Reflexive and voluntary control of smooth eye movements

Jeffrey B. Mulligan[a], Scott B. Stevenson[b] and Lawrence K. Cormack[c]

[a]NASA Ames Research Center, MS 262-2, Moffett Field CA 94035, USA
[b]University of Houston College of Optometry, 4901 Calhoun Rd., Houston TX 77204, USA
[c]University of Texas at Austin, SEA 4.328A, Mailcode A8000, Austin TX 78712, USA

ABSTRACT

An understanding of visually evoked smooth eye movements is required to predict the visibility and legibility of moving displays, such as might be encountered in vehicles like aircraft and automobiles. We have studied the response of the oculomotor system to various classes of visual stimuli, and analyzed the results separately for horizontal and vertical version (in which the two eyes move together), and horizontal and vertical vergence (where they move in opposite directions). Of the four types of motion, only vertical vergence cannot be performed under voluntary control, and certain stimuli (all having relatively long latencies) are incapable of evoking it. In another experiment, we instructed observers to track one of two targets, and measured weak but reliable responses to the unattended target, in which the long-latency component of the response is abolished. Our results are consistent with a system containing two distinct processes, a fast reflexive process which responds to a restricted class of stimuli, and a slower voluntary process capable of following anything that can be seen, but incapable of controlling vertical vergence.

Keywords: smooth pursuit, optokinetic reflex, vergence, second-order motion

1. INTRODUCTION

A variety of computational models have been developed to predict visibility and legibility of text and symbols on electronic displays. The vast majority of this work, however, has been performed in the context of a stationary observer viewing a stationary display (i.e., a typical office environment). Generalizing these models to the case of a moving vehicle requires an estimate of the amount of retinal image motion, both that due to relative motion between the vehicle and the observer, as well as that produced by the resulting eye movements.

The human eye is constantly in motion, even when it appears to be trained on a stationary object, a special type of "movement" known as a *fixation*. Much eye movement behavior consists of a series of fixations to different objects connected by rapid jumps of gaze direction known as *saccades*. Sometimes the object of interest is not stationary, but moving, and the eye is required to move accordingly to keep the image of the object centered on the fovea; this type of movement is referred to as *smooth pursuit*. The eye also moves smoothly in response to vestibular stimulation (the *vestibulo-ocular reflex* or VOR) and full field visual motion (*optokinesis*, or the *optokinetic reflex*). When full field motion continues in a single direction for an extended period of time, the eye tracks the motion using a sawtooth-like motion known as *optokinetic nystagmus* (OKN), consisting of a "slow phase" in which the motion is tracked as in optokinesis, and a "fast phase" resembling a small saccade in which the eye position is reset. A good overview of these phenomena and the underlying physiology can be found in a number of texts.[1, 2]

The vestibulo-ocular reflex (VOR) aids in the fixation of stationary objects in the presence of head movement, by driving the eyes in response to head motion as sensed by the otoliths and semi-circular canals of the inner ear. In the case where the motion of the body is produced by the motion of a vehicle, with no relative motion between the head and the vehicle display, the compensatory motion produced by the VOR alone will be inappropriate, and the retinal image will be degraded unless vision overrides the VOR. On the other hand, when there *is* relative

Further author information: (Send correspondence to J.B.M.)
J.B.M.: E-mail: jeffrey.b.mulligan@nasa.gov, Telephone: 1-650-604-3745
S.B.S.: E-mail: sbstevenson@uh.edu, Telephone: 1-713-743-1960
L.K.C.: E-mail: cormack@psy.utexas.edu, Telephone: 1-512-471-5380

motion between the observer and the vehicle, vision will, in general, be the only source of information that can be used to move the eyes in a way to stabilize the retinal image. Thus, to understand visibility in a dynamic environment, it is necessary to have an understanding of visually-driven eye movements.

Visually-driven eye movements can occur with or without the application of voluntary effort. Pola and Wyatt[3] studied the effect of voluntary effort in "open-loop" conditions, in which the motion of the eye had no effect on the position of the retinal stimulus, and concluded that the mechanisms involved in "passive" involuntary movements were also reflected in "active" voluntary movements. Using a different approach, Kowler and colleagues[4] studied the voluntary selection of a smooth pursuit target in the presence of other visual motions, and found that the small targets could be selected and tracked accurately, independent of other motions. In this paper we will revisit these issues using a new method we call the *eye movement correlogram.*

The two eyes generally move together, but not always by equal amounts. *Vergence* refers to disconjugate (differential) movements of the eyes, while *version* refers to conjugate (common-mode) movements. Horizontal vergence is required to fixate on objects at different depths, and (with a little practice) can be performed voluntarily. Small vertical vergence movements are required when scanning objects that are to one side of the head-relative line-of-sight (known as *primary gaze position*), but, unlike horizontal vergence, vertical vergence cannot be performed voluntarily. When a demand for vertical vergence is introduced, say by the introduction of a prism in front of one eye, the eyes will slowly adopt a posture that brings the left and right eye images into correspondence. The vertical vergence system thus provides us with a window into a purely reflexive system uninfluenced by voluntary processes.

A large body of prior work has attempted to elucidate the physiological underpinnings of voluntary and reflexive eye movements. Using fMRI, it has been shown that the same neural circuits are responsible for smooth pursuit and optokinetic nystagmus (OKN), although slightly different patterns are seen for active "look" OKN and passive "stare" OKN.[5]

2. THE EYE-MOVEMENT CORRELOGRAM

The *eye movement correlogram* (EMC) is an analysis method we have developed to study the latency of visually-driven eye movements with respect to the visual stimulus. Its first use was to compare eye movement latencies to chromatic and achromatic stimuli,[6] using a video ophthalmoscope with a sampling rate of 60 Hz.[7] The EMC has been used subsequently in a number of investigations employing the Dual Purkinje Image (DPI) eye tracker,[8,9] some of which are summarized in this paper.

2.1 Description of the method

The correlogram is obtained by presenting the subject with a visual stimulus moving randomly, and recording the resulting eye movements. The subject is usually (but not always) instructed to follow the track the stimulus movement with their gaze. The subsequent analysis is intimately related to the first-order linear kernel used in nonlinear systems analysis.[10,11] In both cases, the result is obtained by computing the cross-correlation of the input signal with the output signal. In kernel analysis, the goal is to derive a waveform which predicts the output when convolved with the input; for a linear system, this is simply the impulse response. Our analysis differs in that we compute the *normalized* cross-correlation: the result is divided by the norms of the inputs to the calculation, producing a result which is always between 1 and -1. This produces a cleaner result (i.e., a higher signal-to-noise ratio) when averaging across many trials. The normalization loses all information concerning *pursuit gain*: the correlogram will be the same when the eye follows a target perfectly as when it is nearly stationary but tracks the stimulus with a tiny pursuit gain. The peak value represents the degree to which the signal is a faithful (possibly scaled and delayed) replica of the input.

To generate the target motions, we start with equal energy white velocity noise *. Independent signals are generated for the horizontal and vertical directions. These signals are then Gaussian low-pass filtered with a cut-off frequency of 10 Hz, and integrated to obtain a random walk position signal. A scaling factor is applied

---

*Early work used uniform random noise, but most of our experiments have been performed using trajectories derived from Gaussian-distributed white velocity noise, in order to insure complete isotropy. Note that the subsequent low-pass filtering reduces the differences between the two noise distributions.
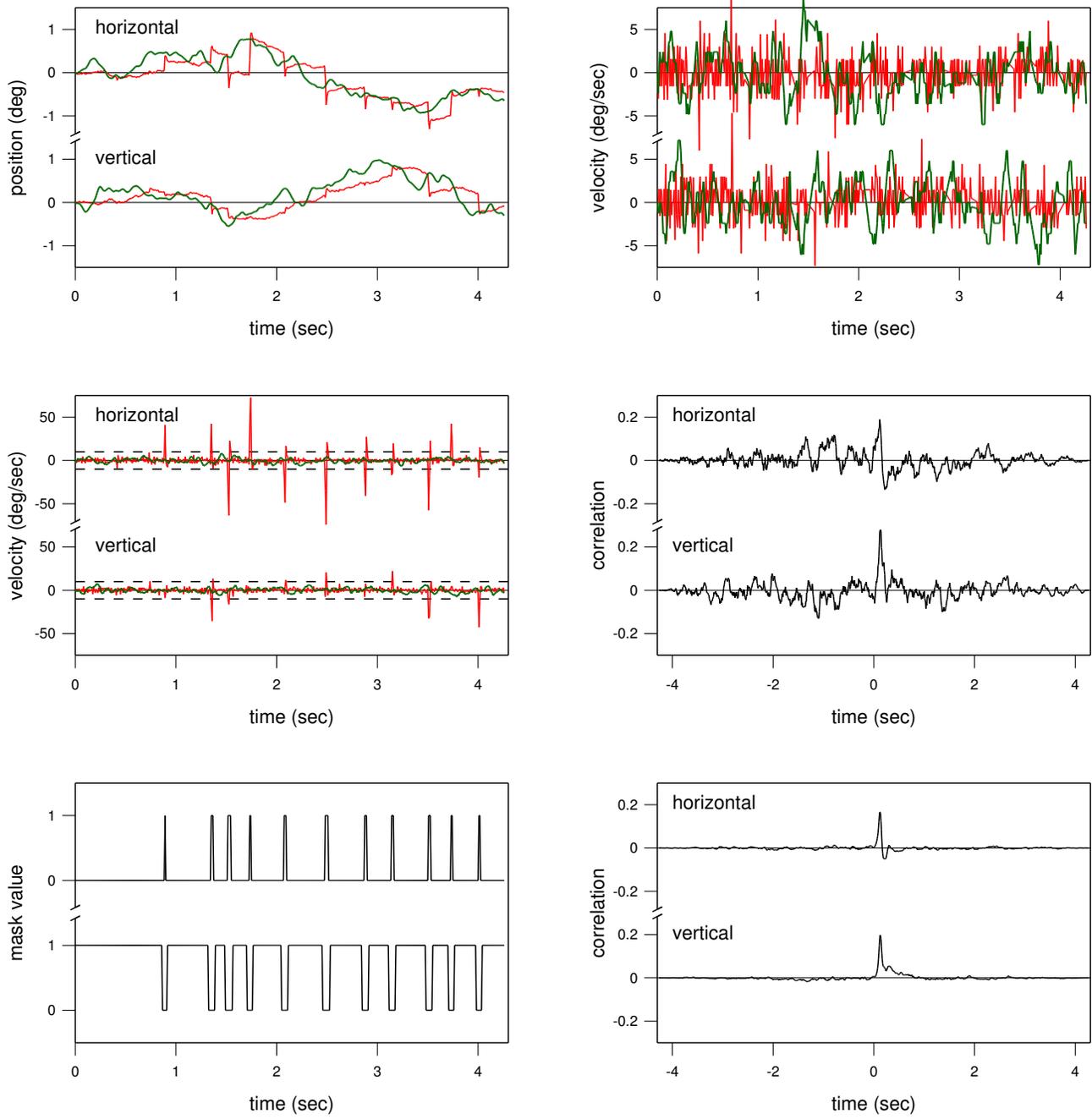
Figure 1: Illustration of the steps in the computation of the velocity correlogram. See text for complete description. Upper left: Eye position signals (red) plotted with stimulus position (green), with the horizontal signals depicted by the upper traces and the vertical by the lower. Center left: Eye and stimulus velocity signals obtained by differentiation; the dashed horizontal lines indicate 10 degree per second threshold for saccade detection. Lower left: Saccade signal and smooth velocity mask. Upper right: Eye velocity after saccade removal (red), plotted with stimulus velocity (green). Center right: Normalized cross correlation of eye and stimulus velocities. Lower right: Velocity correlograms obtained by averaging 50 four-second trials. The horizontal correlogram (upper trace) exhibits a phase of negative correlation following the strong positive peak, while the vertical correlogram (lower trace) does not.
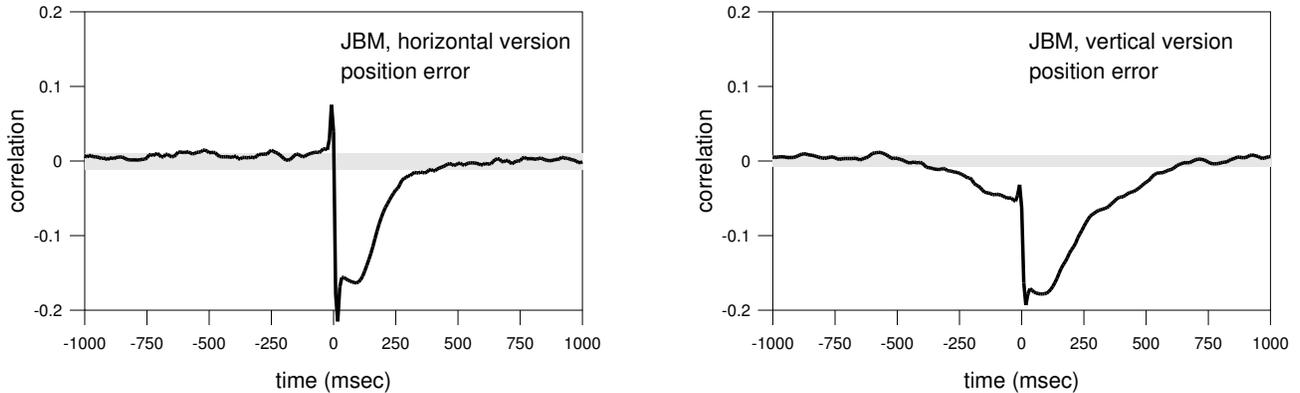
Figure 2: Saccade correlograms for subject JBM tracking spot target.

to the filtered velocity so that the majority of trajectories are bounded by ± 2 degrees. Approximately 2% of the trajectories computed in this way exceeded 2 degrees at some point, and these are discarded.

