1992-02

Neural Representations for Sensory-Motor Control, II: Learning a Head-Centered Visuomotor Representation of 3-D Target Position

NEURAL REPRESENTATIONS FOR SENSORY-MOTOR CONTROL, II: LEARNING A HEAD-CENTERED VISUOMOTOR REPRESENTATION OF 3-D TARGET POSITION

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

February, 1992

Technical Report CAS/CNS–92–030

Boston University Center for Adaptive Systems and
Department of Cognitive and Neural Systems
111 Cummington Street
Boston, MA 02215

NEURAL REPRESENTATIONS FOR SENSORY-MOTOR CONTROL, II: LEARNING A HEAD-CENTERED VISUOMOTOR REPRESENTATION OF 3-D TARGET POSITION

by

Stephen Grossberg†, Frank Guenther‡, Daniel Bullock†, and Douglas Greve‡
Center for Adaptive Systems
and
Department of Cognitive & Neural Systems
Boston University
111 Cummington Street, Room 244
Boston, MA 02215

February 1992

Requests for reprints should be sent to:
Stephen Grossberg
Center for Adaptive Systems
Boston University
111 Cummington Street
Boston, MA 02215

† Supported in part by the National Science Foundation (NSF IRI-87-16960 and NSF IRI-90-24877) and the Office of Naval Research (ONR N00014-92-J-1309).

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

Acknowledgements: The authors wish to thank Kelly A. Dumont and Carol Y. Jefferson for their valuable assistance in the preparation of the manuscript.

Abstract

A neural network model is described for how an invariant head-centered representation of 3-D target position can be autonomously learned by the brain in real time. Once learned, such a target representation may be used to control both eye and limb movements. The target representation is derived from the positions of both eyes in the head, and the locations which the target activates on the retinas of both eyes. A Vector Associative Map, or VAM, learns the many-to-one transformation from multiple combinations of eye-and-retinal position to invariant 3-D target position. Eye position is derived from outflow movement signals to the eye muscles. Two successive stages of opponent processing convert these corollary discharges into a head-centered representation that closely approximates the azimuth, elevation, and vergence of the eyes' gaze position with respect to a cyclopean origin located between the eyes. VAM learning combines this cyclopean representation of present gaze position with binocular retinal information about target position into an invariant representation of 3-D target position with respect to the head. VAM learning can use a teaching vector that is externally derived from the positions of the eyes when they foveate the target. A VAM can also autonomously discover and learn the invariant representation, without an explicit teacher, by generating internal error signals from environmental fluctuations in which these invariant properties are implicit. VAM error signals are computed by Difference Vectors, or DVs, that are zeroed by the VAM learning process. VAMs may be organized into VAM Cascades for learning and performing both sensory-to-spatial maps and spatial-to-motor maps. These multiple uses clarify why DV-type properties are computed by cells in the parietal, frontal, and motor cortices of many mammals. VAMs are modulated by gating signals that express different aspects of the will-to-act. These signals transform a single invariant representation into movements of different speed (GO signal) and size (GRO signal), and thereby enable VAM controllers to match a planned action sequence to variable environmental conditions.

TABLE OF CONTENTS

1. Spatial Representations for the Neural Control of Flexible Movements

The present article introduces a neural network model of how the brain learns spatial representations with which to control sensory-guided and memory-guided eye and limb movements. These spatial 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. The present article describes a model for learning an invariant head-centered representation of 3-D target position. A model for learning an invariant body-centered representation of 3-D target position will be described elsewhere (Guenther, Bullock, Greve, and Grossberg, 1992).

The flexible spatial relationships of the eyes, head, body, and limbs with respect to one another enable humans and other mammals to carry out a remarkable range of skilled behaviors. Understanding how flexible control of multi-link movement systems is achieved during autonomous behavior in real time is one of the most challenging problems in the field of computational neuroscience. Because eye, head, body, and limb segments are not rigidly attached to each other, an object with a fixed location relative to one segment can vary widely in its location relative to other segments. In particular, the sensory systems, such as eyes and ears, typically ride on body segments different than those used to approach or reach for an object in space. The present article analyses the formation and structure 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.

Two examples may be cited to dramatize the central issues. A human can feel an insect crawling up his or her leg while standing or sitting, and can reach accurately without vision to brush away the insect. The leg skin is a sensory surface that assumes different positions relative to the shoulder joint when we move from a sitting to a standing posture. Because the shoulder joint is the origin for the reaching limb, different arm-joint angles are required to reach to the same insect location on the thigh while sitting than while standing. This

defines a cutaneo-motor coordination problem.

Similarly, the eyes are segments containing sensory surfaces that move relative to the head, and the head is a segment that moves relative to the body. As the eyes move in the head and the head moves in a stationary body, the visual representation of a stationary object on the retinas keeps changing, yet the location of object with respect to the body remains fixed. Likewise, if the eyes fixate an object while the body stance is altered, the visual representation of the object may remain unchanged, yet the location of the object with respect to the body changes. Here, different arm-joint angles will be needed to reach an object that is located identically relative to the sensory surfaces by which the object is detected. This defines a visuo-motor coordination problem.

In both of these examples, the information available at the sensory surfaces, whether skin or retina, is insufficient to control accurate sensory-motor coordination across the interposed segments. Additional information is needed to resolve the ambiguity inherent in the one-to-many map between position of a sensory surface and position of a moving limb.

Gibson (1966) has noted that some types of information are inherently superior to others. Information that is naturally generated within the perception-action cycle, and that is capable of acting directly to guide action, is inherently more useful in real-time control than information in the form of "symbolic rules", "assumptions", or "memory images", all of which can be applied to an ongoing sensory-motor control task only by indirect means. Such indirection often requires more processing steps and therefore more processing time, as well as access to types of information that are not available to an animal behaving under uncertain environmental conditions in real time. Schemes that use externally controlled switching between learning and performance episodes, or control event durations to prevent learning instabilities, are also insufficient to model the behavior of freely moving animals or autonomous robots. The neural networks proposed herein rely only on information that is available during an ongoing perception-action cycle. We show how information of several different types may be rapidly combined by an appropriately defined unsupervised learning

system whose properties help to clarify a variety of psychophysical and neurobiological data about movement control.

