On The Visual Input Driving Human Smooth-Pursuit Eye Movements

Leland S. Stone, Brent R. Beutter, and Jean Lorenceau
Ames Research Center, Moffett Field, California

December 1996

LELAND S. STONE, BRENT R. BEUTTER, AND JEAN LORENCEAU*

Ames Research Center

Summary

Current computational models of smooth-pursuit eye movements assume that the primary visual input is local retinal-image motion (often referred to as retinal slip). However, we show that humans can pursue object motion with considerable accuracy, even in the presence of conflicting local image motion. This finding indicates that the visual cortical area(s) controlling pursuit must be able to perform a spatio-temporal integration of local image motion into a signal related to object motion. We also provide evidence that the object-motion signal that drives pursuit is related to the signal that supports perception. We conclude that current models of pursuit should be modified to include a visual input that encodes perceived object motion and not merely retinal image motion. Finally, our findings suggest that the measurement of eye movements can be used to monitor visual perception, with particular value in applied settings as this non-intrusive approach would not require interrupting ongoing work or training.

Introduction

The control of gaze is subserved by two complementary sub-systems that play different roles: the saccadic system rapidly changes gaze position from one location of interest to another and the pursuit system generates smooth-pursuit eye movements that maintain stable foveation of a moving object (ref. 7). Rashbass (ref. 40) demonstrated that pursuit is largely a response to motion and not position. Furthermore, by a direct mechanical link, eye motion affects the image motion on the retina. Thus, pursuit is a negative-feedback system. Young and colleagues (ref. 53) later pointed out that greater stability and accuracy could be achieved if eye-velocity positive feedback were added to the negative-feedback configuration to sustain steady-state pursuit in the absence of visual inputs. Considerable supporting neurophysiological and behavioral evidence for positive feedback through the cerebellum has subsequently been found (ref. 15, 24, 30-32, 48). Current models of pursuit (e.g. ref. 21,42) generally incorporate these two mechanisms (negative feedback and an internal positive-feedback loop) and implicitly or explicitly assume that the input is local retinal motion which the system attempts to drive to zero. In this study, we challenge this notion by using a new stimulus, a moving occluded diamond (ref. 25, 46) to elucidate the nature of the visual signal that drives pursuit.

This work was primarily supported by NASA RTOP 199-16-12-37 to Dr. Stone. Dr. Lorenceau was on sabbatical at Ames Research Center from the Centre National de Recherche Scientifique and received additional support from a NATO travel grant and Stanford-NASA Center of Excellence Grant NCC 2-307. Dr. Beutter is supported by a National Research Council postdoctoral associateship. We thank Dr. B. Chapman for helpful suggestions, Mr. P. Stassart for programming support, and Drs. R. Krauzlis and D. Tomko for comments on an earlier draft.

Methods

Observers were asked to track as best they could the center of a diamond which moved either clockwise (CW) or counter-clockwise (CCW) along an elliptical trajectory (ellipse of height of 1.6° and width of 1.4°) behind an invisible cross aperture which concealed at all times the four corners of the diamond (Fig. 1). Thus, under all conditions, only four line segments (two yoked pairs with the same orientation) were displayed. Each pair of segments oscillated sinusoidally in either sine or cosine phase along a straight line perpendicular to their common orientation (Fig. 1A). Observers performed this task in both a low- and a high-contrast condition: the former in which the luminance of the segments was low (44 cd/m²) and the latter in which it was high (108 cd/m²) with the surrounding luminance at 38 cd/m². Perceptually, our observers confirmed what has previously been reported: the low-contrast condition leads largely to the perception of a diamond moving coherently behind an

* Laboratoire de Physiologie de la Perception et de L'Action, College de France

1 Because the pixels were not exactly square, the motion was ~5° off from perpendicular.
invisible virtual aperture and the high-contrast condition to a jumbled mess of segment motion (ref. 25, 46). The advantage of these two stimuli is that they contain identical local motion (ref. 1, 23, 35, 44), yet are interpreted dramatically differently. References 25 and 46 describe the stimulus and the perceptual effects of luminance contrast in detail and discuss possible explanations. The present study does not attempt to explain why perception is altered by contrast. We merely use the effect as a tool to probe the relationship between perception and pursuit.