In different experiments, we record subjects' eye movements in response to different types of stimuli. We analyze the smooth components of the eye movements separately from the saccades. Saccades are detected using a velocity threshold; this is not reliable in general, because the velocity ranges of saccades and smooth pursuit overlap. In the context of our tracking task, however, the distributions do not overlap, and we obtain robust results using a threshold of 10 degrees per second.

Figure 1 shows the data from a typical record of a four second trial. The upper left panel shows the horizontal and vertical eye gaze positions, in the upper and lower traces, respectively. These are plotted together with the corresponding target positions (in green). It can be seen that most of the positional error is corrected by saccades, but that the eye is not still between saccades. The center left panel shows the velocities of the eye and stimulus, calculated by differentiation of the traces shown in the first panel. The horizontal dashed lines above and below each trace represent the velocity threshold for saccade detection (10 deg/sec). The lower left panel shows two binary signals related to saccade detection and removal: the upper waveform has a value of one for those time bins where the eye velocity exceeds the threshold, and zero otherwise. The pulses in this trace are fattened by three samples (on each side) to ensure that the entire saccade is excluded. We refer to the resulting signal as the *saccade mask*. To select the smooth components of the movement, the saccade mask is complemented, producing the *smooth velocity mask*, shown in the lower trace.

The smooth velocity is computed by taking the raw eye velocity for those samples where the mask has a value of one, and using an interpolated value for those samples where the mask has a value of zero. To minimize the effects of noise, the smooth velocities used to fill saccadic intervals are interpolated between the averages of the five samples immediately before and after the saccade. The resulting de-saccaded velocities are shown in the upper right hand panel of figure 1. In some of our work, a low-pass filter with a cut-off frequency of 10 Hz is applied at this stage, which provides a modest noise reduction. (This example illustrates the processing with no filtering.)

We compute two types of correlograms: the *velocity correlogram* is obtained via the normalized cross correlation of the smooth eye velocity with the stimulus velocity; the *saccade correlogram* uses the saccade velocities, with the intervening smooth velocities masked to a value of zero. These are correlated with the position error, i.e. the difference between the stimulus position and the eye position.

The smooth velocity correlation for a single four second record is shown in the center right panel of figure 1. These signals exhibit a high level of noise, as indicated by non-zero correlations at negative values of the delay variable (which would imply the eye predicted the random motion of the target). The lower right panel of figure 1 shows the average of 50 such trials, representing about five minutes of data collection. As expected, we see a positive peak with a latency of around 100 milliseconds.
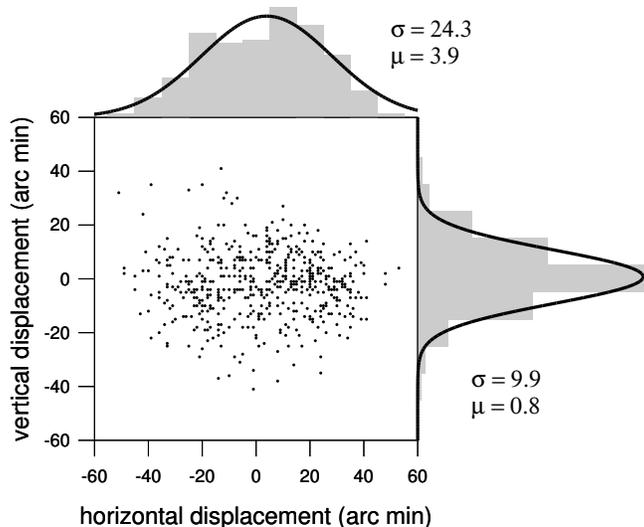
Figure 3: Saccade histogram for the example data set.

Saccade correlograms are shown in figure 2. In order to interpret these figures, some explanation is required. First, we note that the input variable, position error, is defined as eye position minus target position. A positive horizontal position error means that the eye is to the right of the target, which requires a saccade to the left to correct, resulting in a spike of negative horizontal velocity. Thus, a corrective saccade will produce a negative correlation at positive time lags. (Remember that the time axis represents the time of the eye movement relative to a stimulus event.) Negative time lags correspond to position errors *after* the saccade; if the saccade corrects perfectly, the position error (and hence the correlation) should be zero. Unlike the velocity correlogram, where the output variable (eye velocity) can be independent of the input variable (stimulus velocity), the input variable for the saccade correlogram (position error) is necessarily altered by the execution of a saccade (the output variable). This is reflected in the sharp transition seen at time zero. The knee in the curves at around 100 milliseconds indicates the time when the decision to make the saccade occurs. Motions of the target between the time of the decision and the execution of the saccade do not influence the saccade, and so the correlation remains constant or decreases slightly. The spike-like features bracketing the transition at time 0 are the result of the fact that the output of the dual Purkinje image tracker responds to saccade-induced motions of the eye's crystalline lens,[12, 13] which produce ringing-like features at the start and end of most saccades.

The waveforms are similar for horizontal and vertical saccades, with the exception of the negative correlations seen at negative time lags in the vertical signal. Because negative correlations indicate a saccade in the direction opposite to the position error (i.e., a corrective saccade), the negative correlations at negative time lags indicate that the vertical saccades are *hypometric*, e.g. they do not fully-correct the position error. We think that this may account for the difference seen between the horizontal and vertical velocity correlograms, where a positive "shoulder" is seen in the vertical trace but not the horizontal. Large horizontal errors tend to be corrected completely by saccades and do not need subsequent help from the pursuit system, while vertical errors are only partially corrected by saccades, which the pursuit system continues to try to correct.

The gray bars represent the noise level of the measurement, and span a range of ± 3 standard deviations of the baseline correlation values, taken from the range of lags from -2 to -1 seconds. For the velocity correlograms, where significant non-zero correlations are never seen for negative lags, we use the interval -1 to 0 seconds. (The choice of the range is a compromise between using as many samples as possible, and avoiding samples from the extreme lags where the noise level is reduced because of the smaller number of time bins contributing to the cross-correlation.)

The distribution of saccade vectors for this data set is shown in figure 3. Each saccade is transformed to a
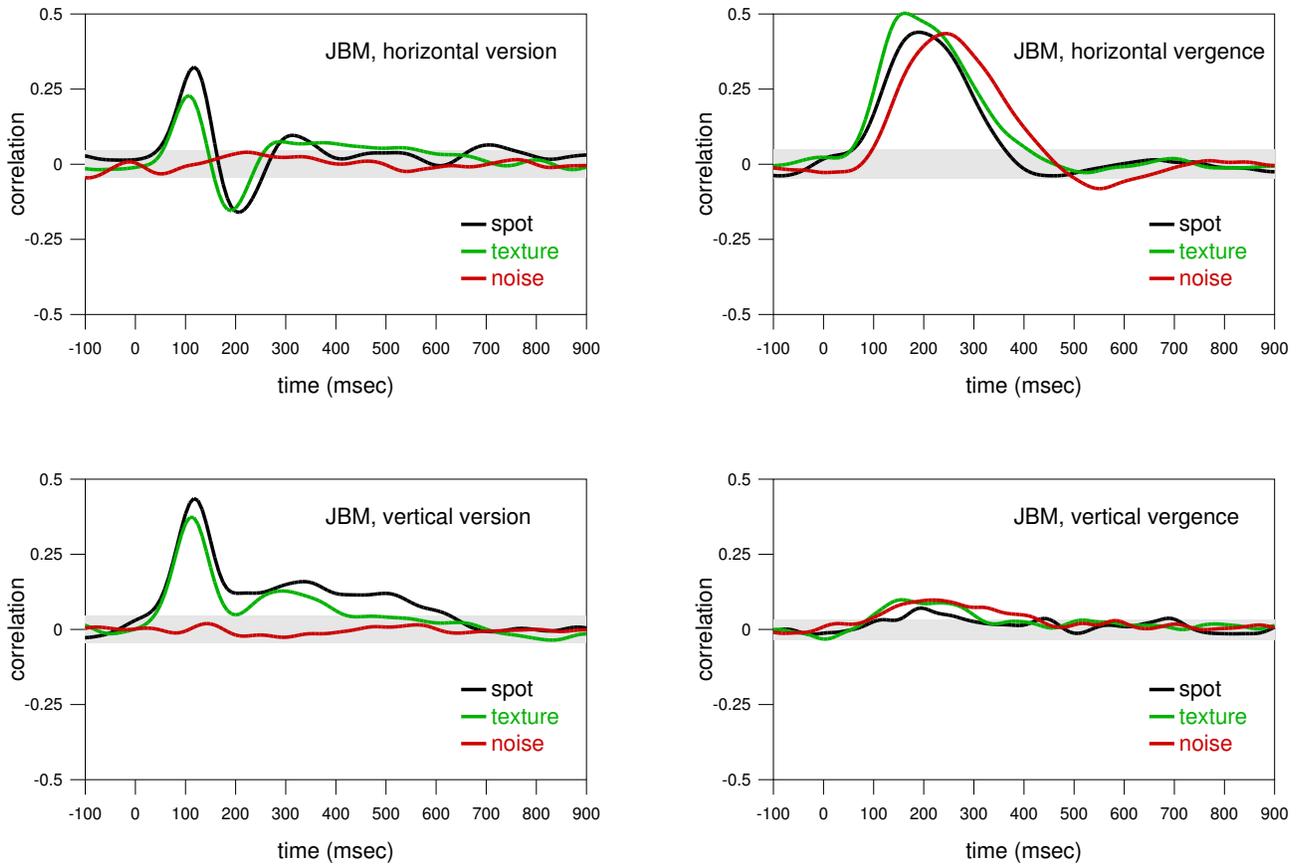
Figure 4: Velocity correlograms for subject JBM in response to three different stimuli. See text for details.

vector by taking the difference between the starting and ending positions. All of the saccades from this set of trials have an amplitude of less than one degree. Although saccades are observed in all directions, the horizontal amplitudes tend to be larger. This has been quantified by computing histograms of the horizontal and vertical components (depicted by the bar graphs at the right and top edges of the scatter plot), using a bin width of 10 minutes of arc. These histograms have in turn been fit with Gaussian distributions (plotted over the bar graphs of the histograms). The fit to the distribution of horizontal components has a standard deviation of 24.3 minutes of arc and a mean of 3.9 minutes, while the corresponding values for the vertical components are 9.9 and 0.8. The smaller amplitude in the vertical dimension is consistent with the observation made above (on the basis of the correlogram shown in figure 2) that the vertical saccades are hypometric. We believe that this may also account for the difference in the shapes of the horizontal and vertical velocity correlograms: the vertical velocity correlogram remains positive for all lags because the saccades are insufficient to get all of the way to the target, and the smooth pursuit system must make up the difference. In the horizontal dimension, on the other hand, the saccades are efficient at eliminating position errors, giving the smooth pursuit system the chance to overshoot the target, producing the negative lobes seen in the velocity correlograms. This pattern has been observed in all of the subjects tested to date.

