Acta Psychologica 82 (1993) 115–138
North-Holland

Neural representations for sensory-motor control, I: Head-centered 3-D target positions from opponent eye commands

Douglas Greve *, Stephen Grossberg **, Frank Guenther * and Daniel Bullock *

Boston University, Boston, USA

This article describes how corollary discharges from outflow eye movement commands can be transformed by two stages of opponent neural processing into a head-centered representation of 3-D target position. This representation implicitly defines a cyclopean coordinate system whose variables approximate the binocular vergence and spherical horizontal and vertical angles with respect to the observer's head. Various psychophysical data concerning binocular distance perception and reaching behavior are clarified by this representation. The representation provides a foundation for learning head-centered and body-centered invariant representations of both foveated and non-foveated 3-D target positions. It also enables a solution to be developed of the classical motor equivalence problem, whereby many different joint configurations of a redundant manipulator can all be used to realize a desired trajectory in 3-D space.

Spatial representations for the neural control of flexible movements

The present article introduces a neural network model of how the brain forms spatial representations with which to control sensory-guided and memory-guided eye and limb movements. These spatial

Correspondence to: S. Grossberg, Center for Adaptive Systems, Boston University, 111 Cummington Street, Boston, MA 02215, USA.

* Supported in part by National Science Foundation (NSF IRI-87-16960 and NSF IRI-90-24877).

** Supported in part by Air Force Office of Scientific Research (AFOSR UR1 90-0175), Defense Advanced Research Projects Agency (DARPA AFOSR-90-0083), and National Science Foundation (NSF IRI-87-16960 and NSF IRI-90-24877)

The authors wish to thank Carol Y. Jefferson for her valuable assistance in the preparation of the manuscript.

representations are expressed in both head-centered coordinates and body-centered coordinates because the eyes move within the head, whereas the head, arms, and legs move with respect to the body. We analyze a key process in the formation of spatial representations whereby humans and other mammals can skillfully act upon objects in 3-dimensional space despite the variable relative location of sensing and acting segments.

The spatial relationships to which we have been led are built up from the same types of computations that are used to control motor commands. This observation leads to a general design theme of our work. We inquire into the natural form of neural computations that are appropriate for representation and control of a bilaterally symmetric body. Bilateral symmetry leads to the use of competitive and cooperative interactions among bilaterally symmetric body segments. These include opponent interactions between pairs of antagonistic neurons that measure one or another type of spatial or motor offset with respect to an axis of symmetry. Based on the present results, we show elsewhere how a 3-dimensional space can self-organize and learn to control synchronous variable-speed and variable-size trajectories of a 3-joint arm, with or without a tool of variable length.

Geometry and psychophysics of object localization

This section surveys key geometrical and psychophysical data pertinent to the model. The following two sections describe how two successive stages of opponent interactions based on eye movement outflow commands can generate the type of head-centered representation that is suggested by these data. These results have been announced in Bullock et al. (1992a).

During eye–hand coordination, both eyes typically fixate a target before or while a hand reaches towards it. Vision, in particular the binocular disparity of an object's image on the retinas of both eyes, provides important cues to the relative 3-D positions of objects with respect to the head. However, such visual information cannot explain all cases of accurate reaching toward binocularly foveated targets. In particular, it cannot explain cases of 'blind reaching' (e.g. Soechting and Flanders 1989). In blind reaching experiments, a subject first binocularly foveates a target, then reaches to the target without

on-line visual feedback regarding the relative positions of the target and the moving hand. The approximate accuracy of such blind reaches suggests the availability of absolute distance information.

This is important because binocular disparity does not provide unambiguous information about absolute distance. For example, suppose that the eyes binocularly foveate a succession of locations in the interior of an object. During this series of fixations, a binocular disparity can be computed for any point P on the boundary. However, though distance from P to the observer is invariant during the fixation series, the binocular disparities computed for P may be quite variable. Moreover, conditions are even worse at the fixation point itself. When both eyes fixate the same location in space, then the binocular disparity of this location on the retinas equals zero, no matter how near or far the object may be from the observer. Thus, fixated points cannot accurately be reached using information about such points' own binocular disparity. Since our primary goal in the present article is to analyse how reaching towards fixated objects is controlled, we need to consider other sources of information than retinal, or visual, information.

The bilaterally symmetric organization of the body provides another, non-visual source of information for computing absolute distance of a fixated target from an observer's head and body. When both eyes binocularly fixate a target, the point of intersection of the lines of gaze may be used to compute the absolute distance and direction of the fixation point with respect to the head. Such extraretinal information may also be used to complement visual processing to derive better estimates of the absolute distance and direction of visually detected but non-fixated objects.

The intersection point of the lines of gaze moves with the mobile eyes within a roughly conical 3-D volume that opens out in front of the head with apex between the eyes and horizontal and vertical bounds determined by the limits of ocular rotation. Clues to the nature of this 3-D coordinate system can be found in the experimental literature on the role of extraretinal information in visual object localization. For example, Foley (1980) has summarized evidence that extraretinal signals are used to compute the absolute distance between a binocular fixation point and the midpoint between the eyes, which we will call the cranial egocenter. If we take this radial distance, R_H, as one dimension of a 3-D coordinate system, it suggests the

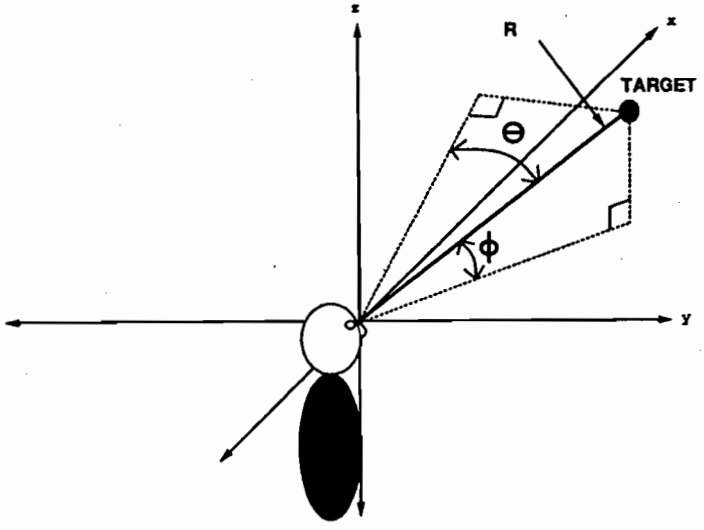

Fig. 1. Illustration of relationships between spherical coordinates ϕ, θ, R and Cartesian coordinates x, y, z. Both coordinate systems have origins centered between the eyes. The $x-z$ plane origin is the midpoint of a y-axis segment drawn between the ocular centers of rotation, and the z-axis is parallel to the gravity vector during upright posture. Thus the x-axis always points 'straight ahead'. Radius R is measured from the origin to the binocular fixation point on the object. Elevation ϕ (left panel) is the angle between the radius and a line in the $x-y$ plane. This line connects the origin to the point where a ray from the fixation point is normal to the $x-y$ plane. Azimuth θ is defined similarly, but with respect to the $x-z$ plane.

