Short-Latency Visual Stabilization
Mechanisms that Help to Compensate for
Translational Disturbances of Gaze

F. A. MILES

Laboratory of Sensorimotor Research, National Eye Institute, National
Institutes of Health, Building 49, Room 2A50, 49 Convert Drive,
Bethesda, Maryland 20892, USA

ABSTRACT: Recent studies in primates have revealed short-latency visual tracking
mechanisms that help to stabilize the eyes during translational disturbances of the
observer, and so operate as backups to otolith-mediated vestibulo-ocular reflexes.
One such mechanism generates version eye movements to help stabilize gaze when
the moving observer looks off to one side, utilizing binocular disparity to help single
out the images in the plane of fixation (ocular following). Two others generate ver-
gence eye movements to help maintain binocular alignment on objects that lie ahead:
one responds to the radial patterns of optic flow (radial-flow vergence) and the other
to the changes in binocular parallax (disparity vergence). Accumulating evidence
suggests that, despite their short latency, all are mediated by the medial superior tem-
poral area of cortex.

INTRODUCTION

Most studies concerned with the visual stabilization of gaze have used stimuli in the
form of an optokinetic drum or a planetarium-like rotating array of spots, with the object
of simulating the optic flow associated with shortcomings of the canal–ocular reflexes
during head rotations. The presumption that rotational disturbances were the adequate stim-
ulus for the visual reflexes as well as for the canal–ocular reflex was nicely reinforced by
the finding in rabbits that the optokinetic system was organized in canal coordinates,¹ an
arrangement that was assumed to foster the veridical summation of the visual and vestibular
inputs. It is possible that the geometry is in fact dictated by the pulling directions of the
eye muscles rather than the orientations of the canals² but, regardless, there seems to be no
question that the optokinetic system of the lateral-eyed rabbit is organized in rotational
coordinates. Curiously, such a shared organization for canal–ocular and optokinetic
reflexes has yet to be demonstrated in primates, an oversight worthy of a doctoral thesis.
In recent years, however, it has become apparent that primates like ourselves with frontal
eyes have (in addition?) visual and vestibular reflexes for which translational disturbances
are the adequate stimulus. While my major concern here is with these newly discovered
visual mechanisms in primates, it is useful to first touch on some aspects of vestibulo-
ocular function in primates during translational disturbances.

¹Address for telecommunication: Phone: 301/496-2455; fax: 301/402-0511; e-mail:
fam@lsr.nei.nih.gov
THE TRANSLATIONAL VESTIBULO-OCULAR REFLEXES

A Visual or Gaze-Centric View

The canal–ocular and otolith–ocular reflexes are often referred to as the angular and linear vestibulo-ocular reflexes (VORs) in accordance with the adequate stimulus for activation of the endorgans, angular and linear accelerations, respectively. There are two linear VORs, however, one compensating for translations and the other for changes in the orientation of the head with respect to gravity, and I shall refer to them as the translational vestibulo-ocular reflex (TVOR) and the orientational vestibulo-ocular reflex (OVOR), respectively. In keeping with these functional descriptors, I shall refer to the canal-ocular reflex as the rotational vestibulo-ocular reflex (RVOR).

The available evidence suggests that the TVOR operates in all directions,3,5 stimuli and responses typically being defined in head-centric coordinates with naso–occipital, interaural, and dorsoventral components. Of course, this is simply a convenient descriptive convention, and the reference frame used by the nervous system is unknown. However, the visual backup mechanisms appear to be organized in a gaze-centric reference frame, and it simplifies matters to assume that this organization is shared by the otolith reflexes dealing with translational disturbances. One visual backup, ocular following, generates conjugate (version) eye movements to compensate for translational disturbances that are orthogonal to the direction of gaze (as when the moving observer looks off to one side) and, I suggest, works in concert with a labyrinthine mechanism that I shall term the gaze-orthogonal-TVOR (GO-TVOR). A second visual backup, radial-flow vergence, generates disjunctive (vergence) eye movements to compensate for translational disturbances that are isogonal to the direction of gaze (as when the moving observer looks in the direction of heading) and, I suggest, works in concert with a labyrinthine mechanism that I shall term the gaze-isogonal-TVOR (GI-TVOR). Note that version (Vs), the average position of the two eyes [(L + R)/2], and vergence (Vg), the difference in the positions of the two eyes [L − R], provide a complete description of binocular eye movements, the position of the left eye being given by [Vs + (Vg/2)] and of the right eye by [Vs - (Vg/2)]. While drawing a clear distinction between version and vergence mechanisms, I do not wish to imply that they are totally independent subsystems.6