2.2 Representative results for various stimuli

Figure 4 shows correlograms for subject JBM in response to three types of stimuli: a bright spot (produced by shining a laser pointer at the otherwise dark screen); a full-field "melting" texture (designed to minimize stable positional cues which might draw saccades); and binocularly correlated dynamic random noise, which produces no discernible motion when viewed monocularly, but provides a strong stimulus to vergence. In all cases, the

subject was actively looking at the display, although only the spot target provided a well-defined positional cue. Although we eliminated persistent positional cues with the "melting" texture, there were random features, persisting with durations approaching 1 second, which carried the velocity information. We suspect that the main difference between the spot and the texture stimuli was simply the amount of retinal area stimulated.

Because the dynamic noise stimulus is binocularly, but not temporally correlated, it provides only one of two cues to motion-in-depth, namely *changing disparity over time* (CDOT), while providing no *inter-ocular velocity difference* (IOVD). On the other hand, the spot and texture stimuli each provide both cues. The horizontal vergence responses in figure 4 indicate that responses mediated by CDOT alone are delayed by at least 50 milliseconds compared to responses mediated by both CDOT and IOVD, although the pattern is less clear for vertical vergence where the signals are weaker.

Correlogram analysis as described here has many similarities to the linear kernel analysis applied to the smooth pursuit responses to perturbations of the speed of a moving target by Tavassoli and Ringach.[14] Their results, however, differ from ours in that their kernels show no negative lobes, whereas we consistently find negative lobes for horizontal version (as seen in the upper-left panel of figure 4), and similar oscillations riding on a positive "shoulder" for vertical version. Tavassoli and Ringach speculated that the negative lobes seen in our results might be a by-product of the filtering applied to our signals, but we obtain the same results whether or not filtering is applied. Furthermore, when low-pass filtering *is* performed, it is with a Gaussian filter that does not produce any ringing. The damped oscillations seen in the correlogram signals are not at a fixed frequency, but vary depending on the stimulus; in general, stimuli with longer delays to the first correlogram peak have longer oscillation periods, suggesting that the oscillations are produced by a feedback loop involving the delay. Such oscillations have been produced in monkeys,[15] and we have replicated the effect in humans.[16] Ringing such as we observe in our correlograms might be produced by responses overshooting the target, necessitating a corrective movement in the opposite direction. Perhaps the fact that the perturbations studied by Tavassoli and Ringach were imposed on top of an average velocity in one direction (tracked with a pursuit gain of less than one) is responsible for the difference in our results.

3. MULTIPLE TARGETS AND THE EFFECT OF ATTENTION/EFFORT

A distinguishing feature of smooth pursuit is the ability to select a single target from a number of moving objects and follow the chosen object while ignoring the irrelevant stimuli.[4] Here we ask, to what extent it is possible to suppress the eye movement response to an irrelevant target.

3.1 Methods

The stimuli consisted of a pair of initially concentric circles, presented on a 10 degree field (see fig 5), generated with Microsoft Visual Basic and a VSG board (Cambridge Research Systems, http://www.crsltd.com/), and displayed on an Image Systems monitor at 120 Hz (http://www.imagesystemscorp.com/). The circles were 1 and 2 degrees in diameter and the thin lines forming the rings were dark on a bright background of 24 candelas per meter squared, with a contrast near 100%. Image motion was produced by redrawing the rings on each frame. Rings for the left and right eyes were presented side by side on a single monitor. Ten degree diameter apertures in the viewing path served to isolate the left and right targets to the appropriate eye. Targets were viewed through the stimulator optics of the eye tracker, and focus and vergence angle were adjusted to far focus and phoria for each subject. The caption of figure 5 contains a link to a short movie illustrating the stimuli, which depicts several trials as they appeared to the subjects, followed by several trials rendered with fixation cursors depicting one subject's gaze locations for three replications of each of the two attention conditions.

The motion trajectories were produced by low-pass filtering white velocity noise at 10 Hz and then integrating to generate position information. Independent trajectories were used for horizontal, vertical, vergence and version axes, and for the large and small rings. Thus, a total of eight independent signals were generated for each 4 second trial. The total excursion of the targets varied from trial to trial, but never exceeded two degrees.

Eye movements were recorded binocularly with an SRI dual-Purkinje image eye tracker[8,9] (now sold and serviced by Fourward Technologies http://www.fourward.com). Left and right horizontal and vertical eye positions were sampled at 120 Hz in synchrony with the video frames. Version eye position was calculated as
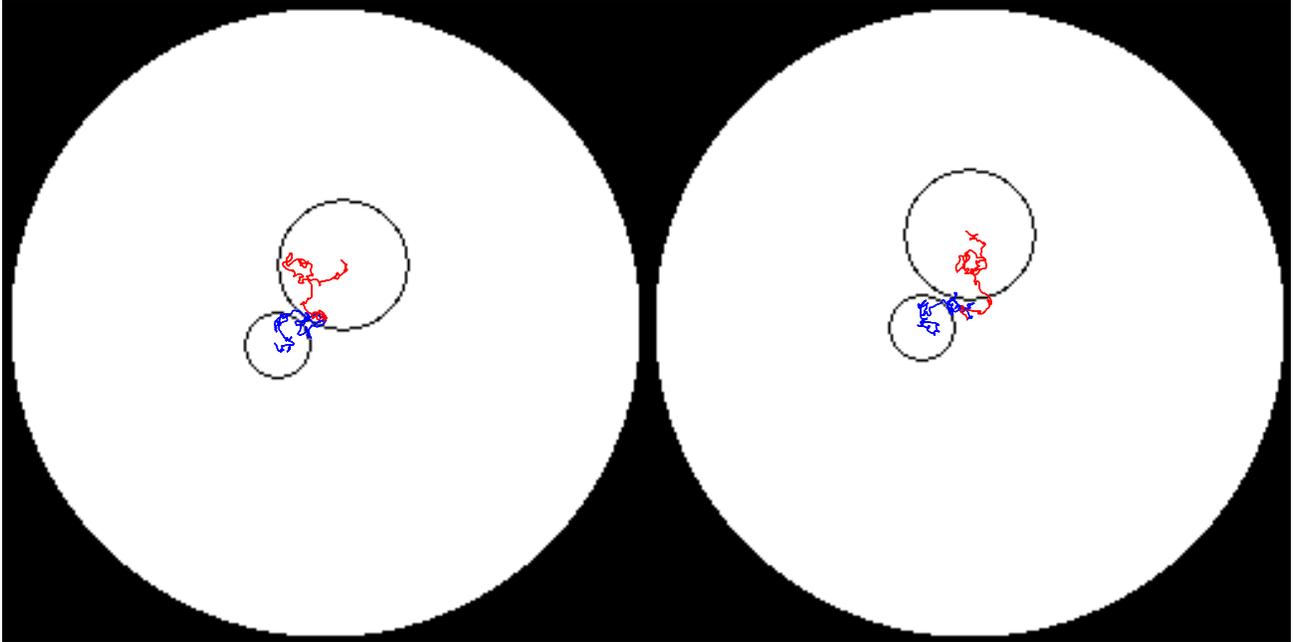
Figure 5: Representation of the dual-ring stimulus. The large white disks represent the 10 degree field stops in the eye tracker viewing optics, and the small black circles show the appearance of the two rings at the end of the four second trial. The red and blue squiggles represent the paths of the centers of the large and small rings, respectively; these are displayed here to illustrate the motion paths but were not part of the stimuli. The two fields are the left- and right-eye images. The two images may be free-fused to get an idea of the binocular stimulus. An animated video version is available at http://dx.doi.org/doi.number.goes.here.

the arithmetic mean of left and right eye positions, and vergence eye position was calculated as the difference between left and right eye positions. Eye position was differentiated to velocity and saccades were removed by zeroing any part of the trace where velocity exceeded 8 degrees per second. The smooth tracking velocity data were cross-correlated to the target velocity to produce a velocity correlogram for each target and each of the four motion axes (horizontal/vertical, vergence/version). Fifty unique trajectories were presented from six to twelve times per subject and the resulting correlograms were averaged together. The two rings were alternately designated as "track" or "ignore" for one block of 50 trials.

The subjects were two of the authors and five other healthy, visually normal individuals from the University of Houston College of Optometry. Degree of experience as a subject or observer in vision experiments varied from none to more than 25 years. All subjects gave informed consent and all procedures were reviewed and approved by the University of Houston Committee for the Protection of Human Subjects.

For a given block of 50 trials, subjects were instructed to track the center of one target, and ignore the other. Each subject performed three to six such blocks for each of the two targets.

3.2 Results

The instruction condition made a substantial difference in the responses. Figure 6 shows velocity correlograms for the attended and ignored rings for one of the seven subjects (SBS). The traces in figure 6 are the average of 6 blocks of 50 trials; results are similar for the remaining subjects who ran only 3 blocks, with the exception of a slightly higher noise level.