relevance of a spherical egocentric coordinate system for 3-D object localization, in which the other two coordinates are horizontal angle or azimuth, θ_H, and vertical angle or elevation, ϕ_H, as shown in fig. 1. More direct evidence for the use of a representation akin to a spherical egocentric coordinate system came from a recent study of the accuracy of pointing to objects in nearby space without simultaneous vision of hand and object (Soechting and Flanders 1989). This task is pertinent because, without simultaneous vision of object and hand, pointing errors are sensitive to errors in locating the point of fixation on the object relative to the body prior to the reach. Soechting and Flanders concluded that an egocentric spherical coordinate representation of spatial location with respect to the shoulder gave a more parsimonious account of variability in the data than either a Cartesian or a cylindrical coordinate representation.

An internal spatial representation akin to spherical coordinates has

several advantages for the control of reaching. Movements of the arm with wrist and finger joints fixed are readily represented in a spherical coordinate frame based at the shoulder. Arm movements due to shoulder rotation correspond to changes primarily in the spherical angles θ and ϕ, whereas bending of the elbow relates primarily to the spherical coordinate R. These properties simplify the task of transforming from spatial coordinates to arm trajectories, as hypothesized by Soechting and Flanders (1989). The close relationship between spherical spatial coordinates and joint coordinates is illustrated in fig. 2. Hollerbach et al. (1986) plotted trajectories for the fingertip, wrist, elbow, and shoulder during free reaches in the sagittal plane through the shoulder. The form of these trajectories, reprinted in fig. 2a, led Hollerbach et al. to propose that they were produced by linearly interpolating between initial and final coordinates in the joint space defined by arm geometry. Fig. 2b shows trajectories that we have simulated between the endpoints of the fingertip paths of Hollerbach et al. (1986). However, our simulations used linear interpolation in a head-centered spherical coordinate space. The correspondence between the simulated trajectories in fig. 2b and the human trajectories in fig. 2a indicates that the observed trajectories are as consistent with spherical space interpolation as with joint-space interpolation. The same could not be said for Cartesian space interpolation.

Although comparing the relative merits of spherical and Cartesian coordinates is of heuristic value, such comparisons ultimately cannot solve the problem of 3-D spatial representation by the brain. Coordinate values such as R_H, θ_H, and ϕ_H can, at best, be computed implicitly as an emergent property of interacting, locally computed neural variables. The values of these variables can, moreover, change adaptively as a behaving animal develops and grows in ever-changing internal and external environments. Our task is to understand how the data may approximate values expected from spherical coordinates in some situations, but not in others. In particular, fingertip trajectories forming nearly straight lines in Cartesian coordinates are found in some parts of the arm movement workspace (Morasso 1981).

Our analysis considers the types of information about position of the eyes in the head, the head in the body, and the arms in the body that may combine to generate an implicit representation of target position with respect to the head, body, or arm. In order to understand how these representations may arise, it is useful to consider

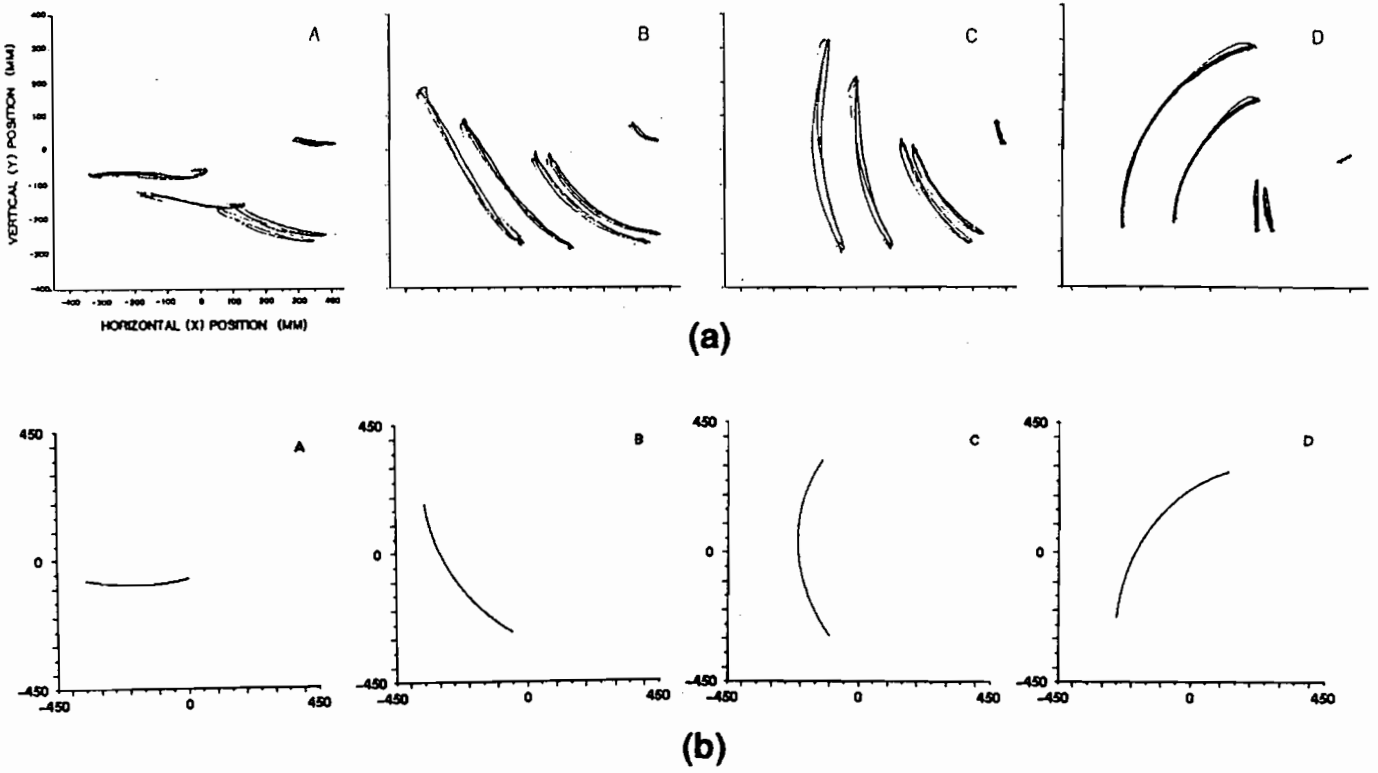

Fig. 2. (a) Hollerbach et al. (1986) data from fingertip, wrist, elbow, and shoulder trajectories during free reaches in the sagittal plane through the shoulder (reprinted with permission); (b) a simulation using linear interpolation in spherical coordinate space can reproduce the Hollerbach et al. (1986) fingertip trajectories. Units are millimeters, with shoulder located near (400,0).