Using a video-based eye tracking system (ISCAN RK-426) calibrated to yield a resolution of ~0.2°, we measured eye movements in four subjects (one naive) in response to a 3.2 s presentation of the moving-diamond stimulus, following a 500-ms presentation of a fixation cross. On a given trial, one of four possible trajectories was presented in a randomly interleaved order: two CW and two CCW with two starting locations each (for a total of four absolute horizontal-versus-vertical phase configurations of the moving diamond). The saccade-free portions of the horizontal and vertical eye-position data from the last 2.4s of each trial were fit separately using a least-squares procedure to find the amplitude and phase of the best-fitting sinusoid at the stimulus temporal frequency (0.9 Hz).

We used a digital-filter implementation of an acceleration threshold to identify saccades. For the data presented here, we set the threshold to ~0.9° as this value was most consistent with our subjective identification of saccades, although some small saccades might have been missed. We therefore reanalyzed the data using a threshold of 0.3°. Although this worst-case analysis clearly identified false saccades, nonetheless 86% of trials still had more than 0.5 cycles of data identified as saccade-free. Pursuit gain was diminished only by ~5% and phase was shifted by only a few degrees. This demonstrates that the response is largely pursuit and that saccadic contamination in our quantitative analysis is small.

**Results**

We found that all four subjects were able to track reliably the high-coherence diamond, but not the low-coherence diamond. This is illustrated by the raw data shown in figure 1. Panels B and D show single examples of trials of a naive observer and of an author, respectively, and illustrate the elliptical trajectories elicited in the high-

---

2 The direction/orientation and speed tuning of striate cortical neurons appears largely independent of contrast. It is not clear to what extent this is true for MT neurons.

---

We therefore reanalyzed the data using a threshold of −0.9° as this value was most consistent with our subjective identification of saccades, although some small saccades might have been missed. This suggests that either the size of the virtual trajectory was underestimated by a less than perfect spatio-temporal integration process, or that the virtual motion was occasionally incoherent thereby reducing the mean gain, or both.

Despite the less than perfect gain, pursuit phase was both largely accurate and highly precise. Figures 2A-D shows plots of the vertical versus horizontal phase of the pursuit response for all trials (open circles) for all four observers, as well as that expected from perfect tracking of the diamond (solid circles) or of a segment (solid squares). The solid circles at horizontal and vertical phase combinations of (45°, 135°), (225°, 315°), (135°, 45°), and (315°, 225°) are shown in four different colors representing the four possible trajectories of the diamond. The first two phase pairs correspond to CCW and the latter two to CW motion. The two solid squares adjacent...
Figure 1. The occluded-diamond stimulus and raw oculomotor responses. A. Two frames of the stimulus which consisted of four moving oblique 3.2° line segments which mimic the motion of a 11.5° by 9.7° diamond behind an invisible cross-shaped aperture. The segments moved sinusoidally in quadrature-phase along linear paths quasi-perpendicular to their orientations. When the contrast of the segments is low, the stimulus is seen as a diamond moving elliptically at -8°/s. When the contrast of the segments is high, the motion is not integrated and only the separate motion of the segment pairs is seen. B & C. Raw eye-position traces from 64 frames (one cycle) of a single trial of naive observer DG in response in the high- and low-coherence conditions, respectively. Note that in B pursuit is elliptical (as is the motion of the diamond) while in C it is linear and oblique (as is the motion of segments). D & E. Same as B & C for highly practiced observer JL.
Figure 2. Temporal correlation between the virtual elliptical trajectory and pursuit for all 4 observers. Note that the graphs are inherently periodic and wrap-around along both axes. In all panels, the open circles represent the best-fitting horizontal and vertical phase of the smooth portion of the eye movement. The solid circles represent perfect tracking of the diamond with the four colors (black, red, green, and blue) representing the four possible phase configurations. For each configuration, there are two segment phase configurations, one corresponding to each segment pair, shown as two purple squares adjoining each solid circle. There were 60 trials for each observer except BB for whom there were 120. Every trial for each observer is shown in A-D for the high-coherence and in E-F for the low-coherence condition.
to each solid circle represent the diagonal linear motion of the two segment pairs for that condition (either exactly in phase, 0°, or exactly out-of-phase, 180°). Note that for all four observers, the data from all trials nearly superimpose on the appropriate solid circle, although there is a slight deviation downward and leftward as expected from a slight phase lag. The mean absolute phase lag with respect to perfect elliptical tracking for all four subjects was \(-11° \pm 16°\) (±RMS error, defined as the square root of the mean of the squared differences between eye and diamond phase) and the mean differential phase (horizontal minus vertical phase) was \(+85° \pm 15°\) for CW trials and \(-88° \pm 13°\) for CCW trials. These latter values are close to the +90° and -90° for perfect elliptical tracking and to the +85° ± 9° and -93° ± 9° observed in the two subjects tested with the fully visible diamond, and far from the 0° or 180° expected from tracking a single segment. These data show that, in the high-coherence condition, observers were pursuing largely elliptical trajectories.