A striking result from this experiment is that attention makes little or no difference to the overall shape of the vertical vergence response. The absolute level of correlation was higher for the attended target, but the response pattern is nearly the same for both.
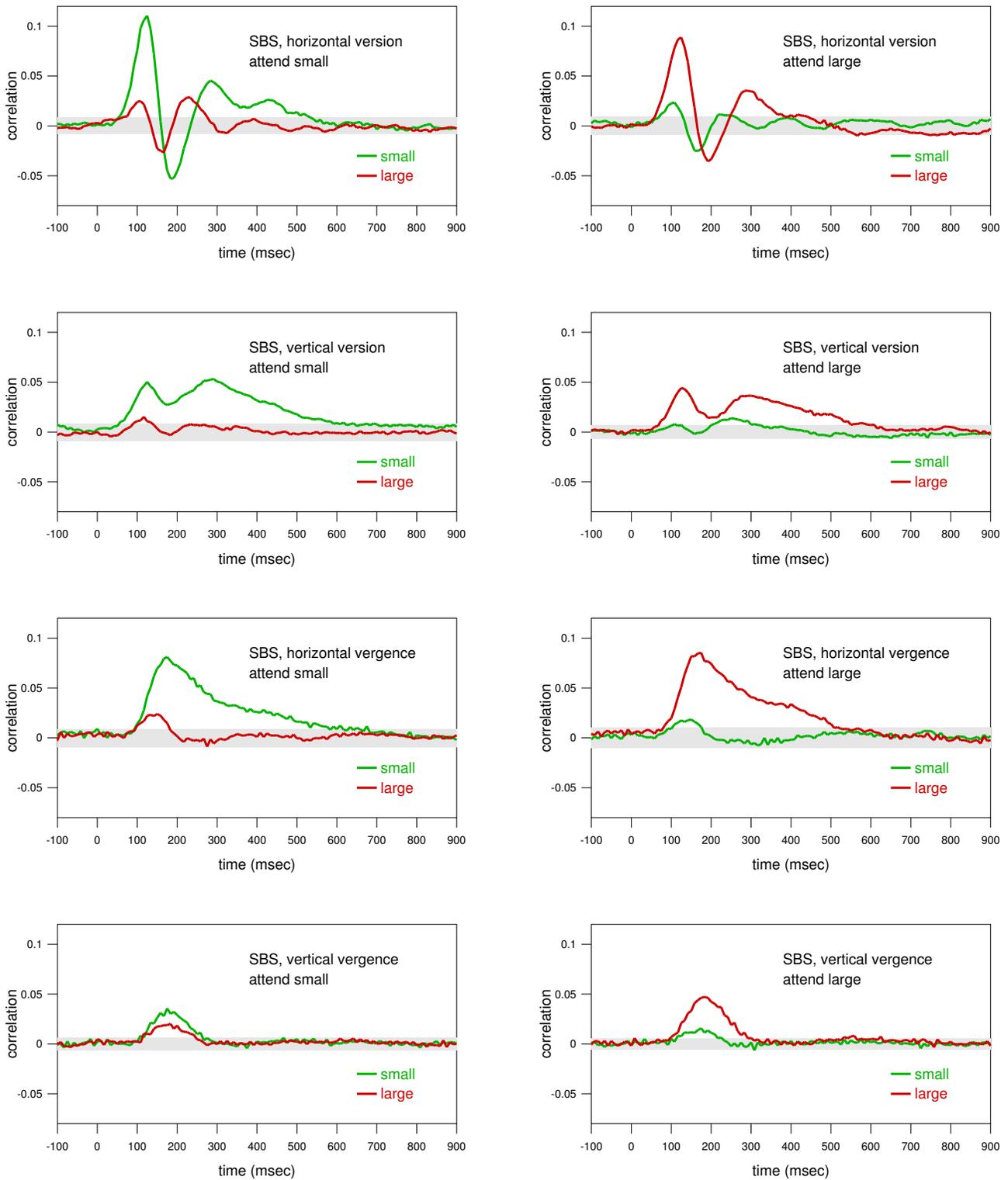
Figure 6: Velocity correlograms for subject SBS in the dual ring target experiment. See text for complete details. Plots in the left-hand column depict results for the "attend small" condition, while the right-hand column shows the "attend large" results.
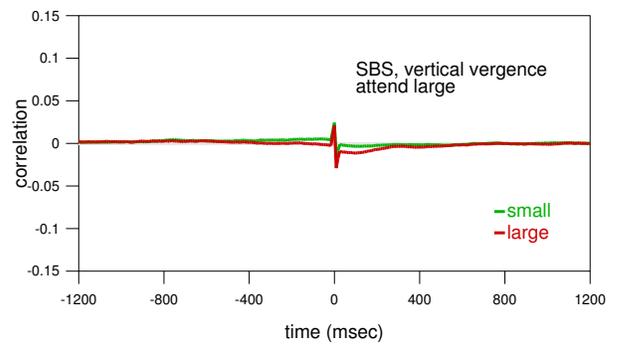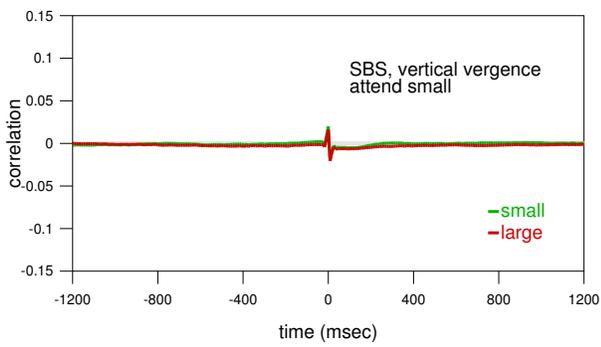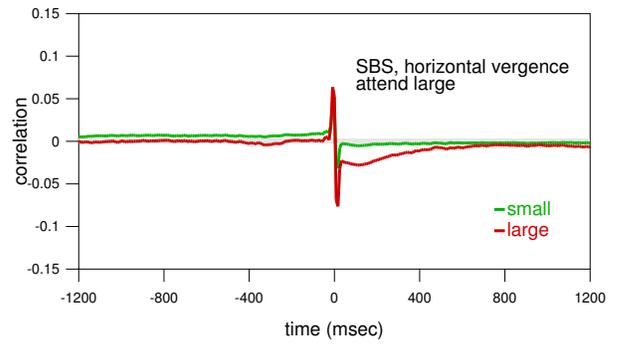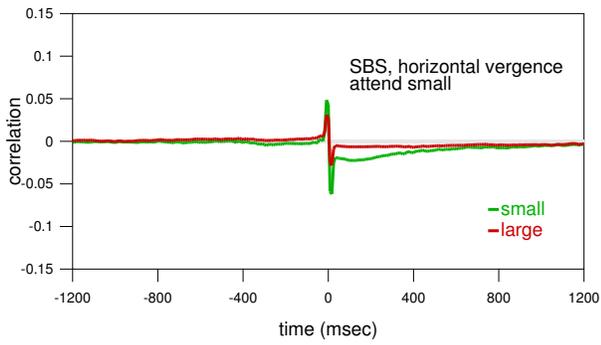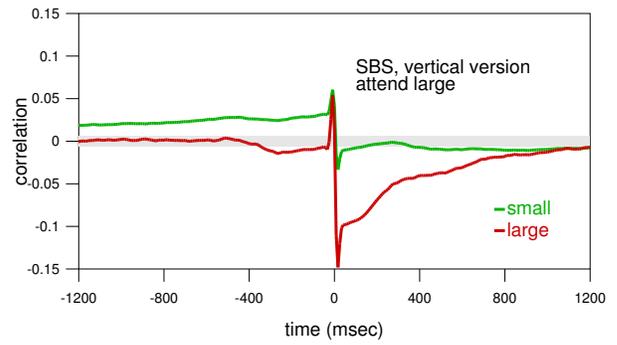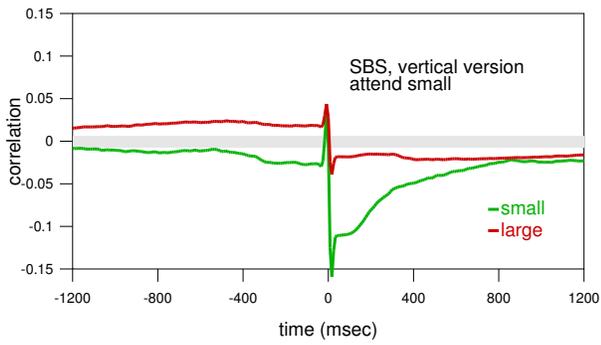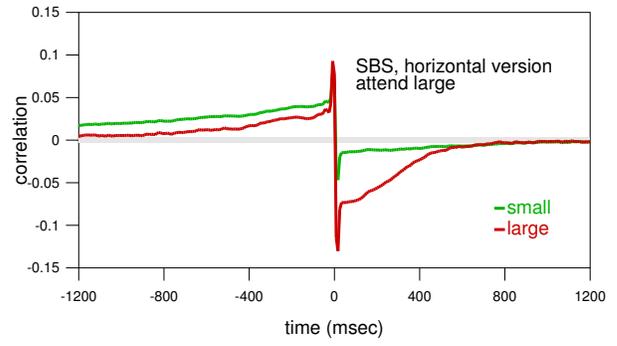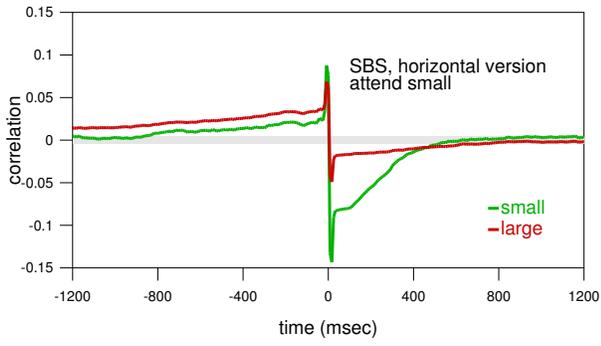
Figure 7: Saccade correlograms for subject SBS in the dual ring target experiment.

The situation is quite different with horizontal vergence, and with both axes of version. Attention alters the dynamics of the response dramatically, adding longer latency components in the 200 to 600 millisecond range. The horizontal vergence response to the ignored target is very similar to the vertical vergence responses, suggesting that the eye alignment reflex is directionally isotropic. Voluntary effort, however, adds longer latency components to the horizontal vergence only. We suggest that this is directly related to the fact that horizontal disparity change is perceived as depth and therefore provides perceptual information on which to base the voluntary effort. The long latency component is thus the temporal signature of attention-based ocular following mechanisms.

The influence of active tracking on version responses is a bit more complicated than that seen for horizontal vergence because the version responses are multiphasic. Responses to the ignored target show considerable ringing, with a peak frequency of 7-8 Hz. Horizontal and vertical version are very similar for the ignored target, and presumably reflect the isolated action of the OKR system. The addition of tracking effort adds long latency components, but also seems to reduce the ringing frequency of the response to around 5-6 Hz. It is unclear whether this is just an artifact of the addition of a broad, monophasic response or whether it reflects a change in the underlying resonant dynamics of the feedback control system driving OKR.

Figure 7 shows the saccade correlograms for the same data set. The traces corresponding to the attended target look more-or-less like the single-target traces presented above in figure 2: large negative correlations for positive time lags, with the extreme value obtained about 100 milliseconds before the saccade. In this data set, the vertical saccades again appear to be hypometric (as seen in figure 2), but, for this subject and stimulus, the horizontal saccades appear to be slightly hypermetric. The pattern of saccade correlations for the unattended target is the opposite: small or no correlations at positive lags (corresponding to position errors before the saccade), and where non-zero correlations are seen there is no temporal structure (unlike the case of the attended target), indicating that the saccades are not triggered by position errors with respect to the unattended target. For negative lags (corresponding to position errors *after* the saccade), however, the correlations with the unattended target are larger, and more persistent in time than those with the attended target, indicating that the saccades on average act to move the gaze away from the unattended target. We have also computed saccade correlograms with respect to the center of the aperture, and obtain results similar to those for the unattended targets in figure 2, providing additional evidence that the small residual correlations are not driven occasional looks to the unattended target.

The overall pattern of results allows us to reject the hypothesis that the responses to the unattended target result merely from the subject's failure to follow the instructions. If the subjects adopted a strategy of alternation between tracking the two targets, then we would expect to see features of the observed responses to the attended target (such as long-latency correlations and saccade targeting) in the responses to the unattended target.

4. HIGHER-ORDER MOTION

4.1 Non-Fourier motion and eye movements

Visual motion, as it usually occurs in natural vision, consists of translation on the retina of image patches, so that all properties move together. Occasionally, a global motion percept may be produced from local changes in image properties, such as the appearance of "amber waves of grain." In the laboratory, it is relatively easy to create artificial stimuli in which various properties are either eliminated or made to move in opposition to other properties, and a body of evidence has accumulated suggesting the existence of multiple mechanisms sensitive to various image properties. A three-tiered system has been proposed by Lu and Sperling[17, 18] to explain these results. In their hierarchy, the first-order ("Fourier") system calculates motion based on the spatio-temporal content of the luminance channel, as proposed by a number of authors.[19–21] The second-order system performs a similar computation to the first-order system, but on a higher-level property instead of the luminance signal used by the first-order system. The third-order system, on the other hand, is a feature-tracking system which operates on a high-order property related to image salience. The third-order system is distinguished from the other two by its behavior in the *pedestal test*, in which a weak moving modulation is added to a strong stationary modulation. Under normal circumstances, all three systems will be excited by a stimulus, and work together. While the third-order system may be isolated by careful selection of the stimulus, it is impossible to produce a
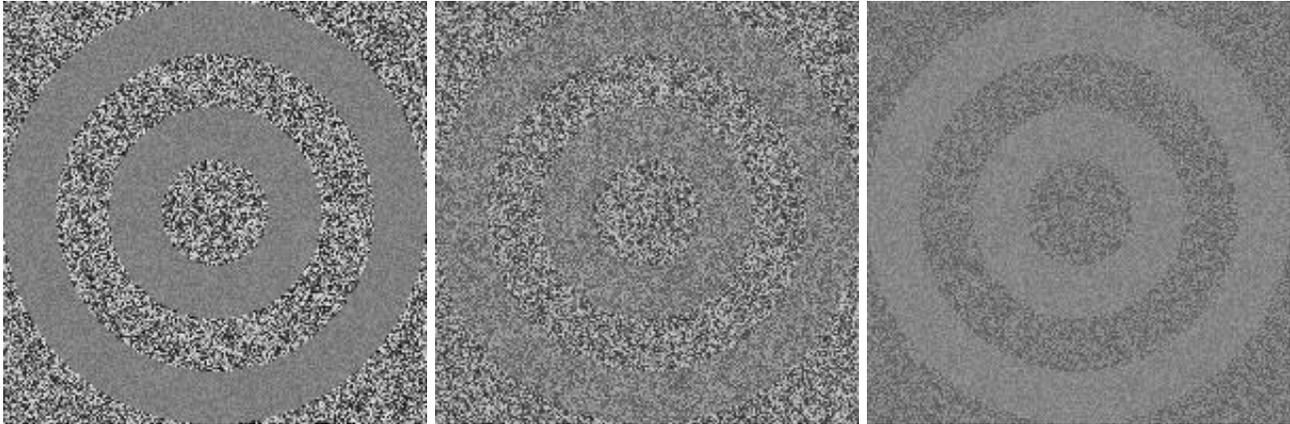
Figure 8: Representative still frames depicting a second-order pattern. Left: 70% texture contrast modulation (15%-80%). Center: 30% texture contrast modulation (35%-65%). Right: 30% texture contrast modulation, with added luminance contrast. Video at http://dx.doi.org/doi.number.goes.here.

first-order stimulus which doesn't also excite the second- and third-order stimuli; similarly most "second-order" stimuli also necessarily excite the third-order mechanism. That being said, stimuli such as a contrast-modulated carrier texture are commonly referred to as "second-order" stimuli, in spite of the fact that they also activate the third-order mechanism, and we will continue that practice here.