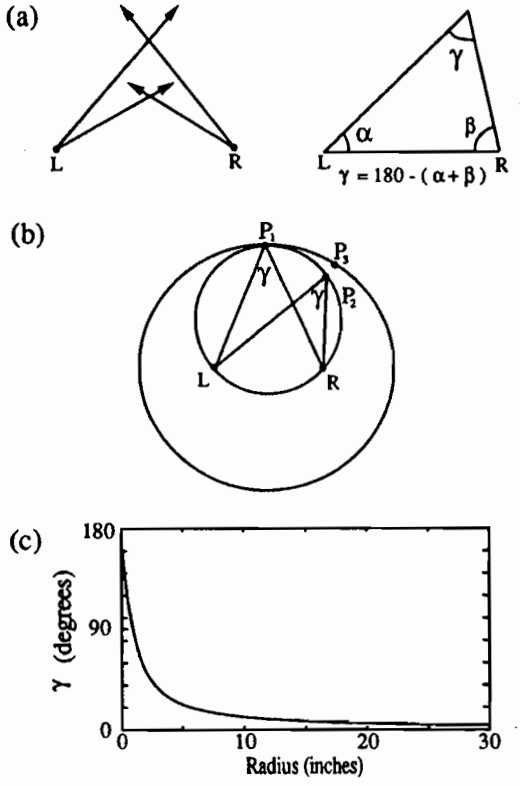

Fig. 3. The geometry of binocular fixation of 3-D target positions by the left (L) and right (R) eyes. See the text for details.

geometrical relations that obtain once a binocular system has successfully foveated an object.

Fig. 3a illustrates that the intersection point of the lines of sight of the two eyes converges toward the nose as the two eyes rotate to foveate increasingly close objects that are straight ahead. The rotation centers of the two eyes together with the fixated point on the object form a triangle, whose three angles add to 180 degrees. The angles of the two eyes in their orbits thus jointly specify the angle γ between the lines of sight that intersect at the fixation point. Angle γ is commonly referred to as the *binocular parallax* (Foley 1980). This triangular structure linking observer and object also allows an internal measure of net ocular *vergence* – the extent to which the eyes are rotated towards the nose – to serve as one basis for estimating the distance from egocenter to a binocularly foveated object.

In order to understand how the parameters α, β, and γ in fig. 3a scale with viewing distance and the line of sight, two questions need to be answered: How well does a fixed binocular parallax, γ, or equivalently a constant sum $\alpha + \beta$ of the nasal deviations of the two eyes in their orbits, represent a constant distance from the egocenter to a fixation point on the object as α and β vary? How well does binocular parallax estimate the distance along any given line of sight?

To answer the first question, consider the two circles in fig. 3b. The smaller circle, known as a Vieth-Müller circle (Blank 1978), is a curve passing through a point P_1 that is located directly in front of an observer's eyes, and through the rotation centers, L and R, of the eyes themselves. The Vieth-Müller (V-M) circle is special because all binocularly fixated points on the circle have the same binocular parallax γ. The larger circle in fig. 3b is the curve formed by points whose distances equal that of P_1 from the egocenter. The divergence of these two circles indicates that binocular parallax, by itself, is not an accurate measure of absolute radial distance from the egocenter. Likewise, internal signals that measure only net binocular vergence cannot veridically specify absolute radial distance.

The V-M circle shows that binocular parallax is an increasingly poor measure of radial distance at extreme gaze angles (large α or β). Fig. 3c shows a second potential shortcoming of using binocular parallax by itself as an indicator of distance from egocenter, even for objects that lie straight ahead. The function relating an object's absolute distance from egocenter to its binocular parallax is markedly nonlinear. This means that any neural sites whose activity levels veridically register vergence angle would use a disproportionate part of their dynamic range to represent a relatively small zone of nearby R_H values. Although this nonlinearity can be compensated by a nonlinear neural transduction, far-distance estimates would remain intrinsically more prone to errors caused by inaccuracies of vergence or by noise at the cellular sites that register vergence.

Although vergence, by itself, is insufficient to measure binocular distance, it is now generally accepted that a vergence-related signal is used in binocular perception. For example, stereoscope experiments have shown that a decrease in apparent stimulus size can be caused by holding the retinal size of an artificial binocular stimulus constant, while increasing the vergence angle needed for fusion (Rock 1984). Although this demonstration shows that the system is susceptible to

illusions, the observed relationship between vergence and apparent size is best explained as a by-product of nature's solution to the problem of computing true stimulus size. Because an object of fixed size subtends a smaller retinal region the more distant it is from an observer, retinal image size is not an accurate measure of true object size. By combining retinal size and binocular disparity with a distance measure derived from vergence angle, the nervous system can compute a better estimate of true object size than that afforded by retinal size alone.

On the basis of an extensive survey of prior theory and data about binocular distance perception, Foley concluded that an egocentric distance signal does exist, and that it 'appears to be of extraretinal origin and ... related to the vergence of the eyes' (1980: 411). He was able to model a large corpus of data from several kinds of experiments on matching of apparent distance. His model computes an effective binocular parallax γ' prior to transformation into perceived distance. The effective binocular parallax γ' does not equal the real binocular parallax. Foley reviewed evidence that the relation between the psychological variable γ' and the physical variable γ is well approximated by:

$$\gamma' = A + B\gamma,\ 0 < A < 2°,\ 0 < B < 1. \tag{1}$$

By (1), veridical registration of binocular parallax obtains if $A = 0$ and $B = 1$. Because the perceptual system acts as if $A > 0$, even objects at optical infinity ($\gamma = 0$) are perceptually registered at finite distance. Because $B < 1$, nearby objects ($\gamma \gg 0$) are registered as farther than they are. In the fourth section, we show that values for A and B in the range observed by Foley emerge naturally within a simple neural network for computing vergence from opponent combinations of oculomotor outflow signals. When optimally tuned, this neural network approximates the ideal values $A = 0$ and $B = 1$.