Three general design themes underly many of our results. One theme explores the need for spatial representations—as distinct from perceptual, cognitive, or motor representations—in the control of goal-oriented behaviors. In this regard, it is well-known that visual inputs activate a "what" processing stream as well as a "where" processing stream within the brain (Goodale and Milner, 1992). The "what" processing stream leads to recognition of external objects, and includes brain regions such as visual cortex and inferotemporal cortex. The "where" processing stream leads to spatial localization of objects, and includes brain regions such as superior colliculus and parietal cortex. "Where" processing is illustrated by the following competence. Imagine that your right hand is moved by an external force to a new position in the dark. Thus neither visual cues nor self-controlled outflow movement commands are available to encode the right hand's new position. Despite the absence of vision and self-controlled volition, it is easy to move your left hand to touch your right hand in its new location. The motor coordinates which represent the position of your right hand are different from the motor coordinates that your left arm realizes in order to touch it. Some representation needs to exist that mediates between the different motor coordinates of the two arms. This mediating scheme is the spatial representation.

This example illustrates that different motor plans, whether for the control of one arm or two, are often used to reach a prescribed position in space. The problem of how animals can reach a fixed target in multiple ways is often called the "problem of motor equivalence". A properly defined spatial representation is a prerequisite to discovering a biologically relevant solution of the motor equivalence problem. The model introduced herein forms part of a proposed solution to the motor equivalence problem (Bullock, Grossberg, and Guenther, 1992). In this regard, our research program has sought to characterize spatial representations that can be embedded in a larger neural system capable of autonomously learning to perform skilled arm movement sequences, such as handwriting and visually-guided object

manipulation, at any reachable positions and size scales with respect to the body. Such a spatial representation should enable planned action sequences to be performed with a tool of variable length and mass, such as a pen or hook, either in response to visual guidance or from memory. We also require that the ability to perform an action starting with a different initial position, size scale, or tool can be achieved without having to learn each of these variations as a different motor plan. Rather, these different trajectories should emerge as natural invariants of the interaction between spatial and motor representations, modulated by state-dependent parameter changes such as "acts of will", and by appropriate sensory feedback. Thus we seek to define an action-oriented spatial representation that has evolved for the control of skilled motor behavior.

The spatial representations 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 second 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. Such an opponent model of 3-D target position was introduced in Bullock, Greve, Grossberg, and Guenther (1992) and developed in Greve, Grossberg, Guenther, and Bullock (1992). It describes a head-centered spatial representation of 3-D targets that are foveated by both eyes. This model is used herein as part of the present model, which learns how to combine visual and motor information to generate an invariant head-centered spatial representation for both foveated and non-foveated 3-D target positions. A head-centered spatial representation of *non*-foveated targets is needed both to look at new targets with the eyes and to reach towards these targets with the limbs.

What type of learning is appropriate to generate such a spatial representation? An answer to this question is described below as part of the third design theme of our work, which

asks, more generally, how to define action-oriented spatial representations. In particular, what type of learning gives rise to spatial representations that are computationally consistent with the motor trajectory generators that they control? Such consistency cannot be taken for granted in a self-organizing system whose behavioral properties emerge from distributed interactions among many system components. Remarkably, spatial representations and trajectory generators seem to use the same type of circuit module, and thus the same type of learning law. The fact that networks for representing space can use the same type of neural circuit, called a Vector Associative Map, as networks for the control of variable-speed synchronous control of a multi-joint limb was first demonstrated by Gaudiano and Grossberg (1991). In this work, it was shown how a 1-dimensional space could self-organize and learn to control synchronous variable-speed trajectories of a 2-joint arm. The present article begins to show how a 3-dimensional space can self-organize and learn to control synchronous variable-speed and variable-size trajectories of a 4-joint arm, with or without a tool of variable length (see Bullock, Grossberg, and Guenther, 1992).

The next section surveys key geometrical and psychophysical considerations pertinent to the model. For completeness, Sections 3 and 4 describe how two successive stages of opponent interactions can generate the type of head-centered representation that is suggested by psychophysical and neurobiological data. Sections 5–14 describe relevant properties of Vector Associative Maps. Section 15 begins specification of a neural network model for learning invariant head-centered visuomotor target positions. Six versions of this model will be described to highlight invariant model properties while also acknowledging the existence of variations on a theme.

2. Geometry of Object Localization

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 position of an object

with respect to the head. Such visual information is, however, often insufficient for accurate reaching towards a binocularly fixed target. One reason for this limitation is that binocular disparity, by itself, does not provide unambiguous information about absolute distance. For example, if each eye fixates the interior of a homogeneous object at a different location, then the two monocular images of the object's interior can be binocularly fused. However, the binocular disparities of the object's boundaries will change with every change in the fixation points of the two eyes. These binocular disparity changes occur without a change in the object's distance from the observer. Thus binocular disparity is not a reliable cue to absolute distance in any situation of this type.

Another limitation of binocular disparity cues arises whenever the object is a target that both eyes binocularly fixate. 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, small fixated objects cannot accurately be reached using only information about 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.

Figure 1

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 (Blank, 1978; Foley, 1980; Hollerbach, Moore, and Atkeson, 1986; Soechting and Flanders, 1989). This evidence is reviewed in Greve, Grossberg, Guenther, and Bullock (1992). A self-contained formal description of such a neurally generated 3-D coordinate system is described herein.

Figure 2

Figure 1a shows how the intersection point of the lines of sight of the two eyes converge 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. 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, which is called the *binocular parallax* (Foley, 1980). This triangular structure also allows an internal measure of net ocular *vergence*—the extent to which the eyes are rotated towards the nose—to serve as one coordinate for estimating the distance from egocenter to a binocularly foveated object. The angle γ will henceforth be used as a measure of vergence. The two other coordinates in this 3-D representation are also derived from estimates of the position of both eyes in their orbits. Figure 1b shows the relation between γ and the radial distance of a target from the radial egocenter that is defined in Figure 2. Figure 2 describes the geometry of 3-D target localization in terms of spherical coordinates. The origin of this coordinate system, called the cranial egocenter, lies at the midpoint between the two eyes. Thus the representation is "cyclopean". The head-centered horizontal angle or azimuth, θ_H, and the vertical angle or elevation, ϕ_H, measure deviations from straight-ahead gaze. The radial distance R_H is replaced by the vergence, as in Figure 1b. Figure 3 describes the geometry of the cyclopean angle θ_H with respect to the angles θ_L and θ_R subtended by the left eye and right eye, respectively.

Figure 3

3. Opponent Interactions for Representation of Foveated 3-D Target Positions

We now summarize how to binocularly combine outflow signals from the tonically active cells that control the position of each eye (Figure 4) to form a head-centered representation of a foveated target. This can be done in two stages of opponent processing. First, opponent interactions combine the outputs of the cells that control the agonist and antagonist muscles of each eye (Figure 5). These opponent interactions give rise to opponent pairs of cells the sum of whose activity is approximately constant, or normalized. Next, the normalized outputs from both eyes are combined in two different ways to generate a head-centered spatial representation of the binocular fixation point. 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 Figure 5, one combination gives rise to a cell population whose activity h_2 approximates the angular spherical coordinate θ_H. The other combination gives rise to a cell population whose activity Γ approximates 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.