It seems obvious that a voluntary eye movement can be made in response to any stimulus that can be seen. It is less clear, however, which aspects of the visual input are available to the systems controlling involuntary, reflexive eye movements. Harris and Smith[22] reported that a second-order stimulus was ineffective at producing OKN, but later reported that a second-order stimulus could alter the OKN response to a simultaneously-presented first-order stimulus.[23]

The inability of a second-order stimulus to excite reflexive eye movements on its own, as suggested by Harris and Smith,[22] is in apparent conflict with a recent demonstration by Rambold, Sheliga and Miles[24] of vertical vergence eye movements elicited by second-order contrast modulation gratings. One notable aspect of the latter study is the size (extent) of the stimuli: the stimuli used by Rambold *et al.* subtended 40 degrees horizontally and 30 degrees vertically. They applied a first-order luminance modulation to null visual distortion products produced by the nominally pure second-order stimulus having constant local mean luminance, and found best cancellation with approximately 1% first-order contrast. But, for such a large stimulus, it may have been that the canceling luminance modulation was not appropriate over the entire field. The range of eccentricities for which typical peripheral defocus is less than 0.5 diopters is approximately $\pm$ 10 degrees about optical axis,[25] half the width of their stimulus. For a 5 mm pupil, defocus of 0.5 diopters corresponds to a blur circle diameter of around 0.04 millimeters on the retina.[26] Using a conversion factor of 3.5 degrees per millimeter,[27] this corresponds to a blur circle diameter of 0.14 degrees, or 8.4 arc minutes. The texture elements in the second-order stimuli used by Rambold *et al.* were square elements with a size of 6.6 arc minutes; thus, it seems likely that the effective texture contrast was reduced in the peripheral parts of the display due to variations in optical quality with eccentricity. A luminance modulation appropriate to null visual distortion products in the central field, might therefore be excessive in the periphery, contributing an unwanted first-order component. We offer this back-of-the-envelope calculation simply as a caveat that obtaining true isolation of a putative second-order mechanism over a 50 degree field is an undertaking fraught with challenges.

To address the question of whether higher-order motion is capable of driving reflexive eye movements, we applied our correlogram method, measuring the eye movements evoked by motion of a contrast-modulation pattern. The following subsections describe the experiment and results.

4.2 Methods

We examined voluntary tracking of an extended target consisting of a dynamic noise carrier pattern modulated in contrast with a square-wave bulls-eye pattern. Stimuli were projected onto a screen using a data projector

(NEC LT156) with a native resolution of 768x1024, under control of a dual processor G4 Macintosh computer. Stimuli were generated using Matlab programs in association with the Psychophysics Toolbox software suite.[28] The carrier pattern was a 12 degree circular field of dynamic random dots with mean luminance of 50 candelas per square meter, and a refresh rate of 60 Hz. Contrast of the dots was square-wave modulated between 0.25 and 0.75 on alternate rings to produce a second-order bulls-eye grating of 0.5 cycles per degree. A luminance component was added to the bulls-eye in contrasts ranging from +10% (high contrast rings brighter) through 0% (high and low contrast rings equal mean luminance) to -10% (low contrast rings brighter), in order to be sure that the equiluminant point was bracketed. Image motion was produced by driving galvanometer-controlled mirrors in the viewing path of the eye tracker.[29] The viewing path included 10 degree circular field stops which did not move, so the edges of the stimulus image were not visible to the subject. Temporal changes to the carrier pattern, as well as control of first- and second-order contrast levels were accomplished with hardware color look-up tables. A separate stimulus was presented to each eye, and the carrier texture was binocularly uncorrelated. Left- and right-eye stimuli were incorporated into a single image, which were brought into binocular correspondence using the DPI tracker's stimulus deflection mirrors during experiment set-up. Typical stimulus images are shown in figure 8; the figure caption contains a link to a short movie illustrating comparable stimuli. which depicts a single trial, rendered in two ways: the upper stereo pair presents a target with a strong first-order luminance component, while the lower pair presents a target with little or no luminance contrast between the rings. The example stimuli in the movie contain static noise, but in the actual experiment the noise was dynamic, so that the movement of the contrast modulation was the only motion.

The experiment consisted of a series of 50 4-second trials during which the left- and right-eye targets moved (via the stimulus deflection mirrors) on independent trajectories. The trajectories were precomputed by synthesizing white noise velocity signals, which were then integrated to produce a position signal. Trajectories with positional excursions greater than 2 degrees from the center were discarded. The experiment was controlled by a microcomputer running the Windows 98 operating system. The experimental software was written in Microsoft Visual Basic, and handled sequencing the trials, monitored a button box used by the subject to initiate trials, sent motion commands to the stimulus deflection mirrors via a digital-to-analog converter, and read the eye tracker gaze position signals using an analog-to-digital converter. For each trial, a data file was stored containing both the stimulus position and recorded gaze position for both the left and right eyes.

Presentation of a pure second-order stimulus requires careful calibration of display gray-levels, to ensure that there are no differences in the mean luminance between the high- and low-contrast regions (which would constitute a first-order stimulus). This was done using a photometer (Minolta LS-100) prior to the experiment. Even when calibration is done perfectly, nonlinearities in the early visual system have the potential to introduce first-order distortion products prior to the initial motion analysis. For this reason, in addition to the nominally pure second-order stimulus, we also tested with stimuli containing small amounts of deliberately added first-order contrast.

Subjects were two of the authors and one other healthy, visually normal individual from the University of Houston College of Optometry; all were experienced psychophysical observers. All subjects gave informed consent, and all procedures were reviewed and approved by the University of Houston Committee for the Protection of Human Subjects.

The stimuli were viewed binocularly and moved in a random walk pattern that stimulated both version and vergence. Subjects made their best effort to maintain fixation on the target while eye movements were recorded. Each subject completed a block of 50 trials lasting four seconds each, at each level of added luminance contrast. The presentation order of the different luminance contrast blocks was randomized.