Another important psychophysical issue is raised by the need, illustrated in fig. 3b, to increasingly amplify the effect of the vergence signal on perceived closeness as the angle of the foveated object relative to straight-ahead increases. Such an amplification has been observed by Blank, who noted that 'an actual circle of apparent equidistance from the observer is somewhat flatter than a Vieth-Müller circle' (1978: 89). This means that in a perceived-distance matching

task, a target initially at P_2 in fig. 3b would be pushed outward by the subject toward point P_3 to create a target perceived to be at a distance from the cranial egocenter equal to the distance of fixed point P_1. Thus a target at P_2 is perceived closer than P_1, consistent with physical reality and with the hypothesis that apparent distance is influenced both by the vergence γ and the head-centered deviation θ_H. For one well-studied subject who matched the apparent distances of a set of targets to a reference target while elevation ϕ_H was held constant, Blank (1978: 95) found this interaction to be well approximated by:

$$G(\gamma,\theta_H) = \frac{\gamma}{1 - 0.19\theta_H^2}. \tag{2}$$

In the fourth section, we describe a neural network for 3-D localization that clarifies how variables such as γ, θ_H, and ϕ_H are neurally estimated.

Adaptive linearization via cerebellar learning of motor responses to outflow eye movement commands

An internal representation of the 3-D location of a target with respect to the head can be derived from the commands that hold the eyes in place while they are fixating the target. These commands are calibrated in motor coordinates that are capable of controlling eye muscle contractions and relaxations. Fig. 4 indicates the movements of each eye that are caused by each of its six extraocular muscles when it operates alone. The agonist–antagonist muscle pair consisting of the lateral rectus and medial rectus controls the horizontal angle, θ, of an eye. Two agonist–antagonist pairs, the inferior oblique and superior oblique, and the superior rectus and inferior rectus, work together to control the vertical angle, ϕ, of the eye. These angles were defined in fig. 1, calibrated with respect to the egocenter. Fig. 5 illustrates that a target has different angles with respect to the egocenter and each of the two ocular centers of rotation. The subscript L denotes spherical coordinates with the origin at the left eye, subscript R denotes an origin at the right eye, and subscript H denotes an origin centered between the left and right eyes.

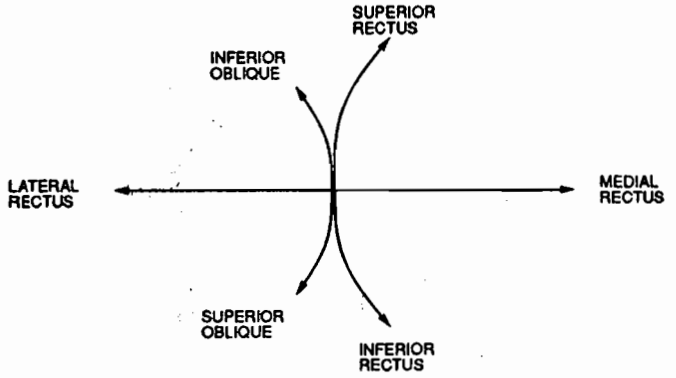

Fig. 4. The direction of movement caused by each muscle of the eye when acting alone. The diagram corresponds to the right eye as seen by an observer. Thus the medical rectus pulls the eye noseward.

It is shown below how signals derived from the commands that control the left eye and the right eye can be combined to create a head-centered 'cyclopean' representation that is centered between the two eyes. These signals are corollary discharges of the outflow movement commands (von Helmholtz 1962). Corollary discharges change linearly with outflow movement commands, but do not themselves cause movements. In order for a corollary discharge to accurately represent eye position, the eye muscles must contract linearly in response to these outflow movement commands. Equal changes in these commands need to cause (approximately) equal eye rotations regardless of the eye's initial position in its orbit. It is, however, known that the eye muscle plant is nonlinear (Robinson 1970; Schiller 1970), and that its characteristics can change during the lifespan of an individual. Grossberg and Kuperstein (1986, 1989: chapter 5) have shown how the mismatch between outflow signals (derived from movement commands) and inflow signals (derived from muscle sensors) may combine to define an error signal that calibrates how nonlinear the muscle plant is at each commanded eye position. These mismatch signals drive an error-based learning process that is suggested to take place in the cerebellum. Learned cerebellar output signals are suggested to modify the total movement command in a way that adaptively linearizes the response of a nonlinear muscle plant to the movement commands that it received. As a result of this learned compensation, corollary discharges of outflow movement commands

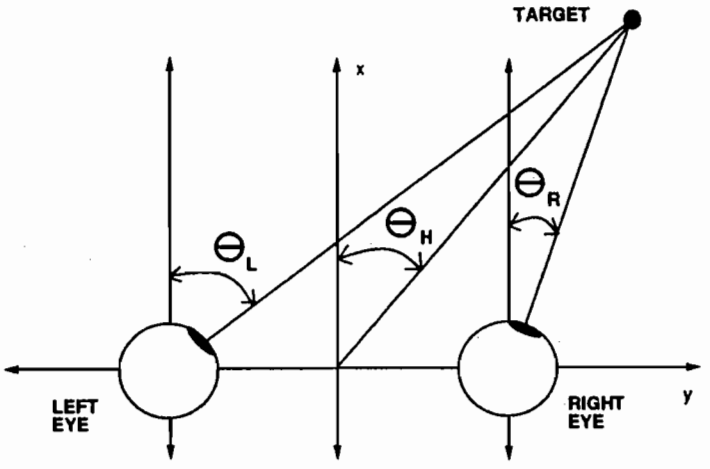

Fig. 5. The angles θ for the three different origins corresponding to a target projected onto the $x-y$ plane.

can be used as eye position signals even if the muscle plant's nonlinear characteristics change through time.

Grossberg and Kuperstein have used this model to explain a variety of behavioral and neural data concerning eye and arm movements, including data about pointing behavior after strabismus surgery (Steinbach and Smith 1981), the role of the cerebellum in preventing dysmetria (Robinson 1973; Vilis et al. 1983), and the existence of saccade-related direct response cells in the dentate nucleus of the cerebellum (Ron and Robinson 1973).