Figure 4

The neural mechanism for normalizing the total activity of opponent cells uses a shunting on-center off-surround network (Grossberg, 1982); 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. Let the normalized opponent cells in the shunting network have activities l_1 and l_2. Suppose that

$$\frac{d}{dt}l_1 = -Al_1 + (1 - l_1)L_1 - l_1L_2 \tag{1}$$

and

$$\frac{d}{dt}l_2 = -Al_2 + (1 - l_2)L_2 - l_2L_1. \tag{2}$$

By equation (1), activity L_1 excites l_1 whereas activity L_2 inhibits l_1. The opposite is true in equation (2). Parameter A is the decay rate. At equilibrium, $\frac{d}{dt}l_1 = \frac{d}{dt}l_2 = 0$, so (1) and (2) imply that

$$l_1 = \frac{L_1}{A + L_1 + L_2} \tag{3}$$

and

$$l_2 = \frac{L_2}{A + L_1 + L_2}. \tag{4}$$

Adding (3) and (4) shows that

$$l_1 + l_2 = \frac{L_1 + L_2}{A + L_1 + L_2}. \tag{5}$$

Thus if $A \ll L_1 + L_2$,

$$l_1 + l_2 \cong 1. \tag{6}$$

The approximation (6) 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. Figure 5 shows the two cellular pairs needed to control θ_L and θ_R. These pairs are labeled 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:

$$l_1 + l_2 = 1 \tag{7}$$

$$\theta_L = -90° + 180° \times l_2 \tag{8}$$

$$r_1 + r_2 = 1 \tag{9}$$

$$\theta_R = -90° + 180° \times r_2 \tag{10}$$

where l_i indicates the activity of left eye cell population i and r_i indicates the activity of right eye cell population i.

Figure 5

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

$$l_3 + l_4 = 1 \tag{11}$$

$$\phi_L = -90° + 180° \times l_4 \tag{12}$$

$$r_3 + r_4 = 1 \tag{13}$$

$$\phi_R = -90° + 180° \times r_4. \tag{14}$$

To provide a head-centered representation of foveated 3-D target positions, the outflow signals l_1, l_2, l_3, and l_4 are binocularly combined. Let the cell populations $h_i, i = 1, 2, ..., 6$, form the basis for this head-centered spatial representation. These populations are also arranged in antagonistic pairs. First we define cell activities h_1, h_2, h_3, and h_4 that linearly approximate the following estimates of θ_H and ϕ_H:

$$h_1 + h_2 = 1 \tag{15}$$

$$\theta_H = -90° + 180° \times h_2 \tag{16}$$

$$h_3 + h_4 = 1 \tag{17}$$

$$\phi_H = -90° + 180° \times h_4. \tag{18}$$

These 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. Figure 5 shows the connectivity of a network for the cell activity h_2 which represents θ_H. In particular,

$$\frac{d}{dt}h_2 = -Bh_2 + (1 - h_2)(l_2 + r_2) - h_2(l_1 + r_1), \tag{19}$$

where B is the decay rate. Solving this equation at equilibrium ($dh_2/dt = 0$) yields

$$h_2 = \frac{l_2 + r_2}{B + l_1 + r_1 + l_2 + r_2}. \tag{20}$$

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

$$h_2 \cong \frac{l_2 + r_2}{2}. \tag{21}$$

Likewise,

$$h_1 \cong \frac{l_1 + r_1}{2} \tag{22}$$

so that, by (21) and (22),

$$h_1 + h_2 \cong 1. \tag{23}$$

To evaluate the adequacy of this internal representation of θ_H, a distortion measure was calculated in Greve, Grossberg, Guenther, and Bullock (1992) 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. 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. The distortion in this range is less than 15%, with essentially 0% distortion for $R_H > 5$ inches. Thus, the opponent network defined above provides an accurate mechanism for computing an internal representation of θ_H. Likewise, the distortion measure for ϕ_H showed that the normalized binocular opponent network provides an accurate internal representation of ϕ_H in all but the most extreme portions of the workspace.

To review how opponent computation leads to a representation of vergence, note that vergence is equal to the difference between r_1 (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 Figure 5, define a cell population with activity Γ (for internal representation of vergence γ) 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

$$\frac{d\Gamma}{dt} = -C\Gamma + (1-\Gamma)(r_1 + l_2) - (\Gamma + D)(l_1 + r_2). \tag{24}$$

At equilibrium,

$$\Gamma = \frac{r_1 - l_2 - Dl_1 - Dr_2}{C + r_1 + r_2 + l_1 + l_2}. \tag{25}$$

Because $r_1 + r_2 = 1$ and $l_1 + l_2 = 1$, equation (25) can be rewritten as

$$\Gamma = \frac{1-D}{C+2} + \frac{1+D}{C+2}(r_1 - l_1). \tag{26}$$

If $D = 1$ and $C = 0$, then

$$\Gamma = r_1 - l_1. \tag{27}$$

In this case, subjective parallax equaled physical parallax. If, however, $C > 0$ and $D < 1$, then the slope $(1 + D)(C + 2)^{-1}$ of Γ versus $r_1 - l_1$ is less than one, and the intercept $(1-D)(C+2)^{-1}$ of the function is positive. Such values are compatible with the Foley (1980) estimate from psychophysical data of the internal representation of Γ. See Greve *et al* (1992) for further discussion of psychophysical data that are consistent with this representation.

4. Converting Motor Representations of Foveated Target Positions into Visuomotor Representations of Non-Foveated Target Positions

This section summarizes computational issues that help to motivate the model. The central question is: How can a motor representation of foveated target positions be used to learn a visuomotor representation of both foveated and non-foveated target positions? In

order to answer this question, the following ingredients are needed: a motor representation of where the two eyes are looking; a retinal visual representation of a nonfoveated target in 3-D space; a head-centered representation of target position in 3-D space; and a learning law that can combine the first two types of information so that they can jointly predict the third.

The next section discusses the learning module. After that, an analysis of how the three types of information are computed and combined during real-time learning conditions will be considered. Of particular importance is the issue of how an invariant head-centered representation of 3-D space can be self-organized even though no part of the system is endowed with such a head-centered representation before learning occurs. The core problem is that *many* combinations of eye position and retinal target position correspond to *one* head-centered target position. What sort of teaching signal can sort out this many-to-one relationship to discover the correct head-centered invariant representation?