Dependence on Viewing Distance

In the case of the RVOR, if we ignore the separation between the axes of eye rotation and the axis of head rotation, perfect compensation would require that any head rotation (the input) be exactly matched by an equal and opposite eye rotation (the output), in which event the gain (given by the ratio, output/input) would be simply unity for all head rotations. The desired eye rotations here are always conjugate, hence the expected output is always version. In contrast, for the TVOR to be optimally effective, its gain should accord with the proximity of the object of interest, nearby objects necessitating much greater compensatory eye movements than distant ones in order for their retinal images to be stabilized during translation. If the observer looks off to one side during translation so that his or her gaze is orthogonal to the direction of heading (pure activation of the GO-TVOR), then the required compensatory eye movements are conjugate (version) and should have a gain that is inversely proportional to the viewing distance. There is ample evidence that this is, indeed, the case.3,5,11-12 On the other hand, when the observer looks ahead during translation

Annals of the New York Academy of Sciences, Volume 871
so that his/her (cyclopean) gaze is in the direction of heading (pure activation of the GI-TVOR), then the required compensatory eye movements are disconjugate (vergence) and should have a gain that is inversely proportional to the square of the viewing distance. Although it is intuitively obvious that the instantaneous vergence angle required for binocular alignment is inversely proportional to the viewing distance, it is not so readily apparent that, as the observer moves toward or away from the object of regard, this vergence angle must change at a rate that is inversely proportional to the square of the viewing distance: if we assume an observer with interpupillary separation, \( S \), fixating an object at a distance, \( D \), then as the observer moves a distance, \( \Delta D \), toward that object, the required rate of increase in the vergence angle with respect to time is approximately by \( \frac{(S/(D-\Delta D))}{s/D/s} \Delta D/\Delta t \), which simplifies in the limit to \( s/D^2 \Delta D/\Delta t \). Perusal of the available data suggests that this might be the case, although no formal attempt has been made to fit such a function to the data. I shall return to this issue of the dependence on viewing distance later.

OPTIC FLOW PATTERNS ASSOCIATED WITH ROTATIONS AND TRANSLATIONS

The vestibular system's decomposition of head movements into rotational and translational components results directly from the contrasting physical properties of the endorgans, the semicircular canals and the otolith organs, respectively. However, there is no such decomposition of the optic flow by the visual endorgans: the two retinas see the full consequences of both translational and rotational disturbances of the observer. Any visual decomposition must be done by signal processing within the CNS, presumably utilizing the different patterns of optic flow associated with rotational and translational disturbances of the observer. An observer who undergoes pure rotation (without compensating experiences) en masse motion of his or her entire visual world, the direction and the speed of the optic flow at all points being dictated solely by the observer's motion. The overall pattern of optic flow resembles the lines of latitude on a globe (Fig. 1A) but, of course, the