Opponent interactions generate head-centered coordinates for representation of 3-D space

We now show how to binocularly combine outflow signals from the tonically active cells that control the position of each eye (fig. 6) in such a way as to form a head-centered representation of a foveated target. First, opponent interactions combine the outputs of the cells that control the agonist and antagonist muscle of each eye (fig. 7). These opponent interactions give rise to opponent pairs of cells whose total activity is approximately constant, or normalized. This normalization property renders the outflow movement commands of each eye

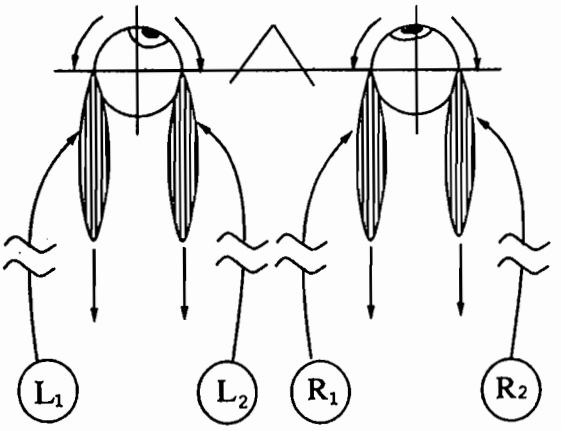

Fig. 6. Notation for the horizontal outflow signals that control the opponent muscles of both eyes.

'dimensionless', and allows the normalized opponent cells to control a variety of eye movement processes, notably corollary discharges, that might otherwise be improperly scaled (Grossberg and Kuperstein 1989).

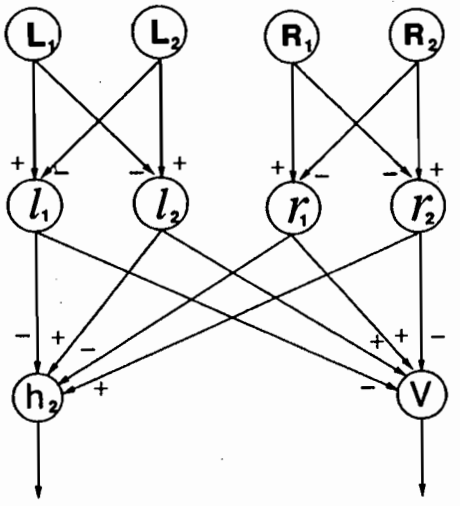

Fig. 7. Network for combining corollary discharges from both eyes, via two stages of opponent processing, into two components of a head-centered representation of 3-D target position. The first stage computes normalized activities l_i and r_i. The second stage uses these normalized activities as inputs to compute an estimate h_2 of horizontal angle and an estimate V of binocular vergence, which correlates with radial distance. Plus and minus signs indicate excitation and inhibition, not addition and subtraction.

Next, the normalized outputs from both eyes are combined in two different ways to generate the type of head-centered spatial representation of the binocular fixation point that was suggested by the data reviewed in the second section. This head-centered representation arises from simple rules for opponent processing of the normalized eye movement corollary discharges. In particular, opponent cells from each eye generate inputs of opposite sign (excitatory and inhibitory) to their target cells at the next processing stage. As illustrated in fig. 7, one combination gives rise to cells whose activities approximate the angular spherical coordinate θ_H. The other combination gives rise to cells whose activities approximate the binocular vergence γ, which in turn can be used to estimate the radial distance R_H. The two combinations generate head-centered coordinates by computing a sum and a difference of the normalized opponent inputs from both eyes. Such a general strategy for combining signals is well-known in other neural systems, such as color vision. For example, a sum $L+M$ of signals from two color vision channels estimates luminance, whereas a difference $L-M$ estimates color (DeValois and DeValois 1975; Mollon and Sharpe 1983). Thus the computations that may be used to control reaching in 3-D space seem to derive from a broadly used principle of neural computation.

A neural mechanism for normalizing the total activity of opponent cells is well-known (Grossberg 1982). It uses a shunting on-center off-surround network; that is, an opponent interaction wherein the target cells obey a membrane equation (Hodgkin 1964; Katz 1966). In particular, suppose that the agonist and antagonist cells that control the horizontal position of the left eye have activities L_1 and L_2, respectively, as in fig. 6. Let the normalized opponent cells in the shunting network have activities l_1 and l_2. Suppose that

$$\frac{\mathrm{d}}{\mathrm{d}t}l_1 = -Cl_1 + (1-l_1)L_1 - l_1L_2, \tag{3}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}l_2 = -Cl_2 + (1-l_2)L_2 - l_2L_1. \tag{4}$$

By eq. (3), activity L_1 excites l_1 whereas activity L_2 inhibits l_1. The

opposite is true in eq. (4). Parameter C is the decay rate. At equilibrium, $(\mathrm{d}/\mathrm{d}t)l_1 = (\mathrm{d}/\mathrm{d}t)l_2 = 0$, so (3) and (4) imply that

$$l_1 = \frac{L_1}{C + L_1 + L_2}, \tag{5}$$

and

$$l_2 = \frac{L_2}{C + L_1 + L_2}. \tag{6}$$

Adding (5) and (6) shows that

$$l_1 + l_2 = \frac{L_1 + L_2}{C + L_1 + L_2}. \tag{7}$$

Thus if $C \ll L_1 + L_2$,

$$l_1 + l_2 \cong 1. \tag{8}$$

The approximation (8) will be used below for all normalized pairs of opponent cells. In particular, we assume that the activities of opponent cell populations that control agonist–antagonist muscle pairs are normalized so that the total activity of each cellular pair is fixed at unity. This ensures that increasing the activity of the agonist control cell results in a corresponding decrease in the activity of its antagonist control cell. To illustrate these relationships, fig. 7 shows the two cellular pairs needed to control θ_{L} and θ_{R}. These pairs are labelled by the variables l_1, l_2 and r_1, r_2, which measure corresponding cellular activities. Thus, the following equations define the internal representations of the horizontal angle of each eye:

$$l_1 + l_2 = 1 \tag{9}$$

$$\theta_L = -90° + 180° \times l_2, \tag{10}$$

$$r_1 + r_2 = 1, \tag{11}$$

$$\theta_R = -90° + 180° \times r_2, \tag{12}$$

where l_i indicates the activity of left eye cell population i and r_i indicates the activity of right eye cell population i. These equations assume that the eye muscles respond linearly to their outflow movement signals as a result of adaptive linearization (see previous section).