5. Vector Associative Maps: A Unified Format for Learning Spatial and Motor Representations

The same type of module, used at different processing stages, is capable of learning parameters for the trajectory controllers of multi-joint limb movements, and the spatial representations that activate the trajectory controllers. Thus, replication of a common design at different stages of brain processing can learn both spatial and motor transformations. The existence of such a module, called a Vector Associative Map, or VAM (Gaudiano and Grossberg, 1991, 1992), clarifies how spatial representations can interact in a computationally consistent way with motor trajectory controllers. The main concepts needed to motivate our development of VAM systems are provided below.

VAM dynamics clarify how a child learns to reach for objects that it sees. This problem requires understanding the interactions between two distinct modalities: vision (seeing an object) and motor control (moving a limb). In particular, how does an individual stably learn

transformations within and between the two different modalities that are capable of controlling accurate goal-oriented movements? The behavioral events that enable such learning to occur were called a *circular reaction* by the Swiss psychologist Jean Piaget (1963).

The circular reaction is an autonomously controlled behavioral cycle with two components: *production* and *perception*, with learning linking the two modalities to enable sensory-guided action to occur. Such a circular reaction is *intermodal*; that is, it consists of the coupling of two systems operating in different modalities. In order for the intermodal circular reaction to generate stable learning of the parameters that couple the two systems, the control parameters within each system must already be capable of accurate performance. Otherwise, performance may not be consistent across trials and a stable mapping could not be learned between different modalities. Thus it is necessary to self-organize the correct *intramodal* control parameters before a stable *intermodal* mapping can be learned.

Grossberg and Kuperstein (1986, 1989) modeled how such intramodal control parameters can be learned within the eye movement system. During early development, eye movements are made reactively in response to visual inputs. When these eye movements do not lead to foveation of the visual target, the nonfoveated position of the target generates a visual error signal. The Grossberg-Kuperstein model suggests how such error signals can be used by the cerebellum to learn eye movement control parameters that lead to accurate foveations.

The VAM model clarifies how the arm movement system can endogenously generate movements during a "motor babbling" phase. "Motor babbling" describes the spontaneous arm movements of an infant during an early developmental phase. As explained below, these movements help to generate the data needed to learn correct arm movement control parameters. For example, they activate target position representations that are used to learn a visuomotor transformation that controls visually guided reaching. The simplest example of a VAM is a model called the AVITE (or Adaptive Vector Integration To Endpoint) model (Figure 6) for variable-speed adaptive control of multi-joint limb trajectories. The AVITE model is, in turn, a self-organizing version of the VITE model of Bullock and Grossberg

(1988a) for variable-speed control of multi-joint trajectories.

Figure 6

6. Trajectory Properties as Emergent Invariants

Bullock and Grossberg (1988a) suggested that arm movement trajectory properties emerge through interactions among two broad types of control mechanisms: planned control and automatic control. Planned control variables include target position, or where we want to move; and speed of movement, or how fast we want to move to the desired position, and the "will" to move at all. Automatic control variables compensate for the present position of the arm, unexpected inertial forces and external loads, and changes in the physiognomy of the motor plant, say due to growth, injury, exercise, and aging.

The VITE model of Bullock and Grossberg implements part of such a strategy of trajectory control and has been used to explain a large behavioral and neurobiological data base (see Bullock and Grossberg, 1988a, 1988b, 1989, 1991). The model clarifies how motor synergies can be dynamically bound and unbound in real-time, and how multiple joints within a synergy can be synchronously moved at variable speeds. The synchrony with which different muscles of a synergy contract by different amounts in equal time emerges from the interactive dynamics of the network, as do many other trajectory properties, such as empirically observed velocity profiles; they are not externally controlled or programmed into the network.

7. The VITE Model

Figure 7 summarizes the main components of the VITE circuit. At the top of the figure, inputs to the Target Position Command (TPC) populations, represent the desired final position of the arm. At the bottom of the figure, the Present Position Command (PPC) populations code an internal representation of where the arm actually is. Outflow movement commands to the arm are generated by the PPC. These outflow signals, supplemented

by spinal circuitry and cerebellar learning (Bullock and Contreras-Vidal, 1991; Bullock, Contreras-Vidal, and Grossberg, 1992; Bullock and Grossberg, 1989, 1991) move the hand to the location relative to the body that is coded by the PPC.

Signals from the TPC and the PPC enable the Difference Vector (DV) populations to continuously compute the discrepancy between present position (PPC) and desired position (TPC). DV activation is integrated by the PPC until the latter becomes equal to the TPC, at which time the DV will be equal to zero and PPC integration stops. Hence the VITE circuit embodies an automatic process that moves the PPC continuously to the TPC. The Adaptive VITE (AVITE) model summarized herein explains how "motor babbling" endogenously generates PPC representations that move the arm through a full range of positions, and activate TPCs whose signals to the DV are adaptively tuned to be dimensionally consistent with the corresponding PPCs, by using the DVs as source of error signals during learning.

Figure 7

8. Coding Movement Speed and Intentionality: The GO Signal

If the PPC were always allowed to integrate the DV, then a movement would begin as soon as the TPC becomes active. Somehow it must be possible to "prime" a target position without moving the arm until another signal indicates the intent to carry out the movement. A related issue concerns how the overall speed of a movement can be varied without changing the desired TPC. "Priming" denotes the limiting case of zero speed.

Trajectory-preserving speed control can be achieved by multiplying the output of the DV with a nonspecific gating signal. This is the GO signal depicted in Figure 7. Because of its location within the VITE model, the GO signal affects the rate at which the PPC is continuously moved toward the TPC, without altering the resulting trajectory.

For example, as long as the GO signal is zero, instatement of a TPC generates a non-zero DV, but the PPC remains unaltered. This "primed" DV codes the difference between

the arm's present position and desired position. When the GO signal is nonzero, the DV is integrated by the PPC at a rate proportional to the product $(DV) \cdot (GO)$. Integration ceases when the PPC equals the TPC and the DV equals zero, even if the GO signal remains positive. Other things being equal, a larger GO signal causes the PPC to integrate at a faster rate, so the same target is reached in a shorter time.

The synchrony of synergetic movement control by a VITE circuit is preserved in response to an arbitrary GO signal, and the main qualitative properties of VITE-controlled velocity profiles are preserved in response to a wide class of increasing GO signals (Bullock and Grossberg, 1988a). The model's prediction of a reversal in the direction of velocity profile asymmetry with increasing speed was confirmed in an explicit test by Nagasaki (1989), and its prediction of a late-acting execution-gating signal was confirmed in an explicit test by DeJong, Coles, Logan, and Gratton (1990).