The low-coherence condition yielded a dramatically different pattern of results. Figure 2E-H shows the phase data for all four subjects. Pursuit in the low-coherence condition was much more variable: observers were not able to track the diamond reliably (if at all), despite the fact the stimulus motion was identical to that in the high-coherence condition. While observer JL (the most highly practiced) appeared to track a particular segment consistently (Fig. 1E and 2E), the remaining observers occasionally tracked a single segment (Fig. 1C) but generally did not (Fig. 2FGH). The mean RMS phase error with respect to ellipse tracking is 36°, which is more than twice the 16° found in the high-coherence condition, and close to the 45° expected for perfect segment tracking. These results indicate that, at the very least, pursuit of the diamond in the low-coherence condition was greatly impaired. Subjects may even have been unable to pursue the diamond at all. The few phase responses that appear close to that of the diamond may be due either to the average tracking of two different segments within a single trial or to the occasional coherence even at high contrast.

The above data demonstrate that elliptical pursuit is on average correlated with perceptual coherence. In addition, the fact that observers see the coherent percept even when instructed to suppress pursuit (ref. 25) allows us to infer a causal direction for this correlation. The most parsimonious explanation of our results is that a signal related to perceived object motion drives pursuit. The fact that a mere contrast manipulation can dramatically change the shape of the gaze trajectory rules out simple tracking of a linearly low-pass filtered (blurred) image. However, three reasonable alternate explanations deserve further exploration. First, it could be argued that, although the accurate tracking of the high-coherence diamond does not drive retinal motion to zero, at low contrast it might drive local-motion neuronal signals to some minimum at the level of striate cortex (V1) or the Middle Temporal cortical area (MT). Thus, local retinal motion might drive pursuit with perceptual coherence being merely an epiphenomenon. Indeed, although there is residual retinal motion during perfect steady-state elliptical pursuit, it is entirely parallel to the orientation of the segments. At low contrast, such motion is not a potent stimulus for V1 or most MT neurons (ref. 2, 45) and may be near or below threshold4. At high contrast, the residual motion may vigorously drive many V1 and MT neurons and may thereby disrupt elliptical pursuit. Second, a related hypothesis is that pursuit may be driven by a low spatial-frequency channel that processes the segments together and is not related to object motion per se. High spatial frequencies in the stimulus would then disrupt pursuit only at high contrast, because they would then be above threshold. To test these two hypotheses, we used a visible as opposed to virtual stationary cross-shaped aperture and ran a control experiment with high-contrast segments. The aperture was made dark (0.2 cd/m²) and therefore visible, rather than equiluminant with the surround as before. This condition produced a coherently moving diamond as has been reported previously (ref. 25). Despite the high contrast of the segments and a visible foreground which would be expected to impair tracking (ref. 9, 17, 18), pursuit remained robust and elliptical. The mean horizontal and vertical gains were 0.73 and 0.52, respectively. The mean phase lag was \(-5° \pm 16°\) and the mean relative phase was \(+95° \pm 19°\) for CW and \(-85° \pm 20°\) for CCW motion. These results show that the presence of visible high spatial-frequency and/or high contrast components per se cannot explain the impaired pursuit in the low-coherence condition. Third, an alternate interpretation is that our observers were not tracking the diamond's trajectory, but merely tracking the centroid of the segments. We ruled out this hypothesis by performing a second control experiment using two parallel vertical apertures rather than the cross-shaped aperture shown in figure 1. In this configuration, the four line segments oscillate up and down within visible vertical apertures. The eye movements in the high-coherence condition were qualitatively unchanged. Although the segments' centroid moved purely vertically, the three observers tested reliably generated elliptical tracking (mean relative phase:

4 Ref. 2 shows that ~29% of MT neurons prefer motion parallel to their preferred orientation and ref. 45 that many MT neurons respond well at the "low" contrast (16%) we used. These two facts provide the caveat that some MT neurons may respond well to residual segment motion even during perfect steady-state pursuit in the high-coherence condition.
+92° ± 15° CW and -75° ± 23° CCW) with a vigorous horizontal component (mean horizontal gain: 0.82). This finding rules out the possibility that the elliptical tracking in the high-coherence condition is merely the result of tracking the centroid of the segments.

Discussion

The issue of perceptual versus retinal motion driving pursuit has been addressed previously, but has remained unresolved largely for want of an appropriate stimulus. Yasui and Young (ref. 52) showed that a stabilized foveal image could enhance eye movement responses during vestibular stimulation as well as generate perceived target motion without any retinal motion. Although their interpretation is that perceived target motion generates pursuit which then augments the vestibulo-ocular reflex (VOR), they themselves point out that their data are inconclusive as the enhanced eye movements could merely have resulted from an attentional modification of VOR gain. Steinbach (ref. 47) also presented qualitative data that humans could pursue the perceived motion of a wagon wheel defined by four illuminated points along its circumference and the horizontal motion of an object presented behind a narrow vertical slit. Although his data show that perceived motion influences pursuit, they did not evaluate whether observers were accurately tracking the object motion. Finally, Pola and Wyatt (ref. 39) also presented evidence that pursuit can be influenced by perceptual manipulations during pursuit of stabilized motion (open-loop). However, they also do not show whether their observers were actually tracking the perceived motion and it is unclear to what extent the open-loop nature of their stimulus may have altered normal pursuit strategies.

On the other hand, Mack and colleagues (ref. 27, 28) disputed the view that perceptual motion drives pursuit by showing that humans pursue retinal motion even in the presence of illusory (induced) motion. Their finding is however inconclusive for two reasons. First, accurate performance in their psychophysical task did not necessarily reflect a perception of induced motion, as it could have been performed quite accurately using static displacement information. Second, their stimulus made perceived motion different than retinal "target" motion by adding additional "non-target" motion, so the perceptual-retinal dichotomy is confounded with target selection and potential motion-motion interactions. Because we used contrast (or the presence of a static aperture) to manipulate perceived motion, our stimulus does not suffer from this problem.

Our experiments provide the first quantitative analysis indicating that humans can pursue a virtual trajectory defined only from the integration of motion signals across space and time. Our results complement the recent report showing that smooth vergence eye movements can be generated to track changes in perceived depth defined only by the spatio-temporal integration of motion cues (the kinetic depth effect) (ref. 41). In particular, we found that when retinal and perceived object motion were both present but different (as was generally the case in our high-coherence condition), pursuit was precisely linked to object motion, rather than to raw retinal motion. Furthermore, this tight link was lost in our low-coherence condition when diamond motion was not generally perceived. The front-end of present models of pursuit should therefore be extended to reflect these facts (Fig. 3).