Internal representations for the vertical angles of left and right eyes may be defined similarly. Thus

$$l_3 + l_4 = 1, \tag{13}$$

$$\phi_L = -90° + 180° \times l_4, \tag{14}$$

$$r_3 + r_4 = 1, \tag{15}$$

$$\phi_R = -90° + 180° \times r_4. \tag{16}$$

To provide a head-centered representation of 3-D space – in particular, of the binocular fixation point – corollary discharges of the two eyes are combined. Let the cell populations $h_i, i = 1, 2, \ldots, 6$, form the basis for this head-centered spatial representation. These populations are also arranged in antagonistic pairs. First we show how h_1, h_2, h_3, and h_4 closely approximate the following linear relationships with respect to θ_H and ϕ_H:

$$h_1 + h_2 = 1, \tag{17}$$

$$\theta_H = -90° + 180° \times h_2, \tag{18}$$

$$h_3 + h_4 = 1, \tag{19}$$

$$\phi_H = -90° + 180° \times h_4. \tag{20}$$

These veridical head-centered binocular representations of θ_H and ϕ_H emerge by simply averaging the corresponding monocular components derived from the left and right eye muscle command corollary discharges using a shunting on-center off-surround network. The connectivity of such a network is shown in fig. 7 for the cell activity h_2 which represents θ_H. In particular,

$$\frac{\mathrm{d}}{\mathrm{d}t} h_2 = -Dh_2 + (1 - h_2)(l_2 + r_2) - h_2(l_1 + r_1), \tag{21}$$

where D is the decay rate. Solving this equation at equilibrium ($dh_1/dt = 0$) yields

$$h_2 = \frac{l_2 + r_2}{D + l_1 + r_1 + l_2 + r_2}. \tag{22}$$

Since $l_1 + l_2 \cong 1$ and $r_1 + r_2 \cong 1$, choosing a small decay parameter D leads to the approximation:

$$h_2 \cong \frac{l_2 + r_2}{2}. \tag{23}$$

Likewise, an antagonist cell with activity h_1 can be created by exchanging indices 1 and 2 throughout eq. (21). The equilibrium activity h_1 of this cell would be approximated by

$$h_1 \cong \frac{l_1 + r_1}{2}, \tag{24}$$

so that, by (23) and (24),

$$h_1 + h_2 \cong 1. \tag{25}$$

Activities h_3 and h_4 can be similarly computed from corollary discharges of vertical muscle command cells to form an antagonistic internal representation of ϕ_H. To evaluate the adequacy of the internal representation of θ_H and ϕ_H, a distortion measure was calculated by dividing the change in the internally represented angle of two successively foveated points by the actual change in angle of the successively foveated points for small changes throughout the workspace. For example, Let $\theta_{\text{int}}(\theta_H, R_H)$ be the internally represented horizontal angle, as defined by (18) and (23), that arises within the model when the muscle length commands l_i and r_i are set to the values needed to achieve binocular foveation of a point in space with an actual horizontal angle θ_H and radius R_H. The internal representation for θ_H is independent of vertical angle ϕ_H. The distortion measure at (θ_H, R_H) is then calculated as:

$$\%\ \text{distortion}(\theta_H, R_H) = \left(\frac{\theta_{\text{int}}(\theta_H + \Delta\theta_H, R_H) - \theta_{\text{int}}(\theta_H, R_H)}{\Delta\theta_H} - 1 \right) \times 100. \tag{26}$$

Here the ratio of internal to external angular changes is compared to a reference value of one because a gain of one between input and output is required for veridical registration of θ_H. A $\Delta\theta_H$ of 1° was chosen for the plots. Given some point in head-centered space (R_H, θ_H, ϕ_H) and a distance d between the eyes, the angles that each eye must assume in order to foveate that point are given by the following equations, which are derived from the geometry shown in figs. 1 and 5:

$$\theta_L = \tan^{-1}\left(\frac{R_H \sin\theta_H + \frac{d}{2}}{R_H \cos\theta_H}\right). \tag{27}$$

$$\theta_R = \tan^{-1}\left(\frac{R_H \sin\theta_H - \frac{d}{2}}{R_H \cos\theta_H}\right). \tag{28}$$

$$\phi_L = \sin^{-1}\left(\frac{R_H \sin\phi_H}{\sqrt{\left(R_H \sin\theta_H + \frac{d}{2}\right)^2 + (R_H \cos\theta_H)^2}}\right). \tag{29}$$

$$\phi_R = \sin^{-1}\left(\frac{R_H \sin\phi_H}{\sqrt{\left(R_H \sin\theta_H - \frac{d}{2}\right)^2 + \left(\frac{d}{2} - R_H \cos\theta_H\right)^2}}\right). \tag{30}$$

The distortion measure was calculated for a workspace defined by $-45° < \theta_H < 45°$, $-45° < \phi_H < 45°$, and 3 inches $< R_H <$ 30 inches (7.6 cm $< R_H <$ 76 cm). This workspace was chosen to approximate the cone within which both binocular foveation and reaching to a target are possible in humans.

Fig. 8 shows the distortion measure for the internal representation of θ_H throughout the workspace. The distortion in this range is less than 15%, with essentially 0% distortion for $R_H > 5$ inches. Thus, the

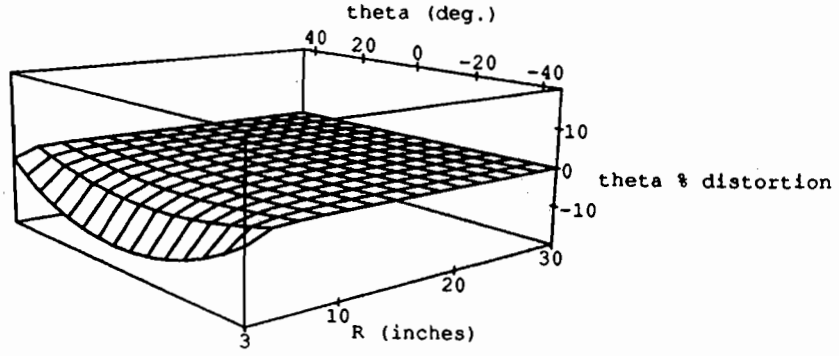

Fig. 8. Percent distortion for θ_H throughout the workspace. See text for details on distortion measure.

opponent network defined above provides an accurate mechanism for computing an internal representation of θ_H.

The distortion measure for ϕ_H depends on R_H, ϕ_H, and θ_H. Figs. 9, 10, and 11 show the distortion measure for ϕ_H with $\theta_H = 0°$, 22.5°, and 45°, respectively. For $\theta_H = 0°$ and 22.5°, distortion is again less than 15% everywhere and essentially 0% for $R_H > 5$ inches. For $\theta_H = 45°$, the distortion goes as high as almost 50% for very small R_H, but again is very small for $R_H > 5$ inches or for $\phi_H < 30°$. Thus, the normalized binocular opponent network depicted in fig. 7 provides an accurate internal representation of ϕ_H in all but the most extreme portions of the workspace.