9. Autonomous Learning of VITE Coordinates

In order for the VITE model to generate correct arm trajectories, the TPC and PPC must be able to activate dimensionally consistent signals $TPC \rightarrow DV$ and $PPC \rightarrow DV$ for comparison at the DV. There is no reason to assume that the gains, or even the coordinates, of these signals are initially correctly matched. Learning of an adaptive coordinate transformation is needed to achieve self-consistent matching of TPC- and PPC-generated signals at the DV.

In order to learn such a transformation, TPCs and PPCs that represent the same target positions must simultaneously be activated. This cannot be accomplished by activating a TPC and then letting the VITE circuit generate a corresponding PPC. Such a scheme would beg the problem being posed; namely, to discover how excitatory $TPC \rightarrow DV$ and inhibitory $PPC \rightarrow DV$ signals are so calibrated that DV stage outputs *can* generate the corresponding PPC. An analysis of all the possibilities that are consistent with VITE constraints suggests that PPCs may initially be generated by an internal, or endogenous, activation source during

a motor babbling phase. This source is called the Endogenous Random Generator, or ERG (Figure 8). After such a babbled PPC is generated and a corresponding action taken, the PPC is itself used to directly instate a TPC that represents the same target position. This occurs via a one-to-one mapping along pathway $PPC \rightarrow NP \rightarrow TPC$ in Figures 6b and 7 (NP = Now Print gate, described below). Thus motor babbling samples the work space and, in so doing, generates a representative set of pairs (TPC, PPC) for learning the VITE coordinate transformation. Such learning enables endogenously generated movements to be supplanted by planned movements.

10. Associative Learning from Parietal Cortex to Motor Cortex during Motor Babbling

Further analysis suggests that the site where an adaptive coordinate change can take place is at the synaptic junctions that connect the TPC to the DV. These junctions are represented as semi-circular synapses in Figure 6. From this perspective, the DV represents an internal measure of error, in the sense that miscalibrated signals $TPC \rightarrow DV$ and $PPC \rightarrow DV$ from TPCs and PPCs that correspond to the same target position will generate a nonzero DV. Learning is designed to change the synaptic weights in the pathways $TPC \rightarrow DV$ in a way that drives the DV to zero. After learning is complete, the DV can only equal zero if the TPC and PPC represent the same target position. If we accept the neural interpretation of the TPC as being computed in the parietal cortex (Anderson, Essick, and Siegel, 1985; Grossberg and Kuperstein, 1986, 1989) and the DV as being computed in the motor cortex (Bullock and Grossberg, 1988a; Georgopoulos *et al.*, 1982, 1984, 1986), then this model predicts that associative learning from parietal cortex to motor cortex takes place during motor babbling, and attenuates activation of the difference vector cells in the motor cortex during postural intervals.

Figure 8

11. Vector Associative Map: On-Line DV-Mediated Learning and Performance

When such a learning law is embedded within a complete AVITE circuit, the DV can be used for on-line regulation of both learning and performance. During a performance phase, a new TPC is read into the VITE circuit from elsewhere in the network, such as when a reaching movement is initiated by a visual representation of a target. The new DV is used to update the PPC to a new setting that represents the same target position as the TPC. As the PPC is updated, the DV is zeroed while the TPC is held constant. During the learning phase, the DV is used to drive a coordinate change in the $TPC \rightarrow DV$ synapses. Zeroing the DV here creates new adaptive weights while both the PPC and TPC are held fixed.

Both the learning and the performance phases use the same AVITE circuitry, notably the same DV, for their respective functions. Thus learning and performance can be carried out on-line in a real-time setting, unlike schemes like back propagation. The operation whereby an endogenously generated PPC activates a corresponding TPC, as in Figure 6, "back propagates" information for use in learning, but does so using local operations without the intervention of an external teacher or a break in on-line processing.

Autonomous control, or gating, of the learning and performance phases is needed to achieve effective on-line dynamics. For example, the network needs to distinguish whether $DV \neq 0$ because the TPC and PPC represent different target positions, or because the $TPC \rightarrow DV$ synapses are improperly calibrated. In the former case, learning should not occur; in the latter case, it should occur. Thus some type of learning gate is needed to prevent spurious associations from forming between TPCs and PPCs that represent different target positions. The design of the total AVITE network shows how such distinctions are computed and used for real-time control of the learning and performance phases. We now explain how this is accomplished.

12. The Motor Babbling Cycle

During the motor babbling stage, an Endogenous Random Generator (ERG) of random

vectors is activated. These vectors are input to the PPC stage, which integrates them, thereby giving rise to outflow signals that move the arm through the workspace (Figure 8a). After each interval of ERG activation and PPC integration, the ERG *automatically* shuts off, so that the arm stops at a particular target position in space.

Offset of the ERG opens a Now Print (NP) gate that copies the PPC into the TPC through some fixed transformation (Figure 8b). The only requirement is that the transformation be one-to-one. It could even be realized through external, notably visual, feedback. The top-down adaptive filter from TPC to DV learns the correct reverse transformation (Figure 8c) by driving the DV toward zero while the NP gate is open (Figure 8d).

Then the cycle repeats itself automatically. When the ERG becomes active again, it shuts off the NP gate and thus inhibits learning. A new PPC vector is integrated and another arm movement is elicited. The ERG is designed such that, across the set of all movement trials, its output vectors generate a set of PPCs that form an unbiased sample of the workspace. This sample of PPCs generates the set of (TPC, PPC) pairs that is used to learn the adaptive coordinate change $TPC \rightarrow DV$ via a vector associative map.

13. The Endogenous Random Generator of Workspace Sampling Bursts

The ERG design embodies an example of opponent interactions (Figure 8). The motor babbling cycle is controlled by two complementary phases in the ERG mechanism: an *active* (ERG ON) and a *quiet* (ERG OFF) phase. The active phase generates random vectors to the PPC. During the quiet phase, input to the PPC from the ERG is zero, thereby providing the opportunity to learn a stable (TPC, PPC) relationship. In addition, there must be a way for the ERG to signal onset of the quiet phase, so that the NP gate can open and copy the PPC into the TPC (Figure 8b). The NP gate must not be open at other times: If it were always open, any incoming commands to the TPC could be distorted by contradictory inputs from the PPC. Offset of the active ERG phase is accompanied by onset of a complementary mechanism whose output energizes opening of the NP gate. The signal that opens the NP

gate can also be used to modulate learning in the adaptive filter. No learning should occur except when the PPC and TPC encode the same position.

Further details concerning ERG design and autonomous learning of AVITE parameters are found in Gaudiano and Grossberg (1991). Gaudiano and Grossberg also reported the first example of how iterated VAM modules, forming a VAM Cascade, could be used to learn a simple head-centered spatial representation for control of a VITE motor trajectory generator (Figure 9). This head-centered representation used a single eye's position and retinal target location to learn a 1-dimensional spatial map. Such a representation is insufficient to control spatial orientation and reaching in 3-D space. For this purpose, positional and retinal information from both eyes needs to be suitably combined. How this can be achieved is the central theme of the present article.