A second implication of our results is that, although the spatio-temporal integration of local motion signals into a signal related to object motion may begin in area MT (ref. 36, 43, 49), visual motion alone cannot always support accurate integration during ongoing pursuit. This is a problem that remains unaddressed by current models of human motion perception. During perfect steady-state pursuit, the retinal motion alone is consistent with a stationary diamond viewed through an aperture moving along an elliptical path, yet the veridical percept is that of a moving diamond viewed through a stationary aperture. Specifically, during perfect steady-state pursuit, the residual retinal segment motion is all parallel to the segment orientations and is inadequate to recover the rigid diamond. The rigid linking of the segments however can be achieved if the system is aware of the ongoing eye movement or retains some visual "memory" of ongoing object motion (perhaps by positive feedback). In the case of our virtual trajectory, such an extra-retinal signal could play two roles. In addition to providing the stabilized target with an ongoing velocity (ref. 52, 53), it could also enable the proper linking of the segments which, during perfect tracking, is not possible from the residual segment motion alone.

The sustained perception of a moving diamond experienced during pursuit of the coherent diamond therefore suggests that a higher cortical area which contains extra-retinal signals related to pursuit, perhaps the Medial Superior Temporal area (MST) (ref. 11, 19, 36, 43, 49), visual motion alone cannot always support accurate integration during ongoing pursuit. This is a problem that remains unaddressed by current models of human motion perception. During perfect steady-state pursuit, the retinal motion alone is consistent with a stationary diamond viewed through an aperture moving along an elliptical path, yet the veridical percept is that of a moving diamond viewed through a stationary aperture. Specifically, during perfect steady-state pursuit, the residual retinal segment motion is all parallel to the segment orientations and is inadequate to recover the rigid diamond. The rigid linking of the segments however can be achieved if the system is aware of the ongoing eye movement or retains some visual "memory" of ongoing object motion (perhaps by positive feedback). In the case of our virtual trajectory, such an extra-retinal signal could play two roles. In addition to providing the stabilized target with an ongoing velocity (ref. 52, 53), it could also enable the proper linking of the segments which, during perfect tracking, is not possible from the residual segment motion alone.

5 It should however be noted that the retinal slip when tracking the virtual center of the occluded diamond is not foveal. It is possible that the presence of foveal retinal motion would have overridden the pursuit signal from perceived object motion.

6 Alternate hypotheses include the possibility that segment linking during steady-state pursuit is accomplished via higher-order visual constraints such as the cognitive concept of a moving invisible aperture, or that residual perpendicular motion, caused by less than unity gain, is used to link the segments.
Figure 3. Block diagram of the primate pursuit system consistent with our results within the constraints of primate extrastriate cortex and brainstem anatomy and physiology. The main point is that there are two positive feedback loops: one cerebral and one cerebellar. The former is involved in the integration of visual motion into an object-motion signal that we postulate drives both pursuit and perception and the latter in the processing of motor commands to compensate for the dynamics of the oculomotor plant (see ref. 22). Given that the output of the cerebellum is inhibitory yet stimulation of the ipsilateral cerebellum in regions where Purkinje cells are excited during ipsiversive pursuit produces ipsiversive eye motion, there must be a second sign change before reaching the plant. If this final sign change is functionally outside of the feedback loop and the gain of the straight-through cerebellar pathway is \( G \), then the overall gain of the efferent limb up to the plant is \( G/(1 + GP') \), which reverts to \(-1/P'\) at reasonably high values of \( G \). Therefore, if \( P' \) (cerebellar feedback transfer function) \(- P \) (plant transfer function), then the overall gain of the efferent limb through the plant becomes \(-1\), which is exactly what is desired if the input driving the efferent limb is object motion. Such a scheme for cerebellar processing is generalizable to motor control in general. Parallel pathways, including direct subcortical and indirect cortical pathways through the accessory optic system (see Fig. 15 of ref. 48), or pathways through the superior colliculus and pulvinar to MT may also contribute, but have been omitted for simplicity. LGN, lateral geniculate nucleus; V1, striate visual cortex; MT, middle temporal area; MST, medial superior temporal area; FEF, frontal eye fields; \( O' \), positive feedback within MST or between MST, FEF, or other areas; \( P' \), positive feedback through the cerebellum; VN, vestibular nucleus.
The framework for pursuit in figure 3 reflects this fact and integrates the segment motion into the object-motion signal used to drive both perception and pursuit. Our high-coherence data (Fig. 2A-D) demonstrate that a signal related to object motion drives pursuit. The effect of MST lesions and stimulation argues that this pursuit signal at least passes through MST (ref. 12, 13, 20, 51). Furthermore, both perceptual coherence (ref. 25, 46) and elliptical pursuit (Fig. 2E-H) are impaired at high-contrast. This strongly argues that perception and pursuit share the same object-motion signal. The fact that stimulation and lesions of MST affect motion perception (ref. 8, 38) as well as pursuit (ref. 12, 13, 20, 51) provides supporting neurophysiological evidence for this view. Moreover, the motion of the diamond (and objects in general) during steady-state pursuit cannot be accurately estimated by the simple linear addition of retinal and eye motion as observed within cerebellar oculomotor structures (ref. 15, 24, 30, 31) and posited by most current models of pursuit (ref. 21, 42, 52, 53). In general, the accurate estimation of target motion during ongoing pursuit will require complex non-linear spatio-temporal integration of visual and oculomotor signals. A visual cerebral cortical site is therefore more likely responsible for the sustained eye velocity as well the perceived object motion experienced during steady-state pursuit.