4.3 Results

Velocity correlograms were computed for each condition; the results for the three subjects are shown in figures 9-11. Each trace in figures 9-11 is based on less than 5 minutes of eye movement data. Each figure contains panels, one for each class of eye movement. Each panel displays 9 different correlograms, corresponding to stimuli with varying small amounts of first-order luminance contrast. For each curve, a latency value was obtained by selecting the time delay corresponding to the maximum correlation. The values of the maximum correlations,
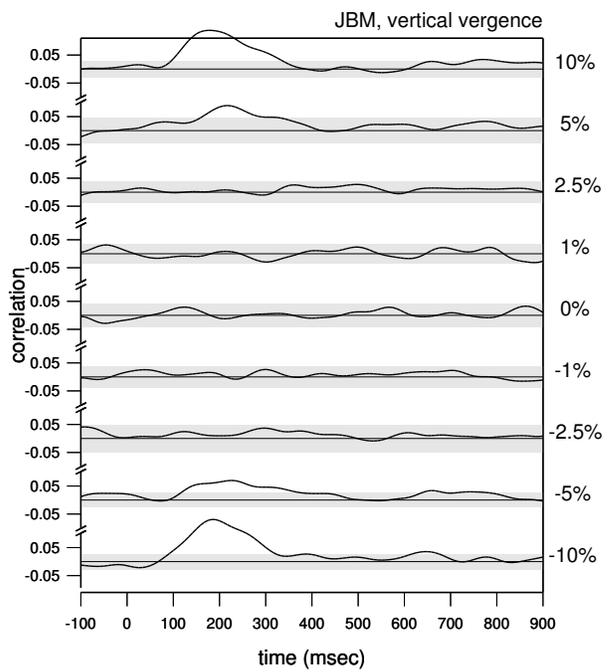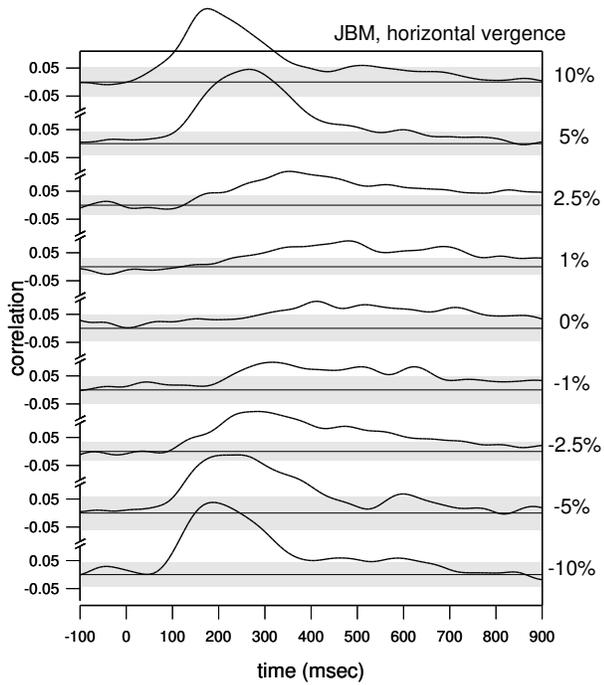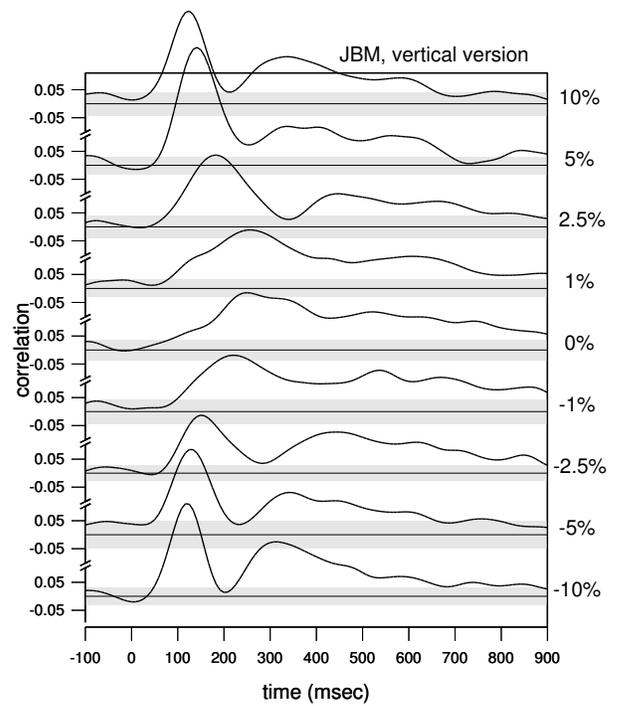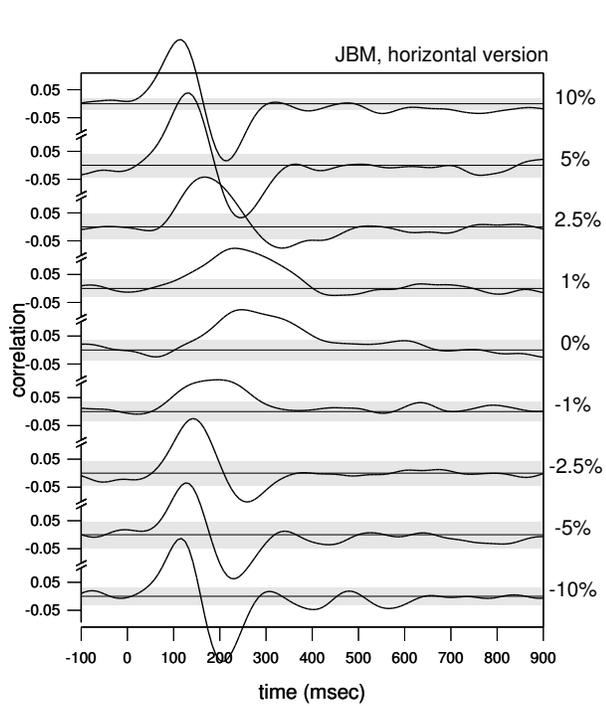
Figure 9: Eye movement correlograms for second-order target, with various amounts of added first-order contrast, subject JBM.
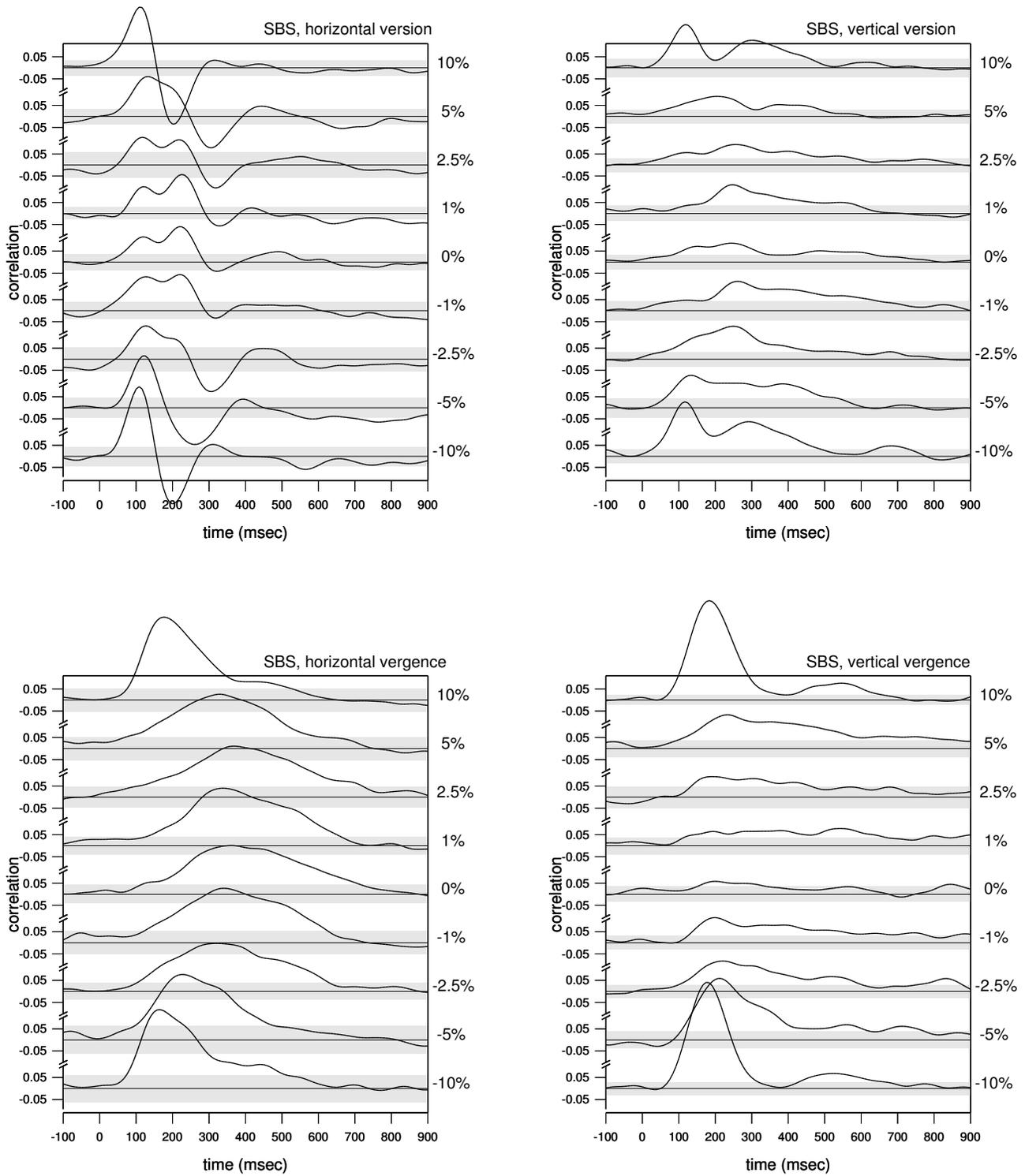
Figure 10: Eye movement correlograms for second-order target, with various amounts of added first-order contrast, subject SBS.
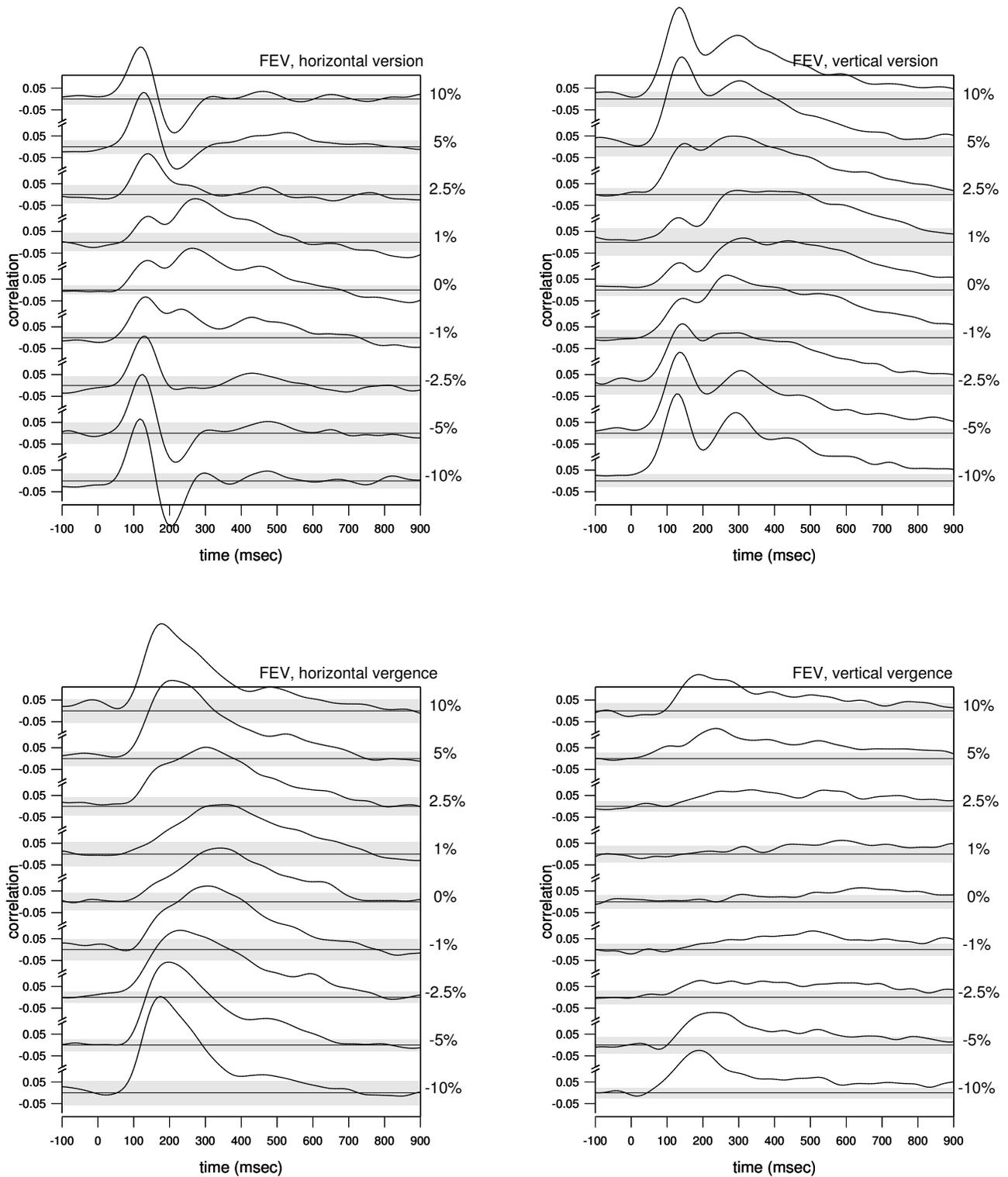
Figure 11: Eye movement correlograms for second-order target, with various amounts of added first-order contrast, subject FEV.
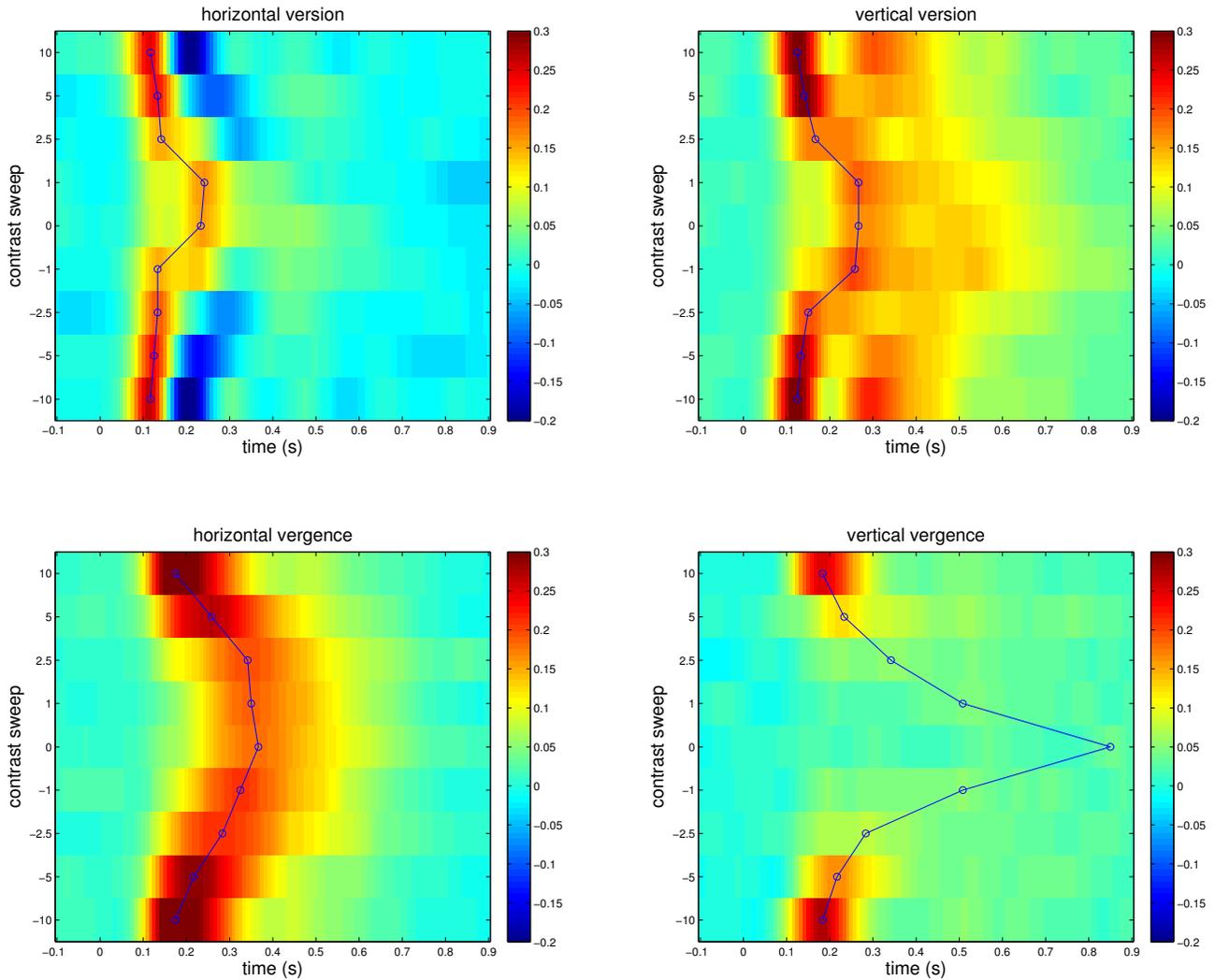
Figure 12: Data from figures 9-11 averaged together and replotted as a heat map. Overlaid symbols indicate the time of maximum correlation for each contrast value.

and the corresponding latencies, are plotted in figure 13. Figure 12 shows the data for all three subjects averaged together and replotted as a "heat map."