Figure 9

14. Voluntary Rescaling of Movement Properties by Nonspecific GO, GRO, and CO Signals

Before describing details of a VAM for computing 3-D head-centered representations, we note an implication of the postulate that such vector representations exist. In particular, vector representations make it relatively easy to use nonspecific control signals to rescale parameters of movement and posture. For example, scalar multiplication of difference vectors can be used to rescale movement speed or amplitude while preserving movement direction. Within an AVITE model for motor trajectory control, the DV is multiplied by a GO signal before the $DV \cdot GO$ product is integrated by the PPC. To control movement speed without changing movement direction, the same scalar GO signal multiplies all components of the DV equally—that is, nonspecifically or without any component-specific bias.

Now consider a case where an AVITE TPC is being updated by a mapping from a DV computed in 3-D spatial coordinates. A multiplicative signal applied to such a DV may be called a GRO signal, because it rescales the amplitude of the movement specified by the

DV without changing its direction. Bullock and Grossberg (1991) have noted that such unbiased rescaling effects are quite difficult to achieve in alternative models that deviate from VITE-like designs.

Even using VITE-like controllers, however, specialized ancillary circuitry is needed to ensure that the nonlinear muscle plant will respond veridically to rescaled VITE commands. The FLETE model (Bullock and Grossberg, 1989, 1991; Bullock and Contreras-Vidal, 1991) clarifies how spinal circuitry works to ensure unbiased motor responses to nonspecific rescaling signals. In addition to explaining how spinal circuits assist speed rescaling, the FLETE model explains how a nonspecific signal sent to all PPC components can achieve equal co-contractions of opponent muscles. This co-contraction, or CO, signal controls joint stiffness to deal with variable force conditions without altering the planned motor trajectory.

These three signals—GO, GRO, CO—enable a stereotyped series of DV's to be transformed into motor performances with variable sizes, speeds, and tensions. In this way, VAM controllers can be used to tailor a planned action sequence to match variable environmental conditions without having to learn a different trajectory for every circumstance. The GO, GRO, and CO signals are under voluntary control. Indeed, they define different dimensions of volition. Their simple, nonspecific mode of action is transformed by the VAM architecture into subtle multi-dimensional movement changes. This interaction helps to clarify how the apparent simplicity of volition may lead to complex biomechanical consequences.

Neural sites pertinent to these three types of scaling signals have been partly identified. The GO signal shares properties with cells in the globus pallidus (Bullock and Grossberg, 1989, 1991; Horak and Anderson, 1984a, 1984b). The CO signal may be expressed in the spinal cord and generated in the precentral motor cortex (Bullock and Grossberg, 1991; Humphrey and Reed, 1983). It remains to determine where GRO signals are computed. Plausible sites include parietal cortex and basal ganglia. These correlations are summarized in Table 1.

Table 1

15. Variations on a Theme: Explicit Teachers for Learning an Invariant Representation

This article shows how six different, but related, ways of combining information about eye position, retinal target position, and head-centered target position can learn an invariant head-centered spatial representation using a VAM network. All six variations are described to provide a better insight into the map learning process, and because different variations may have advantages in different species and applications (Table 2). These models are illustrated in Figures 10-15. In each model, stages analogous to those in an AVITE exist. The analog of the TPC is a distributed representation of 3-D target position that is implicitly defined by converging signals from two types of representations: representations of the 3-D position at which the eyes are initially gazing, expressed in motor coordinates, and representations of a non-foveated 3-D target position, expressed in visual coordinates. The analog of the PPC is a distinct representation of 3-D target position, which acts as a teaching signal. These different representations of the same 3-D target position send signals to a DV stage, at which any discrepancy triggers DV-reducing learning within the adaptive weights corresponding to the visual representation.

Figure 10

Figures 10–12 summarize three models which exploit the fact that an explicit teaching signal exists during learning of a head-centered map. In Model 1 of Figure 10, the two eyes begin by foveating some position in 3-D space. Their respective locations in the head are jointly coded by the 3-D motor vector that represents foveated eye position, as described in Section 3. This representation is stored in short term memory, or STM, throughout the subsequent eye movement. It also sends signals along fixed weight pathways to the DV stage.

Figure 11

A non-foveated target position is represented by activation of two retinotopic spatial maps, one associated with each eye. During the subsequent eye movement, each map stores in STM the position that the target initially excited on the retina of its eye. In Model 1, it is assumed for simplicity that only horizontal eye positions are encoded with respect to the egocenter. A similar analysis can be carried out for vertical and oblique egocentric locations. Each retina is mapped into a coarse-coded one-dimensional horizontal array. Model 1 assumes that, at the DV stages, each retinotopic array adds its own monocular adaptive signals to the non-adaptive signals from the eye position vector in order to learn a head-centered visuomotor representation. In effect, monocular visual signals from two retinotopic maps are adaptively combined through learning into an effective binocular control signal. The pairs of monocular retinotopic signals need to correspond to the same 3-D target position in order for effective learning to occur. It is assumed that such a selection is made by a feedback interaction with a binocular visual representation of the target's position that is computed elsewhere in the network.

The teaching signal in Models 1–3 takes advantage of the fact that the saccadic eye movement system can learn to make accurate visually reactive movements. As noted in Section 5, Grossberg and Kuperstein (1986, 1989) have shown how visual error signals can be used by the cerebellum to learn eye movement parameters that lead to accurate foveation. After such a correct movement takes place, the new positions of both eyes provide a head-centered representation of the desired target position. We assume that this representation is instated at the PPC stage of the spatial VAM, from which it propagates to the DV stage as a teaching signal after the eye movement is complete. This representation is also encoded using the head-centered opponent motor map of eye position that was described in Section 3.

After an accurate eye movement takes place, three types of information are simultaneously available: a motor representation of both eyes' positions before the movement; a retinal representation of the target position on both retinas before the movement; and a mo-

tor representation of both eyes' positions after the movement. A VAM module enables the first two types of information to learn to predict the third. After this happens, all combinations of initial eye position and retinal position that predict the same final eye position will read-out the same representation of this position at the VAM DV Stage. Note that without the retinotopic input, the DV stage measures the difference between the initial and final eye positions needed to foveate a 3-D target in terms of a fixed motor metric. VAM learning calibrates retinotopic inputs to be consistent with this motor metric. Once calibrated, these retinotopic inputs combine with cyclopean eye position inputs to compute head-centered target positions that are invariant under eye rotations and the retinal translations of target images that they induce. VAM learning hereby converts a non-invariant representation of final eye position into an invariant representation of head-centered target position.

