping cues such as proximity, similarity, and luminance on row/column grouping, the results suggest the possibility of a hierarchy in the parameters that facilitate grouping. The results also demonstrate that perceptual organization is time dependent, where grouping by proximity is a fast process (<60 ms), whereas similarity and luminance cues are perceived later (60–160 ms; Wertheimer, 1923; Elder & Goldberg, 2002). If proximity indeed dominates in grouping, it is possible that our results reflect a measure of whether the grid was column-like or not column-like, rather than column-like or row-like. This, however, does not affect our interpretation of the results in that retinal slip systematically influences perceptual grouping, as the results are still contingent on eye movements introducing a retinal image shift to alter the arrangement of grid elements presented in physical alignment and for this retinal shift to lack compensation.

The majority of trials in Wallis' perceptual grouping task recorded eye movement amplitudes between 0.01° and 0.06°. In his results, grouping performance for the small proportion of trials in which eye movement amplitude was 0.06° to 0.07° drastically deteriorated and no longer differed significantly from that for motion in the range of 0.00°–0.01°. This is in contrast with our results, in which grouping performance systematically improved with increasing magnitude of retinal shifts and continued to improve for retinal shifts of 0.06° and beyond. However, there are some key differences between the results of Wallis (2006) and ours, likely due to the former being based on post hoc extraction of eye movement amplitudes introduced by tremor, drifts, and microsaccades, while our results are based on retinal shifts introduced by smooth pursuit added to shifts due to fixation instability. Wallis (2006) posited that retinal shifts exceeding 0.06° were presumably introduced by more rapid microsaccadic or saccadic eye movements. The reduction in grouping performance was argued to be due to microsaccades being fully corrected for based on eye position information (i.e., extraretinal signals fully compensated for the retinal shift occurring between presentations in these trials, leading to a veridical percept of the grid in physical alignment). The possibility of image blur leading to a reduction in observers' capacity to resolve the spatial offsets induced by the eye movement was also proposed, although the experimental paradigm could not distinguish between the two scenarios, as both could lead to a reduction in performance. However, our results lend support for veridical perception of the grid and against image blur, and our results should similarly have demonstrated deterioration in performance for retinal shifts of 0.06° and beyond if image blur reduced our capacity to resolve spatial offsets induced by high-velocity eye movements. Our findings that observers remain sensitive to temporal offsets beyond 0.06° suggest that the perceptual system's access to extraretinal signals for microsaccades and smooth pursuit might differ as they serve two very different functions in visual perception.

Contrary to our hypothesis that grouping performance would be unaffected by velocity if compensation via extraretinal signals resulted in pursuit eye movements being fully compensated for, the results suggest this is not the case. Briefly presented stimuli can reveal the lower limit of compensation by retinal signals, as compensation must occur over a narrow but finite period, during which global motion is estimated (Wallis, 2006). Our results suggest that compensation for smooth pursuits may be constrained by the same temporal limits as fixational eye movements. Previous work (Festinger et al., 1976) attempted to quantita-

tively assess whether the retinal information, the efferent command, or some combination of both are used to compensate for smooth pursuit eye movements. The reported percept of an untracked stimulus was demonstrated to be closer in agreement with the retinal information rather than the actual eye movement during smooth pursuit (Festinger et al., 1976). There was little correspondence between perceived distance and the actual distance by which the eye moved, and the perceived orientation of an untracked spot was closer in orientation to the spot on the retina. The authors concluded that the efferent command for smooth pursuit contains good information about the direction of tracking but only crude information about speed at the stage at which it is monitored by the perceptual system.

There exists further support for the efferent command containing limited information for compensation of smooth pursuit eye movements. Stoper (1967) used judgments of relative spatial location of successive flashes to determine the extent to which extraretinal information and retinal information were used by the perceptual system and found observers' perception of relative spatial location to be almost completely determined by the retinal location of the flashes (Stoper, 1967). Fujii (1943) found that when observers tracked complex motions, the percept of the object path closely resembled the form of the actual retinal path, implying that perception is largely driven by the retinal image motion, and that little of the actual eye movement was taken into account by the perceptual system. The evidence from these studies supports the perceptual system's gross inaccuracies in its compensation for changes in eye position brought about by smooth pursuit and raises the possibility that central commands for pursuit eye movements are general, and lacking specific information. The results from the above studies, however, cannot be used directly to support the results presented here, as they were performed in complete darkness in the absence of other objects in the visual field—a visual environment that is very different from the testing setup used in the current experiments. Despite this, these studies—combined with our results—certainly point to retinal signals doing the heavy lifting in compensating for retinal slip arising from smooth pursuit for briefly presented stimuli and suggest that the same temporal constraints are imposed on the perceptual system when stimuli are briefly presented.