FIGURE 1. Patterns of optic flow experienced by a (passive) moving observer. (A) The retinal optic flow can be considered to be distributed over the surface of a sphere and created by projection through a vantage point at the center. Here, the observer rotates about this vantage point and the pattern of flow resembles the lines of latitude on a globe. In reality things are never as simple as this, voluntary head turns occurring about an axis some distance behind the eyes so that the latter always undergo some slight translation. Such second-order effects are ignored here. (From Miles et al., with permission.) (B) A cartoon showing the observer's limited field of view and the kind of motion experienced during rotation about a vertical axis as the observer looks straight out to the side. The speed of optic flow is greatest at the center ("equator") and decrements as the cosine of the angle of latitude. However, both the pattern and the speed of the optic flow at all points are determined entirely by the observer's motion—the 3-D structure of the scene is irrelevant. (From Miles, with permission.) (C) The pattern of optic flow experienced by the translating observer resembles the lines of longitude on a globe. (From Miles et al., with permission.) (D) A cartoon showing the centrifugal pattern of optic flow experienced by the observer who looks in the direction of heading—the black dot at the foot of the mountain. (From Bussetti et al., with permission.) (E) The optic flow experienced by the moving observer who looks off to the right but makes no compensatory eye movements so that the visual scene appears to pivot about the distant mountains (effective infinity). The speed of image motion is inversely proportional to the viewing distance. (After Miles et al., with permission.) (F) Again, the observer looks off to one side but here attempts to stabilize the retinal image of a particular object in the middle ground (tree), necessitating that he or she track to compensate for his or her own motion, thereby reversing the apparent motion of the more distant objects and creating a swirling pattern of optic flow. The scene now appears to pivot about the tree. (After Miles et al., with permission.)

Annals of the New York Academy of Sciences, Volume 871
observer's restricted field of view means that only a portion will be visible at any given time (as in Fig. 1B). In contrast, the optic flow experienced by the observer who undergoes pure translation (without compensating) consists of streams of images emerging from a focus of expansion straight ahead and disappearing into a focus of contraction behind, the overall pattern resembling the lines of longitude on a globe (see Fig. 1C). As during rotational disturbances, the direction of flow at any given point during translational disturbances depends solely on the motion of the observer. However, the speed of the flow at
any given point during translations depends not only on the observer’s speed of motion but also on the viewing distance at that location as nearby objects move across the field of view much more rapidly than more distant ones: motion parallax.\textsuperscript{16,15} A consequence of this is that, during translation, eye movements can stabilize only the images in one particular depth plane, and if the observer makes no active attempt to compensate, then only the images of the most distant objects are stable. Again, given the observer’s restricted field of view, the pattern of motion actually experienced depends on the direction in which the observer chooses to look: straight ahead the observer sees a radially expanding world (as in Fig. 1D), whereas off to one side the sensation is of the visual world pivoting around the far distance as images move across the field of view in inverse proportion to their viewing distance (as in Fig. 1E). If the observer undergoing translation wishes to fixate an object in the middle ground off to one side, then he or she must now compensate for the motion and, in so doing, his or her visual world will now rotate about this new object (as in Fig. 1F).

**VISUAL STABILIZATION**

The traditional approach to the visual stabilization of gaze is represented by the opto-kinetic response, which has two distinct components: an early component (OKNe) with brisk dynamics and a delayed component (OKNd) with sluggish dynamics.\textsuperscript{16}

*Images that Lie Off to One Side: Ocular Following*

Recent studies of OKNe have often employed large moving patterns backprojected onto a translucent tangent screen facing the observer because it offers much better control of the stimulus parameters and the responses evoked in this situation, which have very short latencies (<60 ms in monkeys and <85 ms in humans), have been termed “ocular following.”\textsuperscript{16,17} My colleagues and I have suggested that OKNd evolved as a visual backup to the RVOR, dealing with residual disturbances of gaze associated with rotations of the observer, whereas ocular following/OKNe evolved as a visual backup to the TVOR, dealing with residual disturbances of gaze associated with lateral translations of the observer (such as in Fig. 1E, 1F).\textsuperscript{11,16-26} Initial support for this idea rested largely on two observations: first, changes in the gain of the RVOR (resulting from exposure to magnifying or minifying spectacles) were associated with proportional changes in the gain of OKNd but not of OKNe.\textsuperscript{27} Second, changes in the gain of the TVOR (resulting from changes in the viewing distance) were associated with proportional changes in the gain of ocular following.\textsuperscript{2,11,18} Such changes in the gains of the visually driven responses were attributed to changes in central pathways that are shared with the vestibular reflexes, presumably reflecting functional synergies between the RVOR and OKNd on the one hand, and the GO-TVOR and OKNe on the other. The block diagrams in A and B of Figure 2 illustrate the two hypothesized visuovestibular mechanisms dealing independently with rotational and translational disturbances, and indicate the shared gain elements: in the case of the RVOR this gain element mediates long-term adaptive gain control, whereas in the case of the GO-TVOR this gain element gives the reflex its dependence on (the inverse of) the current viewing distance. Of course, this latter assumes some internal representation of viewing distance, perhaps using efference copy of cues such as vergence and/or accommodation.