Table 2

Model 2 uses the same teaching signal as Model 1. Instead of using pairs of monocularly activated retinas, Model 2 assumes that binocular vision has converted these monocular activations into a binocular retinotopic representation of target position, as in Figure 11 and Table 2. Such a binocular representation encodes the fused binocular position and the binocular disparity of the target, among other parameters. If only horizontal positions are considered, then horizontal position and binocular disparity may be combined into a coarse-coded two-dimensional spatial map. The fused binocular position is computed as the average of the individual left eye position and right eye position of the target. The binocular disparity is computed as the difference of the monocular target positions. The fused binocular position approximates the property of *displacement*, or *allelotropia* (Kaufman, 1974; von Tschermak-Seyseneg, 1952; Werner, 1937). In this phenomenon, when a pattern of letters **AB C** is viewed through one eye and a pattern **A BC** is viewed through the other eye, the letter **B** can be seen in depth at a position halfway between **A** and **C**. Thus the fused binocular position of **B** averages the left eye and right eye monocular positions of **B**. An explanation of how

allelotropia occurs is given in Grossberg (1992). When the two eyes foveate a target, these visually derived binocular position and disparity perform essentially the same averaging and difference computations as the head-centered estimates of cyclopean azimuth and vergence that are derived from motor outflow commands to the eye muscles. It is of considerable interest that the motor computations of cyclopean eye position and visual computations of binocular target position both estimate the same types of quantities in Models 2 and 5. In Model 3, a simpler two-dimensional binocular spatial map is used for comparison (Figure 12); namely, the $(i,j)^{th}$ map position codes the i^{th} and j^{th} positions in the left and right eye, respectively. In both Model 2 and Model 3, the binocular representation of target position and the binocular representation of initial eye position are stored before the eye movement occurs. After the eye movement is over, the VAM learns to combine these binocular representations into a many-to-one invariant representation of 3-D target position.

Figure 12

16. Variations on a Theme: Autonomous Discovery of an Invariant Representation

Models 4–6 illustrate a remarkable property of VAM learning. A VAM can discover an invariant many-to-one representation of 3-D target position even if an explicit teacher is not used, or does not exist. VAM learning can feed upon DV error signals that are generated by the statistics of the environment in order to discover invariant mapping properties that are implicit in these fluctuations.

Figure 13

Model 4 uses the monocular retinotopic representations of Model 1 (Figure 13). Model 5 uses the binocular representation of Model 2 (Figure 14). Model 6 uses the binocular representation of Model 3 (Figure 15). Models 4–6 each assume that the initial eye position signals and retinotopic signals are combined before an eye movement takes place and that

the combination is stored at the PPC stage throughout the eye movement. This stored vector provides an estimate of target position which may or may not be correct. In order to store this estimate, the model exploits the existence of a *gating*, or multiplicative, operation between the DV and the PPC. In the VITE model, for example, a GO signal gates the DV before the PPC can integrate the $DV \cdot GO$ product (Sections 8 and 14). The GO signal is an example of a *movement gate*, because it is open during a movement. A *posture gate* is a gate that is open between movements, when the system is maintaining a fixed posture. Pauser cells are examples of posture gates that close during saccadic eye movements (Grossberg and Kuperstein, 1989; Keller, 1981; Robinson, 1975; Schlag-Rey and Schlag, 1983).

Figure 14

In Models 4–6, we assume the existence of a posture gate, or pauser cell, between the DV and the PPC (Figures 13–15). This gate opens while the initial eye position-plus-retinal target position estimate is loaded from the distributed TPC stages into the PPC stages via the DV stage. The gate closes during the movement, thereby protecting the stored estimate from being altered by the changing eye positions and retinal positions that are activated during the movement. After the movement is over, a new estimate of eye position and retinal position is read out of their respective TPCs. The DV stage compares this new estimate with the old, stored estimate. Non-zero components of the DV act as error signals that changes the adaptive weights of the $TPC \rightarrow DV$ pathways via VAM learning. It is assumed that the pauser gate stays closed long enough after the movement occurs for some such learning to occur, before the new TPC estimate is loaded into the PPC. Then the process repeats itself. The computer simulations summarized below show that the VAM can learn an invariant many-to-one head-centered representation from the time series of these internally generated error estimates.

Figure 15

17. An Exposition of Model 4

For definiteness, we describe the equations for Model 4 in detail before showing representative simulations of all the models. The network simulations are restricted to movements in the horizontal plane. A mathematical analysis is also provided in Section 19 that demonstrates the existence of an ideal set of adaptive weights. Computer simulations show that the network weights converge to the ideal weights during VAM learning. The simulations also show that the network discovers an invariant and unique representation of target location, which can then be used to generate motor commands to foveate or reach the target.

Model 4 is summarized in Figure 13. Given a target in the horizontal plane at some distance r from the cyclopean egocenter and angle θ from the sagittal plane, the angles that the eyes must realize in order to foveate the target are given by:

$$\theta_L = \tan^{-1}\Big(\frac{R_H \sin\theta_H + d/2}{R_H \cos\theta}\Big) \tag{28}$$

and

$$\theta_R = \tan^{-1}\Big(\frac{R_H \sin\theta_H - d/2}{R_H \cos\theta_H}\Big), \tag{29}$$

where d is the distance between rotation centers of the eyes (set to 2.75 inches in the simulations). Given angle θ_L of the left eye, the corollary discharges of the left extraocular muscles that maintain the eye at this position follow from (7) and (8); namely,

$$l_1 = \frac{1}{2} - \frac{\theta_L}{\pi}, \tag{30}$$

and

$$l_2 = \frac{1}{2} + \frac{\theta_L}{\pi}. \tag{31}$$

Note that the sum $l_1 + l_2$ is constant and equal to 1, independent of the value of θ_L, as in (7). Likewise, for the right angle of θ_R, it follows from (9) and (10) that

$$r_1 = \frac{1}{2} - \frac{\theta_R}{\pi}, \tag{32}$$

and

$$r_2 = \frac{1}{2} + \frac{\theta_R}{\pi}. \tag{33}$$

As in (21) and (22), the opponent head-centered representation of θ_H is given by

$$h_1 = \frac{r_1 + l_1}{2} \tag{34}$$

and

$$h_2 = \frac{r_2 + l_2}{2}. \tag{35}$$

An opponent head-centered representation of target vergence can likewise be derived from

$$h_5 = \frac{1}{2} + r_1 - l_1, \tag{36}$$

and

$$h_6 = \frac{1}{2} + l_2 - r_2. \tag{37}$$

The motor vector (h_1, h_2, h_5, h_6) represents the 3-D position of a foveated target in the horizontal plane.