A number of interesting findings are evident in these correlograms. As seen previously, each type of eye motion produces a distinct latency pattern for first-order targets, which is consistent across subjects. Version responses to first-order contrast are multiphasic with the first peak latency at around 120 milliseconds, while the second-order version responses are considerably broader, and monophasic. Vergence responses are monophasic with a peak latency of around 160 milliseconds.

The data are mixed with regard to the question of whether the pure second-order stimulus is capable of driving vertical vergence. For subject JBM, the correlograms show no correlations above the noise level for a range of first-order contrasts centered at approximately 0.5%. This is not true for subjects SBS and FEV, whose data show small but statistically significant correlations over a broad range of lags. Qualitative differences between the subjects are also seen in the correlograms for horizontal version: for subject JBM, in the vicinity of the first-order null, the correlograms are significantly delayed and strictly monophasic. For the other two
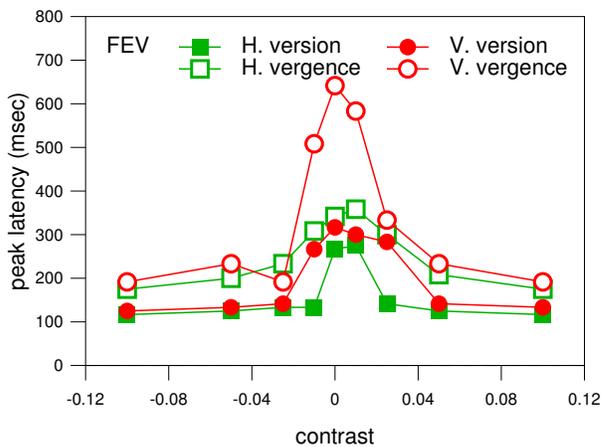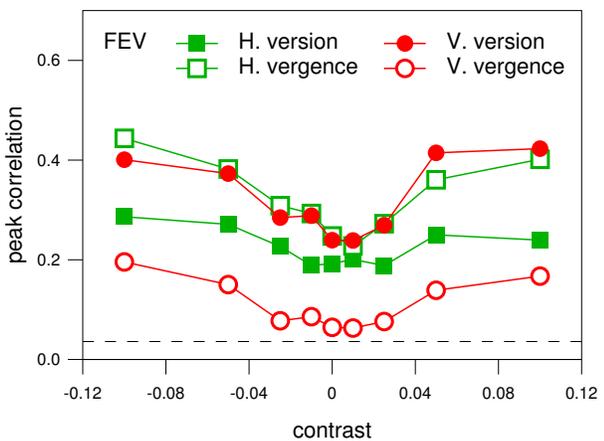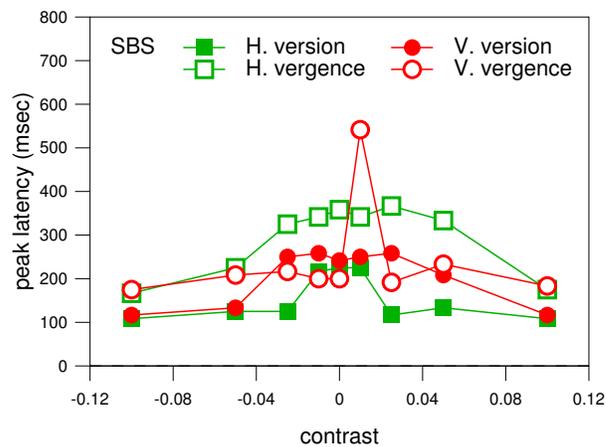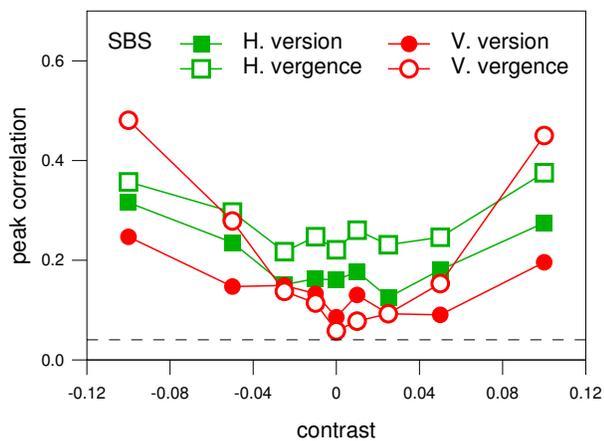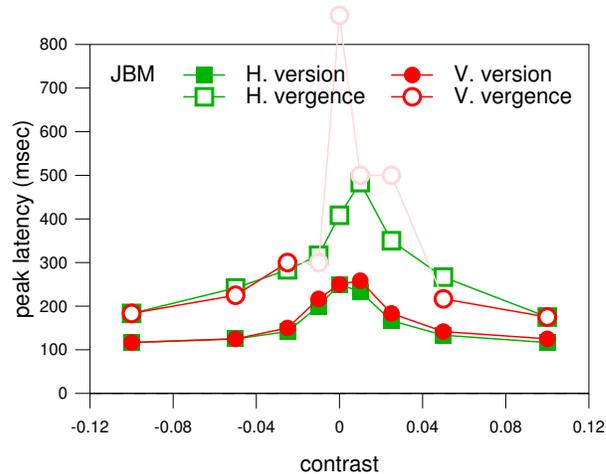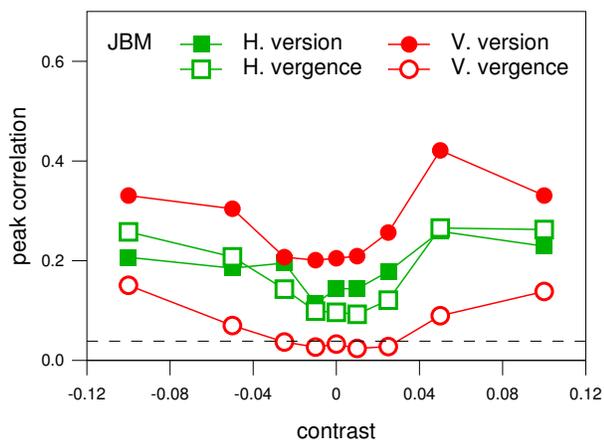
Figure 13: Summary data showing values and times of occurrence of the peak correlations of the data shown in figures 9-11. The dashed lines in the left-hand panels represent 3 times the average standard deviation of the correlogram data for lags from -1 to zero seconds. Times for peak correlations that fall below this threshold are plotted with a lighter color (upper right-hand panel).

subjects, however, although the duration of the correlogram increases, all horizontal version correlograms show an early peak at around 120 milliseconds, regardless of the level of first-order contrast. This suggests that there may have been a residual first-order component to the stimulus in spite of our best efforts to eliminate it. What could be the origin of such an artifact? One possibility is suggested by our discussion of the work of Rambold *et al.* in section 4.1, where we described how spatially-nonuniform optical blur might have introduced an artifact over part of the field. In the present case, the possibility exists of *temporally*-nonuniform blur, caused by the variable retinal motion of the stimulus. If motion blur were sufficient to introduce an effective first-order artifact, why would the data of subject JBM not show it as well? One possibility is that this subject's uncorrected optics (about half a diopter of astigmatism [†]) introduced enough blur to render the additional motion blur insignificant. This hypothesis could be addressed in future work by simply blurring the stimuli themselves before presentation to subjects.

Although our data are inconclusive on the question of second-order motion and vertical vergence, we note that the response delays we observe for the best-isolated second-order stimuli (over 100 milliseconds) are much longer than those observed by Rambold *et al.* (tens of milliseconds), which is consistent with the responses they observed being mediated by a low-contrast first-order signal, either produced by visual nonlinearities, or introduced to cancel the same.

The overall pattern of results suggests to us that the second-order stimuli activate high level, attention-driven mechanisms exclusively. The reflexive mechanisms, as typified by vertical vergence, seem insensitive to this class of stimulus. It is worth noting here that all subjects perceived vertical diplopia in the pattern, but were unable to voluntarily correct the error in vertical eye alignment. Responses to the first-order stimuli are presumably composed of both reflexive and voluntary components, in some unknown combination.

5. A CONCEPTUAL MODEL

The results described in the preceding sections have led us to a model containing independent parallel pathways, both driven by the retina, but only one of which is accessible to consciousness (see figure 14). Our model is similar to that suggested by Pola and Wyatt,[3] with additional elaborations.

The model proposes two distinct functions of attention. The first we refer to as *target strengthening*, by which we mean an early amplification of the representation of an attended target. There are numerous reports in the physiological literature of such effects, although there is some disagreement as to whether they are best characterized as modulation of response gain or contrast gain; these results have been reviewed and unified by Reynolds and Heeger.[30] This aspect of attention is crucial to our explanation of why the vertical vergence response is attenuated for an ignored target; by definition, attention has no effect on purely reflexive mechanisms.

The second role of attention is to perform target selection, a function solely of the voluntary system. Our results, showing an inability to completely ignore the motion of a distractor, are somewhat at odds with other results concerning target selection. Ferrera and Lisberger[31] reported that when two targets were present, each moving with a constant velocity, the distractor affected the latency, but not the direction, of the response to the cued target. Similarly, Kowler *et al.*[4] found that selected targets could be pursued without interference from other motions. In our model, inputs to the the reflexive system are modulated in strength by attention, but never completely quenched, as in a winner-take-all mechanism. One difference between our study and these earlier studies is that we have averaged 50 trials together, revealing correlations that involve potentially very small eye velocities. It is possible that the effects are simply too small to see in other ways. Another difference is our use of unpredictable, constantly accelerating motion for both the target and distractor. For constant motions of target and distractor, predictive components (presumably part of the voluntary system) may counteract contributions from the reflexive system; alternatively, the reflexive system may be primarily sensitive to retinal accelerations. We note that retinal acceleration has been proposed as a component of at least one model of smooth pursuit.[15]

A number of authors have previously investigated the interactions between attention and smooth pursuit. Madelain, Krauzlis and Wallman[32] reported differences in the latencies of catch-up saccades in response to perturbations depending on the spatial allocation of attention. Hutton and Tegally[33] found that divided attention