Recent experiments indicate that ocular following has special built-in features for dealing with the visual problems posed when the moving observer looks off to one side, as in parts E and F of Figure 1. The visual task confronting the visual stabilization mechanisms
here is to single out the motion of particular elements in the scene, such as the mountain in Figure 1E and the tree in Figure 1F, and ignore all of the competing motion elsewhere. One way to achieve this would be to use attentional focusing mechanisms to spotlight the target of interest. Such mechanisms exist and are used by the so-called pursuit system, but have the limitation that they require high-level executive decisions to select the image to be tracked, and this of necessity is very time-consuming. A more rapid alternative might be to use low-level stereomechanisms that perform rapid parallel processing of binocular images, effectively sorting them on the basis of the depth plane that they occupy. This idea uses the fact that the object on which the two eyes are aligned (such as the mountain in Fig. 1E or the tree in Fig. 1F), which is said to reside in the plane of fixation, is imaged at corresponding positions on the two retinas. In contrast, objects that are nearer or farther than the plane of fixation have images that occupy noncorresponding positions on the two retinas and are said to have "binocular disparity." Clearly, a highly reliable algorithm for stabilizing gaze on objects in the plane of fixation would be to track only those images that lack disparity and to ignore all others. This would require neurons sensitive to both motion and binocular disparity, a combination known to be commonplace in the dorsal stream of cortex. Early support for a stereo algorithm was the finding that optokinetic responses deteriorate when the driving images have binocular disparity. However, high-level processing, perhaps involving selective attention, may have been a factor in these studies, which examined the closed-loop, steady-state responses. More recent experiments indicate that the early oculomotor following responses of both monkeys and humans can be disrupted by disparity before there can have been time for attentional mechanisms to operate. This is consistent with the notion that the motion detectors driving oculomotor following are also disparity selective. Thus, the oculomotor-following system helps to stabilize gaze on objects of interest not by selecting a particular one but by stabilizing the image of any object that happens to lie close to the plane of fixation, an implicit assumption therefore being that this plane contains the objects likely to be of most interest. This means that the time-consuming process of selecting the object of interest rests with the oculomotor subsystems that bring images into the plane of fixation, that is, the saccadic system working in concert with the vergence system. These latter systems redirect gaze to objects using higher-level criteria, whereas oculomotor following relies on low-level rapid parallel filters. Thus, the general concept is of low-level reflex systems stabilizing whatever images the high-level systems happen to bring into the plane of fixation.

There is strong evidence that oculomotor following derives at least some of its input from the medial superior temporal (MST) region of the cortex. Thus, chemical lesions in MST result in impairments of even the earliest components of oculomotor following and single-unit recordings in this region indicate the presence of many directionally selective neurons that discharge in close relation to the large-field, high-speed motion stimuli that are optimal for eliciting these motor responses.