The consequences of the proposed change to the front end of models of pursuit are profound (Fig. 3). If the integration of visual and eye motion is performed in extrastriate visual cortex with the observed extra-retinal signal in MST (ref. 37) a manifestation of that fact, then this process need not be performed at the level of the cerebellum as has been previously thought (ref. 24, 31, 48). If the visual cortex provides a signal related to target-object trajectory to drive pursuit, then the documented cerebellar positive feedback loop may be performing another task, perhaps compensation for the dynamics of the oculomotor plant (ref. 22) as this is the only remaining transformation necessary once target motion is derived. The framework for pursuit in figure 3 reflects this fact and also provides new insight into the nature of the non-retinotopic deficits that occur following MST and FEF lesions in monkeys (ref. 12, 13, 16, 26, 51) and occipitoparietal and frontal lesions in humans (e.g. ref. 33, 34, 50).

Conclusions

We have found that human smooth-pursuit eye movements are not merely driven by local retinal motion (retinal slip), but rather by a signal related to perceived target-object motion. Models of human pursuit must therefore be modified to include a front end which is capable of performing the spatio-temporal integration necessary to estimate object motion. Our results together with other recent findings (ref. 3, 4) have considerable implications for the potential application of eye-movement measurements to aerospace human factors. They suggest that the monitoring of eye movements can be used as an indirect, non-invasive, and non-intrusive measure of human perception. Such a method could therefore be used in applied settings to gather critical human-factors data without interfering with the task at hand. For example, oculomotor monitoring could be used to evaluate the quality of displays and virtual environments by measuring their ability to generate accurate motion percepts, or to quantify human perceptual performance in simulated or real visuo-motor tasks as part of an enhanced training paradigm.

References

9. Collewijn, H.; and Tamminga, E.P.: Human smooth and saccadic eye movements during voluntary pursuit of different target motions on different


On The Visual Input Driving Human Smooth-pursuit Eye Movements

Leland S. Stone, Brent R. Beutter, and Jean Lorenceau

Ames Research Center
Moffett Field, CA 94035-1000

Current computational models of smooth-pursuit eye movements assume that the primary visual input is local retinal-image motion (often referred to as retinal slip). However, we show that humans can pursue object motion with considerable accuracy, even in the presence of conflicting local image motion. This finding indicates that the visual cortical area(s) controlling pursuit must be able to perform a spatio-temporal integration of local image motion into a signal related to object motion. We also provide evidence that the object-motion signal that drives pursuit is related to the signal that supports perception. We conclude that current models of pursuit should be modified to include a visual input that encodes perceived object motion and not merely retinal image motion. Finally, our findings suggest that the measurement of eye movements can be used to monitor visual perception, with particular value in applied settings as this non-intrusive approach would not require interrupting ongoing work or training.