---

[†]Subjects adjust the tracker's viewing optics for optimal focus, but the possibility remains that they may not do this perfectly.
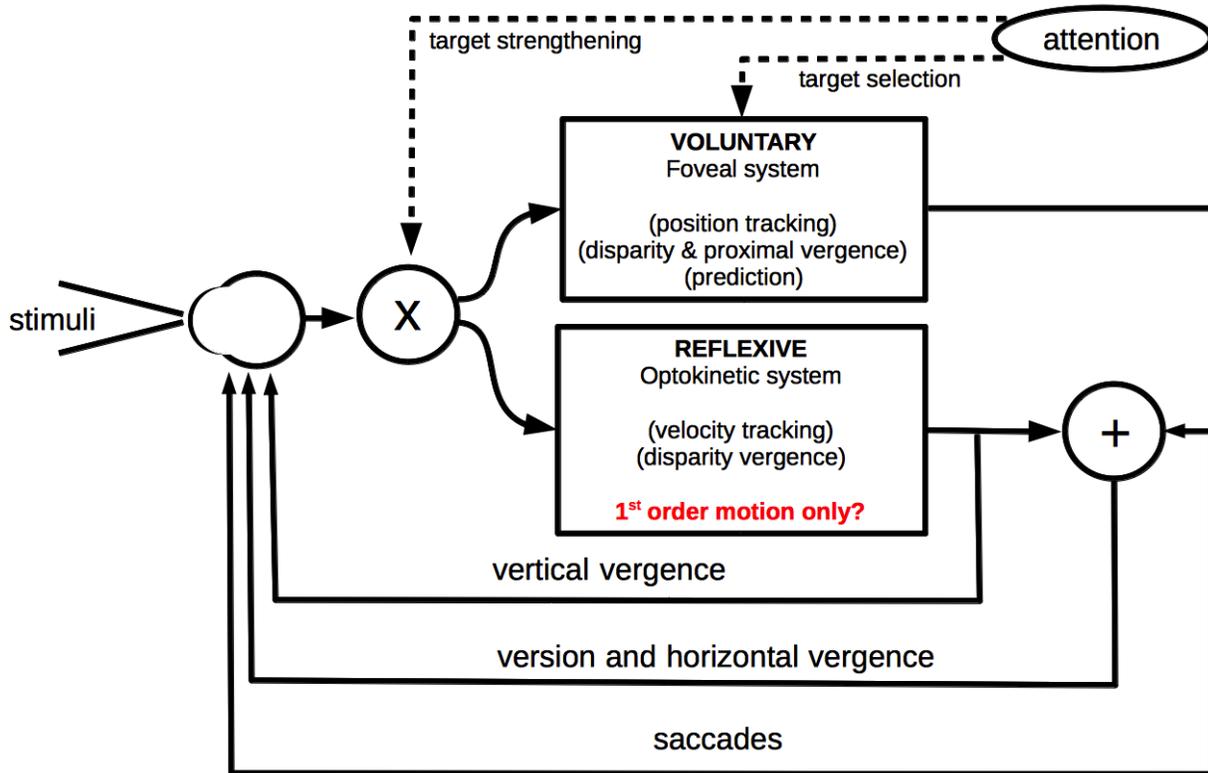
Figure 14: Proposed architecture with dual pathways for voluntary and reflexive movements.

reduced pursuit gain and increased positional errors. Similarly, Souto and Kerzel[34] found that peripheral attention to a static target reduced the gain of closed-loop pursuit, but did not affect the early open-loop phase of pursuit. We attribute the early phase to the reflexive system (relatively unaffected by attentional manipulations), with saccades and sustained pursuit requiring the action of the voluntary system. Kathmann, Hochrein and Uwer,[35] on the other hand, found that smooth pursuit error *decreased* when attention was divided between the pursuit task and an auditory discrimination, attributing this to an "automatic mode," which may correspond to our proposed reflexive system.

A key aspect of our results is that the reflexive and voluntary systems have markedly different response latencies. Montagnini, Spering and Masson[36] reported that when an oblique line segment begins to move horizontally, there is a brief movement of the eye in the "aperture" direction perpendicular to the segment, even when the horizontal motion is completely predictable, and anticipatory horizontal pursuit occurs before the motion onset. Our interpretation of this finding is that the reflexive system, unhindered by cognitive expectations, is responsible for the orthogonal response. This interpretation suggests that the reflexive system may not compute the solution of the "aperture problem."

Several investigators have examined the relationship between perceptual judgments of motion and eye movement responses, with some discrepancies in the results; Stone and Krauzlis[37] found good agreement between perception and eye movements, while Tavassoli and Ringach[38] found that the eyes responded to small motions that were undetected by perception. We would explain these differences in terms of variation in the role of the reflexive system; in cases where the eye movements are dominated by the voluntary system, we expect to see a high correlation between eye movements and perception, while in other cases where the reflexive system plays a more important role, the correlation should be diminished.

6. CONCLUSION

Our results show that the human oculomotor system tracks targets all the time, both those we attend and those we ignore. The effort to track, however, brings sophisticated, attentive mechanisms into play that supplement the reflexive mechanisms when needed. These mechanisms can respond to classes of stimuli, such as the second-order targets we used, that may be invisible to the reflexive systems. Conversely, the reflexive systems can respond to aspects of stimulus motion of which subjects are not aware, such as vertical disparity direction. (Although subjects experienced vertical diplopia some of the time, they had no way of telling which direction the error was or how they might will the eyes to correct it.) The correlogram technique provides a means by which these various mechanisms can be segregated and studied independently.

We have presented data from correlogram analysis supporting a dual process model of human eye movement control. The method has the advantage of high sensitivity (given low noise in the eye movement signal), and provides detailed information about the time course of the response. We hope this method may prove useful in future studies.

We feel the evidence is strong that fast reflexive movements are driven primarily (if not exclusively) by simple first-order luminance motion. We suspect that the pattern of results shown here for second-order motion produced by contrast-defined form may also hold for other types of high-order stimuli such as flicker-defined form, and cyclopean form (defined only by binocular disparity). A strong prediction of our model is that higher-order motion produced by an unattended target should produce no response. We hope to test this prediction in future work.

REFERENCES

[1] Carpenter, R. H. S., [*Movements of the Eyes*], Pion Limited, London (1977).

[2] Kowler, E., "The role of visual and cognitive processes in the control of eye movement," in [*Eye Movements and Their Role in Visual and Cognitive Processes*], Kowler, E., ed., 1–70, Elsevier Science Publishers B.V., Amsterdam (1990).

[3] Pola, J. and Wyatt, H. J., "Active and passive smooth eye movements: effects of stimulus size and location," *Vision Res.* **25**, 1063–1076 (1985).

[4] Kowler, E., van der Steen, J., Tamminga, E. P., and Collewijn, H., "Voluntary selection of the target for smooth eye movement in the presence of superimposed, full-field stationary and moving stimuli," *Vision Res.* **24**(2), 1789–1798 (1984).

[5] Konen, C. S., Kleiser, R., Seitz, R. J., and Bremmer, F., "An fmri study of optokinetic nystagmus and smooth-pursuit eye movements in humans," *Exp. Brain Res.* **165**(2), 203–216 (2005).

[6] Mulligan, J. B., "Pursuit latency for chromatic targets," *Inv. Ophth. Vis. Sci.* **39**, 444 (1998).

[7] Mulligan, J. B., "Image processing for improved eye tracking accuracy," *Behav. Res. Methods Instrum. Comput.* **29**, 54–65 (1997).

[8] Cornsweet, T. N. and Crane, H. D., "Accurate two-dimensional eye tracker using first and fourth purkinje images," *J. Opt. Soc. Am.* **63**(8), 921–928 (1973).

[9] Crane, H. D. and Steele, C. M., "Generation-v dual-purkinje-image eyetracker," *Appl. Opt.* **24**(4), 527 (1985).

[10] Marmarelis, P. Z. and Marmarelis, V. Z., [*Analysis of physiological systems: the white noise approach*], Plenum Press, New York (1978).

[11] Schetzen, M., [*The Volterra and Wiener theories of nonlinear systems*], Krieger Publishing Co. (2006).

[12] Deubel, H. and Bridgeman, B., "Fourth purkinje image signals reveal eye-lens deviations and retinal image distortions during saccades," *Vision Res.* **35**, 529–538 (1995).

[13] He, L., 3rd Donnelly, W. J., Stevenson, S. B., and Glasser, A., "Saccadic lens instability increases with accommodative stimulus in presbyopes," *J. Vis.* **10**(4), 14.1–14.16 (2010).

[14] Tavassoli, A. and Ringach, D. L., "Dynamics of smooth pursuit maintenance," *J. Neurophysiol.* **102**(1), 110–118 (2009).

[15] Goldreich, D., Krauzlis, R. J., and Lisberger, S. G., "Effect of changing feedback delay on spontaneous oscillations in smooth pursuit eye movements of monkeys," *J. Neurophysiol.* **67**(3), 625–38 (1992).

[16] Mulligan, J. B. and Stevenson, S. B., "Spontaneous oculomotor oscillations induced by delayed visual feedback," *J. Vis.* **7**(9), 320 (2007).

[17] Lu, Z. L. and Sperling, G., "The functional architecture of human visual motion perception," *Vision Res.* **35**(19), 2697–2722 (1995).

[18] Lu, Z. L. and Sperling, G., "Three-systems theory of human visual motion perception: review and update," *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **18**(9), 2331–2370 (2001).

[19] Watson, A. B. and Ahumada, A. J. J., "Model of human visual-motion sensing," *J. Opt. Soc. Am. A* **2**(2), 322–341 (1985).

[20] Adelson, E. H. and Bergen, J. R., "Spatiotemporal energy models for the perception of motion," *J. Opt. Soc. Am. A* **2**(2), 284–299 (1985).

[21] van Santen, J. P. and Sperling, G., "Elaborated reichardt detectors," *J. Opt. Soc Am. A* **2**(2), 300–321 (1985).

[22] Harris, L. R. and Smith, A. T., "Motion defined exclusively by second-order characteristics does not evoke optokinetic nystagmus," *Vis. Neurosci.* **9**(6), 565–570 (1992).

[23] Harris, L. R. and Smith, A. T., "Interactions between first- and second-order motion revealed by optokinetic nystagmus," *Exp. Brain Res.* **130**(1), 67–72 (2000).

[24] Rambold, H. A., Sheliga, B. M., and Miles, F. A., "Evidence from vergence eye movements that disparities defined by luminance and contrast are sensed by independent mechanisms," *J. Vis.* **10**(14:31), 1–34 (2010).

[25] Shen, J., Clark, C. A., Soni, P. S., and Thibos, L. N., "Peripheral refraction with and without contact lens correction," *Optom. Vis. Sci.* **87**(9), 642–655 (2010).

[26] Vishwanath, D. and Blaser, E., "Retinal blur and the perception of egocentric distance," *J. Vis.* **10**(10), 1–16 (2010).

[27] Hubel, D., [*Eye, Brain and Vision*], W. H. Freeman & Co. (1988).

[28] Brainard, D. H., "The psychophysics toolbox," *Spat. Vis.* **10**(4), 433–436 (1997).

[29] Crane, H. and Clark, M., "Three-dimensional visual stimulus deflector," *Appl. Opt.* **17**(5), 706–714 (1978).

[30] Reynolds, J. H. and Heeger, D. J., "The normalization model of attention," *Neuron* **61**(2), 168–185 (2009).

[31] Ferrera, V. P. and Lisberger, S. G., "The effect of a moving distractor on the initiation of smooth-pursuit eye movements," *Vis. Neurosci.* **14**(2), 323–338 (1997).

[32] Madelain, L., Krauzlis, R. J., and Wallman, J., "Spatial deployment of attention influences both saccadic and pursuit tracking," *Vision Res.* **45**(20), 2685–2703 (2005).

[33] Hutton, S. B. and Tegally, D., "The effects of dividing attention on smooth pursuit eye tracking," *Exp. Brain Res.* **163**(3), 309–313 (2005).

[34] Souto, D. and Kerzel, D., "Dynamics of attention during the initiation of smooth pursuit eye movements," *J. Vis.* **8**(14), 3.1–3.16 (2008).

[35] Kathmann, N., Hochrein, A., and Uwer, R., "Effects of dual task demands on the accuracy of smooth pursuit eye movements," *Psychophysiology* **36**(2), 158–163 (1999).

[36] Montagnini, A., Spering, M., and Masson, G. S., "Predicting 2d target velocity cannot help 2d motion integration for smooth pursuit initiation," *J. Neurophysiol.* **96**(6), 3545–3550 (2008).

[37] Stone, L. S. and Krauzlis, R. J., "Shared motion signals for human perceptual decisions and oculomotor actions," *J. Vis.* **3**(11), 725–736 (2003).

[38] Tavassoli, A. and Ringach, D. L., "When your eyes see more than you do," *Curr. Biol.* **20**(3), R93–R94 (2010).