To explain how opponent computation leads to a representation of vergence, note that vergence is equal to the difference between r_1

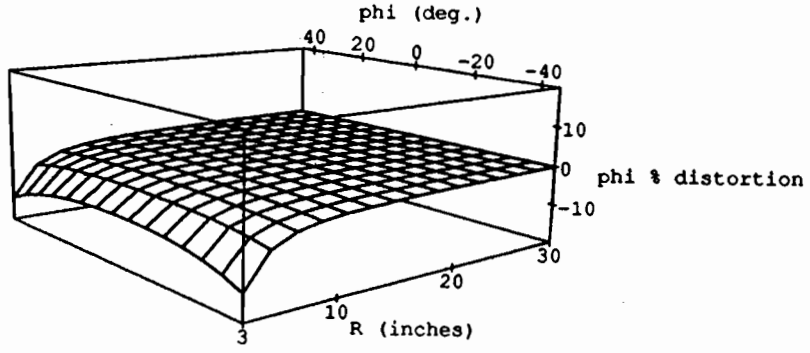

Fig. 9. Percent distortion for ϕ_H with $\theta_H = 0°$. See text for details on distortion measure.

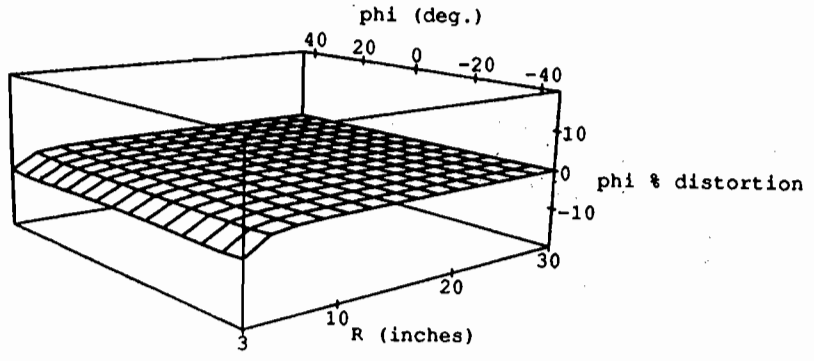

Fig. 10. Percent distortion for ϕ_H with $\theta_H = 22.5°$. See text for details on distortion measure.

(the outflow command to the medial rectus of the right eye) and l_1 (corresponding to the lateral rectus of the left eye). As in fig. 7, define a cell population with activity V (for vergence) which receives excitatory inputs l_2 and r_1 from cells controlling the medial recti of both eyes and inhibitory inputs l_1 and r_2 from cells controlling the lateral recti of both eyes. Then its activity will be governed by

$$\frac{dV}{dt} = -EV + (1-V)(l_2 + r_1) - (V+F)(l_1 + r_2). \tag{31}$$

At equilibrium,

$$V = \frac{l_2 + r_1 - Fl_1 - Fr_2}{E + r_1 + r_2 + l_1 + l_2}. \tag{32}$$

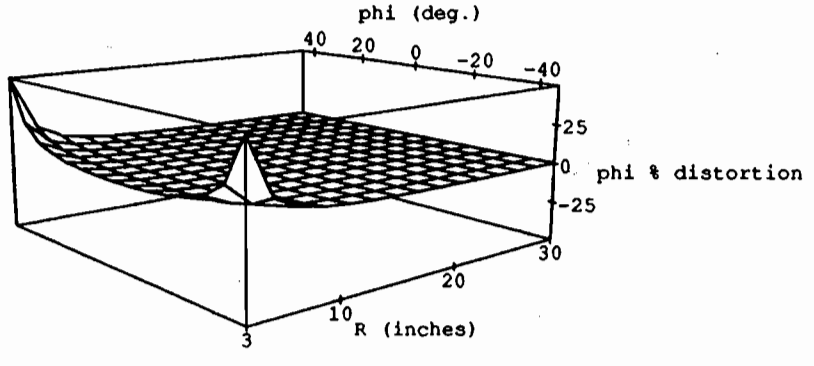

Fig. 11. Percent distortion for ϕ_H with $\theta_H = 45°$. See text for details on distortion measure.

Because $r_1 + r_2 = 1$ and $l_1 + l_2 = 1$, eq. (32) can be rewritten as

$$V = \frac{1-F}{E+2} + \frac{1+F}{E+2}(r_1 - l_1). \tag{33}$$

This expression suggests a mechanistic explanation of eq. (1), Foley's formal psychophysical function for subjective binocular parallax. Note first that if $F = 1$ and $E = 0$, then

$$V = r_1 - l_1. \tag{34}$$

In this case, subjective parallax equalled physical parallax. If, however, $E > 0$ and $F < 1$, then the slope $(1+F)(E+2)^{-1}$ of V versus $r_1 - l_1$ is less than one, consistent with Foley's estimate of $B < 1$ in (1) from the data; and the intercept $(1-F)(E+2)^{-1}$ of the function is positive. A value of F less than 1 is also compatible with Foley's estimate that the psychophysical function has a small positive intercept A, such that $0 < A < 2°$.

Computation of a roughly linear internal representation of the spherical coordinate R_H can also be achieved by shunting competition, as shown in fig. 12. Note that the vergence $V = r_1 - l_1$ is always positive, since the horizontal angle of the left eye is always greater than the horizontal angle of the right eye when a target is foveated. Also, $r_1 - l_1$ reaches a maximum when commands to both medial recti are maximal; that is, when the eyes are maximally converged and R_H

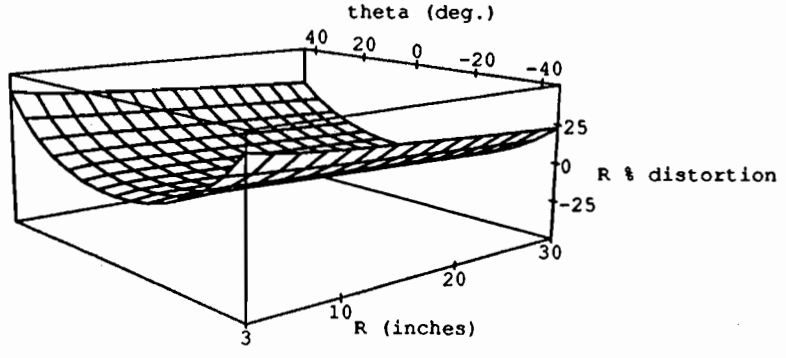

Fig. 12. Percent distortion for R_H throughout the workspace. See text for details on distortion measure.

is minimal. These properties suggest that, if $r_1 - l_1$ is normalized by a shunting competitive interaction, then $r_1 - l_1$ may be transformed into an estimate of R_H using shunting inhibition rather than excitation. The simplest shunting competitive interaction of this type has equilibrium values

$$h_5 = \frac{r_1 - l_1}{G + r_1 - l_1}, \tag{35}$$

and

$$h_6 = \frac{G}{G + r_1 - l_1}, \tag{36}$$

where G is a level of tonic excitation that is inhibited by the vergence signal. Because h_6 is inversely related to vergence, h_6 can estimate the distance R_H. In fact, h_6 is a more linear estimate of R_H than the approximation $R_H \alpha 1/\gamma$ for a wide range of E values. The relevance of such a transformation was shown in fig. 3c.