Images that Lie Ahead: Radial-Flow Vergence

The visual challenge considered in the previous section on oculomotor following was that confronting the moving observer who looks off to one side. I now consider the gaze stability problems of the moving observer who looks in the direction of heading and so experiences radial patterns of optic flow such as that featured in Figure 1D. Insofar as the radial pattern of flow is associated with a change in viewing distance, it should be accompanied by increases in the vergence angle of the two eyes in order for the object of interest in the scene ahead to stay imaged on both foveas. Recent experiments on humans have indicated that radial optic flow elicits vergence eye movements at latencies that are closely comparable with the ultrashort values mentioned earlier for human
A: rotation

B: translation orthogonal to direction of gaze

C: translation isogonal to direction of gaze
FIGURE 2. Block diagrams showing the proposed linkages between the visual and vestibular reflexes operating to stabilize gaze. (A) The open-loop RVOR and the closed-loop OKN generate version eye movements, $E_v$, that compensate for rotational disturbances of the head, $H_v$. These reflexes share (a) a velocity storage element, which is responsible for the slow build-up in OKN and the gradual decay in RVOR with sustained rotational stimuli, and (b) a variable gain element, $G$, that mediates long-term regulation of RVOR gain. The characteristics of the spatial filter remain to be determined in primates. (After Miles et al. with permission.) (B) The open-loop GO-TVOR and the closed-loop OKNe/ocular following generate version eye movements, $E_v$, that compensate for translational disturbances of the head that are orthogonal to the direction of gaze, $H_v$, and affect gaze in inverse proportion to the viewing distance, $d$. These reflexes share (a) a variable gain element, $k/d$, that gives them their dependence on proximity, and (b) a fixed gain element, $k_s$, that generates a small response irrespective of proximity. The spatial filter is centered on the plane of fixation, effectively rejecting image motion in other depth planes. (After Schwartz et al. with permission.) (C) The open-loop GI-TVOR and the open-loop radial-flow–vergence mechanism generate vergence eye movements to help maintain binocular alignment on the object(s) of regard during translational disturbances of the head in the direction of gaze, $H_v$. The GI-TVOR is assumed to have an overall gain that is inversely proportional to the square of the viewing distance, as required by the optical geometry, and this is achieved by two gain elements in series, each having a gain inversely proportional to viewing distance, $d$. Radial-flow vergence affects gaze in inverse proportion to the viewing distance, $d$, and this is achieved by having the radial-flow input share the later part of the GI-TVOR pathway. A fixed gain element, $k_v$, once more generates a small response (this time, radial-flow vergence irrespective of proximity). The spatial filter is tuned to radial patterns of optic flow. (After Yang et al. with permission.) Dashed lines represent physical links: $H_v$, head velocity in linear coordinates; $E_v$, $G_s$, and $W_p$, velocity of head, eyes (in head), gaze and visual surroundings, respectively, in angular coordinates; $H_v$, $E_v$, and $G_v$, velocity of head, eyes (in head) and gaze in depth (equivalent vergence) coordinates. SCC, semicircular canals; OTO, otolith organs.

Occular following (<85 ms). Centrifugal (expanding) flow, which signals a forward approach and hence a decrease in the viewing distance, results in increased convergence, and centripetal (contracting) flow, which signals the converse, resulted in decreased convergence. Interestingly, the apparent changes in the sizes of objects as the observer moves closer or farther away also elicit convergence, but only at much longer, pursuit-like latencies generally estimated to be in excess of 200 ms.

The clear suggestion here is that the brain is able to sense the radial pattern of optic flow and to infer from this that there has been a change in viewing distance. However, a characteristic of the ocular responses to these radial-flow patterns is that each eye always moves in the direction of the net motion vector in the nasal hemifield, and this allows an alternative and less interesting explanation for the responses: the vergence might result from monocular tracking, in which each eye tracks only the motion that it sees and with a preference for the nasal hemifields. For example, with centrifugal flow the net motion vector in the nasal hemifields is toward the nose and each eye moves in that direction: hence, the increased convergence. That this was not the explanation was apparent from the observation that binocular vergence responses persisted, albeit weaker, when the nasal hemifields were masked off, consistent with the idea that the vergence responses result from a true parsing of the radial pattern of flow. Because latencies are so short, it is reasonable to assume that the system must depend on parallel processing to sense the pattern of flow. This is consistent with the idea that there are neurons or networks that act like templates or tuned filters so as to respond to radial patterns of optic flow. Once again the general concept is of a relatively low-level reflex system responding appropriately to whatever region of the optic flow field is brought into view by the high-level saccadic system.