Perceptual grouping of temporally asynchronous stimuli demonstrates that fundamental perceptual processes can be influenced by briefly presented stimuli that are presented with an imperceptible temporal offset (Singer & Gray, 1995; Von der Malsburg, 1995). Our results suggest that eye movements cannot be ignored when attempting to explain perceptual grouping of temporally asynchronous stimuli. Even with good fixation and head restraint, temporal offsets can be transformed into apparent spatial shifts (Dakin & Bex, 2002).

The role of onset and offset transients in grouping performance

Dakin and Bex (2002) showed that onset and offset transients are a significant cause of the improved ability to extract contours when contour and background Gabor elements were presented on alternate movie frames, rather than synchronously on the same frame. It is highly unlikely that such transient effects are responsible for the grouping seen in our experiment, however. The onset and offset transients in our flashed stimulus are fixed, and so grouping performance should remain constant despite changes in pursuit velocity. In contrast, we find a steady improvement in grouping as velocity increases, indicating that retinal shift magnitude is driving the grouping effect. Of note is that our function relating retinal shift to grouping performance increases linearly from our nominal zero shift (fixation) condition (Figure 5). If stimulus transient effects were dominant, one would expect an initial plateau in this function until such time that retinal shifts were of a sufficient magnitude to dominate behavior. The absence of this signature suggests that retinal shifts due to eye movements are the critical determinant of grouping performance in our task, even under conditions of steady fixation. Dakin and Bex (2002) argued that the visibility of onset and offset transients was dependent on high orientation uniformity in their contour stimuli, allowing these contoured elements to be distinguished from the nonuniform background orientations by a transient, orientation-bandpass filter. The spatial characteristics of our stimulus—in which the orientation information on alternate frames are identical—would not be amenable to such processing.

Conclusion

The present study provides direct evidence for the role of retinal slip in grouping of temporally asynchronous stimuli. A systematic relationship was found between the magnitude of retinal slip and observers' sensitivities to the temporal asynchrony in the presentation of alternate grid elements for a row/column grouping task. The findings suggest that perceptual compensation of smooth pursuit eye movements is primarily achieved by retinal signals, with poor compensation via extraretinal signals, and that the ability to use retinal signals is thwarted when stimuli are extremely brief. The study highlights the perceptual

influence of eye movements and the importance of considering eye movements as a confounding factor in studies that use temporally asynchronous stimuli.

Acknowledgments

References

Arathorn, D. W., Stevenson, S. B., Yang, Q., Tiruveedhula, P., & Roorda, A. (2013). How the unstable eye sees a stable and moving world. *Journal of Vision*, 13(10):22, 1–19, https://doi.org/10.1167/13.10.22. [PubMed] [Article]

Ben-Av, M. B., & Sagi, D. (1995). Perceptual grouping by similarity and proximity: Experimental results can be predicted by intensity autocorrelations. *Vision Research*, 35, 853–866.

Bridgeman, B. (1995). A review of the role of efference copy in sensory and oculomotor control systems. *Annals of Biomedical Engineering*, 23, 409–422.

Burak, Y., Rokni, U., Meister, M., & Sompolinsky, H. (2010). Bayesian model of dynamic image stabilization in the visual system. *Proceedings of the National Academy of Sciences, USA*, 107, 19525–19530.

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26, 404–413.

Coren, S., Bradley, D. R., Hoenig, P., & Girgus, J. S. (1975). The effect of smooth tracking and saccadic eye movements on the perception of size: The shrinking circle illusion. *Vision Research*, 15, 49–55.

Dakin, S. C., & Bex, P. J. (2002). Role of synchrony in contour binding: Some transient doubts sustained. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 19, 678–686.

Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*, 2(4):5, 324–353, https://doi.org/10.1167/2.4.5. [PubMed] [Article]

Engbert, R., Mergenthaler, K., Sinn, P., & Pikovsky, A. (2011). An integrated model of fixational eye movements and microsaccades. *Proceedings of the National Academy of Sciences, USA*, 108, E765–E770.

Festinger, L., Sedgwick, H. A., & Holtzman, J. D. (1976). Visual perception during smooth pursuit eye movements. *Vision Research*, 16, 1377–1386.

Freeman, T. C., Champion, R. A., & Warren, P. A. (2010). A Bayesian model of perceived head-centered velocity during smooth pursuit eye movement. *Current Biology*, 20, 757–762.

Fujii, E. (1943). Forming a figure by movement of a luminous point. *Japanese Journal of Psychology*, 18, 196–232.

Herrmann, C. J. J., Metzler, R., & Engbert, R. (2017). A self-avoiding walk with neural delays as a model of fixational eye movements. *Scientific Reports*, 7, 12958.

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology*, 35, 71–98.

Mon-Williams, M., & Tresilian, J. R. (1998). A framework for considering the role of afference and efference in the control and perception of ocular position. *Biological Cybernetics*, 79, 175–189.

Murakami, I. (2003). Illusory jitter in a static stimulus surrounded by a synchronously flickering pattern. *Vision Research*, 43, 957–969.

Murakami, I. (2004). Correlations between fixation stability and visual motion sensitivity. *Vision Research*, 44, 751–761.

Murakami, I., & Cavanagh, P. (1998, October 22). A jitter after-effect reveals motion-based stabilization of vision. *Nature*, 395, 798–801.

Pitkow, X., Sompolinsky, H., & Meister, M. (2007). A neural computation for visual acuity in the presence of eye movements. *PLoS Biology*, 5, e331.

Poletti, M., Listorti, C., & Rucci, M. (2010). Stability of the visual world during eye drift. *Journal of Neuroscience*, 30, 11143–11150.

Rashbass, C. (1961). The relationship between saccadic and smooth tracking eye movements. *Journal of Physiology*, 159, 326–338.

Royden, C. S., Banks, M. S., & Crowell, J. A. (1992, December 10). The perception of heading during eye movements. *Nature*, 360, 583–585.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, 18, 555–586.

Sommer, M. A., & Wurtz, R. H. (2008). Brain circuits

for the internal monitoring of movements. *Annual Review of Neuroscience*, *31*, 317–338.

Souman, J. L., Hooge, I. T., & Wertheim, A. H. (2006). Localization and motion perception during smooth pursuit eye movements. *Experimental Brain Research*, *171*, 448–458.

Stoper, A. E. (1967). *Vision during pursuit movement: The role of oculomotor information* (Unpublished doctoral dissertation). Waltham, MA: Brandeis University.

Sun, L. D., & Goldberg, M. E. (2016). Corollary discharge and oculomotor proprioception: Cortical mechanisms for spatially accurate vision. *Annual Review of Vision Science*, *2*, 61–84.

Tulunay-Keesey, U., & VerHoeve, J. N. (1987). The role of eye movements in motion detection. *Vision Research*, *27*, 747–754.

Usher, M., & Donnelly, N. (1998, July 9). Visual synchrony affects binding and segmentation in perception. *Nature*, *394*, 179–182.

van Beers, R. J., Wolpert, D. M., & Haggard, P. (2001). Sensorimotor integration compensates for visual localization errors during smooth pursuit eye movements. *Journal of Neurophysiology*, *85*, 1914–1922.

Von der Malsburg, C. (1995). Binding in models of perception and brain function. *Current Opinion in Neurobiology*, *5*, 520–526.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, *138*, 1172–1217.

Wallis, G. (2005). A spatial explanation for synchrony biases in perceptual grouping: Consequences for the temporal-binding hypothesis. *Perception & Psychophysics*, *67*, 345–353.

Wallis, G. (2006). The temporal and spatial limits of compensation for fixational eye movements. *Vision Research*, *46*, 2848–2858.

Ward, F. (1976). Pursuit eye movements and visual localization. *Eye Movements and Psychological Processes*, 1976.

Wertheimer, M. (1923). Studies concerning the theory of shape. *Psychologische Forschung*, *4*, 301–350.