Fig. 12 shows the distortion measure for R_H throughout the workspace with $G = 0.001$. The internal representation of R_H is independent of ϕ_H. As expected from our discussion of fig. 3b, the distortion in this case is much larger than that of θ_H and ϕ_H, ranging up to nearly 35% and taking the form of an increasing overestimate of the distances of targets with larger horizontal angles $|\theta_H|$. A subsequent article will show how better estimates of R_H, for purposes of reaching a target with an arm, can be discovered through an on-line learning process.

The many uses of a head-centered representation of 3-D target position

The main result of the present article is that opponent interactions can transform naturally occurring eye movement outflow commands into a representation of 3-D position whose properties are supported by psychophysical evidence. Remarkably, the polar angle (θ_H) and vergence (γ) variables arise as sums and differences of suitably preprocessed eye movement commands.

Such a target representation, by itself, can only be used in situations where the eyes are already foveating the target. Subsequent articles of this series show however, how the brain can utilize such an opponent representation for carrying out a variety of skilled sensory-motor actions. For example, in Grossberg et al. (1992), it is shown how a suitable defined learning process can combined the component 3-D motor representation with binocular visual information to form an invariant head-centered representation of both foveated and non-foveated 3-D target positions. In Guenther et al. (1992), it is shown how an invariant *body*-centered representation of both foveated and non-foveated 3-D target positions can be learned. In Bullock et al. (1992b), a solution of the classical motor equivalence problem is suggested, whereby many different joint configurations of a redundant manipulator can all be used to realize a desired trajectory in 3-D space. In particular, using the opponent model as a foundation, this control system learns a mapping from motion directions in 3-D space to velocity commands in joint space. Computer simulations of the model have demonstrated that, without any additional learning, the network can generate accurate movement commands that compensate for variable tool lengths, clamping of joints, distortions of visual input by a prism, and unexpected limb perturbations. Blind reaches have also been simulated. Thus the opponent model described herein, albeit simple, seems to form a key module in several neural systems that are capable of controlling complex sensory-motor skills.

References

Blank, A.A., 1978. 'Metric geometry in human binocular perception: Theory and fact'. In: E.L.J. Leeuwenberg and H.F.J.M. Buffart (eds.), Formal theories of visual perception. New York: Wiley.

Bullock, D., D. Greve, S. Grossberg and F. Guenther, 1992a. 'A head-centered representation of 3-D target position commands'. In: Proceedings of the 1992 International Joint Conference on Neural Networks, Vol. I (pp. 79–85). Piscataway, NJ: IEEE..

Bullock, D., S. Grossberg and F. Guenther, 1992b. A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm. Journal of Cognitive Neuroscience.

DeValois, R.L. and K.K. DeValois, 1975. 'Neural coding of color'. In: E.C. Carterette and M.P. Friedman (eds.), Handbook of perception, Vol. 5: Seeing. New York: Academic Press.

Foley, J.M., 1980. Binocular distance perception. Psychological Review 87, 411–434.

Grossberg, S., 1982. Studies of mind and brain. Dordrecht: D. Reidel.

Grossberg, S., F. Guenther, D. Bullock and D. Greve, 1993. Neural representations for sensory-motor control, II: Learning a head-centered visuomotor representation of 3-D target positions. Neural Networks (in press).

Grossberg, S. and M. Kuperstein, 1986. Neural dynamics of adaptive sensory-motor control: Ballistic eye movements. Amsterdam: North-Holland.

Grossberg, S. and M. Kuperstein, 1989. Neural dynamics of adaptive sensory-motor control. New York: Pergamon Press.

Guenther, F., D. Bullock, D. Greve and S. Grossberg, 1992. Neural representations for sensory-motor control, III: Learning a body-centered visuomotor representation of 3-D target positions. Submitted for publication.

von Helmholtz, H.L.F., 1962. Treatise on physiological optics. J.P.C. Southall (Trans.). New York: Dover.

Hodgkin, A.L., 1964. The conduction of the nervous impulse. Liverpool University, Liverpool.

Hollerbach, J.M., S.P. Moore and C.G. Atkeson, 1986. 'Workspace effect in arm movement kinematics derived by joint interpolation'. In: G. Gantchev, B. Dimitrov and P. Gatev (eds.), Motor control. New York: Plenum Press.

Katz, B., 1966. Nerve, muscle, and synapse. New York: McGraw Hill.

Mollon, J.D. and L.T. Sharpe (eds.), 1983. Colour vision. New York: Academic Press.

Morasso, P., 1981. Spatial control of arm movements. Experimental Brain Research 42, 223–227.

Robinson, D.A., 1970. Oculomotor unit behavior in the monkey. Journal of Neurophysiology 35, 393–404.

Robinson, D.A., 1973. Models of the saccadic eye movement control system. Kybernetik 14, 71–83.

Rock, I., 1984. Perception. New York: W.H. Freeman.

Ron, S. and D.A. Robinson, 1973. Eye movements evoked by cerebellar stimulation in the alert monkey. Journal of Neurophysiology 36, 1004–1021.

Schiller, P.H., 1970. The discharge characteristics of single units in the oculomotor and abducens nuclei of the unasesthetized monkey. Experimental Brain Research 10, 347–362.

Soechting, J.F. and M. Flanders, 1989. Errors in pointing are due to approximations in sensorimotor transformation. Journal of Neurophysiology 62, 595–608.

Steinbach, M.J. and D.R. Smith, 1981. Spatial localization after strabismus surgery: Evidence for inflow. Science 213, 1407–1408.

Villis, T., R. Snow and J. Hore, 1983. Cerebellar saccadic dysmetria is not equal in the two eyes. Experimental Brain Research 51, 343–350.