Recent experiments have shown that radial-flow vergence responses are a linear function of the pre-existing vergence angle, and hence would be expected to share ocular
following's dependence on the reciprocal of the viewing distance. The etiology of this
dependence on viewing distance is assumed to be similar to that proposed for ocular
following—central pathways that are shared with the synergistic translational vestibular
reflexes, here the GI-TVOR. The block diagram in C of Figure 2 illustrates this hypothe-
sized linkage. The GI-TVOR is assumed to have an overall gain that is inversely propor-
tional to the square of the viewing distance as discussed earlier, and this is achieved by two
gain elements in series, each having a value that varies with the inverse of the viewing dis-
tance, d. However, because the radial-flow vergence responses modulate only in inverse
proportion to the viewing distance, it is assumed that only the later part of the GI-TVOR
pathway is shared.

There is extensive evidence that area MST in the monkey's cortex contains neurons that
are selectively sensitive to radial optic flow patterns such as those now known to evoke
vergence eye movements at ultrashort latencies. In fact, MST is the first stage in the
so-called dorsal (motion) pathway at which global flow is encoded at the level of single
cells; at earlier stages, such as MT, individual cells have much smaller receptive fields and
encode only local motion. It is tempting to assume that radial-flow vergence, like its
sister reflex, ocular following, is also mediated by MST.

Other Short-Latency Mechanisms?

There is at least one more short-latency visual tracking mechanism that can be revealed
with large-field stimuli. This third mechanism generates vergence eye movements in
response to binocular disparity and so would be expected to help maintain binocular align-
ment during motion in depth. However, this is a secondary function rather than a primary
one because this disparity vergence mechanism operates to maintain the vertical—as well
as the horizontal—alignment of the two eyes, a function clearly unrelated to motion of the
observer per se. Nonetheless, this reflex has much in common with the other two that have
been discussed, in addition to an ultrashort latency. Thus, like ocular following and
radial-flow vergence, this disparity—vergence mechanism shows dependence on a prior
saccade, whereby stimuli presented in the immediate wake of a saccade are much more
effective than the same stimuli presented some time later; this post-saccadic enhancement
is largely visual, resulting from the shift in the image of the world on the retina induced by
the saccade and can be simulated by saccadlike shifts of the visual scene. There is also evi-
dence from single-unit recordings that MST contains neurons that discharge closely in relation
to these disparity vergence responses. Last, as with ocular following and radial-flow
vergence, the ultrashort latency suggests that subjects respond before they can even be
aware that there has been a stimulus, that is, the responses are independent of perception.
In this regard, it has recently been shown that the short-latency disparity—vergence
responses can be produced by applying the disparity steps to patterns that do not give rise
to the perception of depth because they have opposite contrast at the two eyes, so-called
anticorrelated images. This is consistent with the idea that these short-latency vergence
responses derive their visual input from an early stage of cortical processing prior to the
level at which depth percepts are elaborated.

CLOSING REMARKS

The three visual stabilization mechanisms share a number of features in addition to
their ultrashort latencies, and it has been argued that all are involved in generating eye
movements to compensate for translational disturbances, two being highly specialized for
this purpose. This has led to the suggestion that these mechanisms constitute a family of
reflexes, and Table 1 summarizes our current knowledge of their fundamental similarities
and differences.

<table>
<thead>
<tr>
<th>Table 1: Major features of the three visual stabilization mechanisms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Function</strong></td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Input</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Binocular motion in plane of fixation</td>
</tr>
<tr>
<td>Horizontal/vertical version</td>
</tr>
<tr>
<td>Latency</td>
</tr>
<tr>
<td>Postnasal enhancement</td>
</tr>
<tr>
<td>Adaptive gain control</td>
</tr>
<tr>
<td>Dependence on viewing distance</td>
</tr>
</tbody>
</table>

Source: After Miles, with permission.

REFERENCES

6. Coia, A. & H. L. Baljana. 1995. Providing distinct vergence and version dynamics in a bi-