When a target is presented in a position in which the eyes are not looking, as in Figure 10, the target image excites the retinas at a certain distance from the fovea. This distance depends upon the angle through which each eye must move to foveate the target. When the eyes are foveating a position (R^P, θ^P) with radius R^P and azimuth θ^P, the present eye angles can be calculated from equations (28) and (29). When a new target is presented at position (R^T, θ^T), the eye angles necessary to foveate the target (θ_L^T, θ_R^T) can also be calculated from equations (28) and (29). The difference between the angles is given by:

$$\Delta\theta_L = \theta_L^T - \theta_L^P \tag{38}$$

and

$$\Delta\theta_R = \theta_R^T - \theta_R^P. \tag{39}$$

Each retina consisted of a one-dimensional array of nodes, since the simulations reported here consider only horizontal eye movements. The target position T that is maximally activated by a light corresponding to angle $\Delta\theta$ (either $\Delta\theta_L$ or $\Delta\theta_R$) is given by

$$T = \frac{(\Delta\theta + \Delta\theta_{\max})(T_{\max} - 1)}{\Delta\theta_{\max}}, \tag{40}$$

where $\Delta\theta_{\max}$ is the maximum angle relative to the fovea at which a target will fall on the retina (set to 100° in the simulations), and $T_{\max}$ is the total number of retinal positions. This formula sweeps out nodal positions between zero and $T_{\max} - 1$. If the analog target position T falls between two discrete positions i and $i + 1$, namely $i \leq T \leq i + 1$, then the retinal activity values V_i and V_{i+1} at these nodes were set equal to $V_i = T - i$ and $V_{i+1} = i + 1 - T$. All other $V_j = 0$. This interpolation scheme defines a continuous linear generalization gradient across the retina, which reduced quantization effects and speeded learning. The subscript indicating the left or the right eye has been dropped because this formula works for both. When there are two one-dimensional monocular retinas, as in Models 1 and 4, two monocular representations are activated. When binocular two-dimensional maps are used, as in Models 2,3,5, and 6, only one representation is activated. In all cases, the retina can be considered to be one large column vector. This vision vector is denoted by V in all models. This notation makes the following equations independent of the type of architecture used. Generalization gradients were also used in the binocular visual representations, as described below.

The activity at the DV stage is given by

$$\Delta\hat{h}_i = h_i + Z_i \cdot V - \hat{h}_i, \tag{41}$$

where $i = 1, 2, 5, 6$; h_i is the present foveated eye position vector; Z_i is the vector of adaptive weights from the retina to component i; V is the vision vector; and $\hat{h}_i$ is the previous internal representation of target location. Notation $Z_i \cdot V$ denotes the dot product of Z_i with V. Conceptually, $h_i + Z_i \cdot V$ represents the prediction of the head-centered representation by the network. It is assumed that $\hat{h}_i$ is zero when a target first appears. Thus $\Delta\hat{h}_i$ stores this

prediction. After the eyes move, the stored $\Delta\hat{h}_i$ vector is compared with the new $h_i + Z_i \cdot V$ vector. Now, $\Delta\hat{h}_i$ codes the difference between two predictions of the location of the *same target.* Any non-zero value indicates an error or, more precisely, an inconsistency in the internal representation. This error is used to change the weights in such a way that the error is reduced by the VAM learning equation:

$$\frac{dz_{ij}}{dt} = -\delta\Delta\hat{h}_j x_i, \tag{42}$$

where z_{ij} is the weight from vision component i to DV component j, δ is the learning rate, $\Delta\hat{h}_j$ is the j^{th} DV component, and x_i is the activity of the j^{th} retinal component.

The simulations were carried out as follows:

(1) The eyes were randomly moved to some fixation point in their work-space (R^P, θ^P).

(2) The head-centered representation of this point was calculated according to equation (34)–(37).

(3) A target was presented at a random position (R^T, θ^T).

(4) The vision indices and activations were calculated as discussed above.

(5) $\Delta\hat{h}_i$ was calculated according to equation (41) with $\hat{h}_i = 0$.

(6) The eyes were moved to a random new location (the target stays the same).

(7) $\hat{h}_i$ was set equal to the previous values of $\Delta\hat{h}_i$.

(8) The new vision and eye position representations were calculated for the new eye positions.

(9) The new values of $\Delta\hat{h}_i$ were calculated according to equation (41) with $\hat{h}_i$ equal to its new value.

(10) The weights were updated according to equation (42).

(11) The cycle was repeated.

18. Computer Simulations

The network was trained for 500,000 trials with a learning rate δ in (42) of 0.5. The workspace was defined by a minimum radius of 10 inches, a maximum radius of 30 inches, a minimum θ_H of $-45°$, and a maximum θ_H of $+45°$. Adaptive weights z_{ij} were initialized to zero. Each retina had 50 discrete positions i.

18.1 Gaze Angle Component

Figure 16 shows the results for the $\hat{h}_1$ component of gaze angle. The target was moved randomly to all points in the workspace and the foveation point was held stationary at $R^P = 20$ and $\theta^P = 0°$. Ideally, $\hat{h}_1$ should change linearly with the target gaze angle. Figures 16a and 16b show that, indeed, the $\hat{h}_1$ component is linear with the target gaze angle and is essentially independent of target vergence. In Figure 16a, $\hat{h}_1$ is shown as the target vergence is changed for different values of the target gaze angle θ^T with the foveation point held stationary at $R^P = 20$ and $\theta^P = 0°$. Note that $\hat{h}_1$ does not change with changes in target vergence. However, it does change for changes in target gaze angle, as shown in Figure 16b. Figure 16b shows that, in fact, $\hat{h}_1$ changes linearly with target gaze angle and the slope is $-\frac{1}{\pi}$ as predicted in the analysis of Section 19 below. The dynamic range of $\hat{h}_1$ is approximately 0.5 in all the models.

In Figure 16c and 16d, $\hat{h}_1$ is shown as the foveation vergence and gaze angle were varied over the entire workspace while the target was stationary ($R^T = 20$, $\theta^T = 0°$). Since the target does not change position, its internal representation should not change. These figures show that the $\hat{h}_1$ component does not change. Together these figures show that the internal representation of target gaze angle is invariant over eye rotations.

Figure 16

18.2 Vergence Component

Figures 17a and 17b show how the internal representation of target vergence $\hat{h}_5$ changes

as the target is moved to all points in the workspace while the present foveation position is fixed at $R^P = 20$, $\theta^P = 0°$. Ideally, $\hat{h}_5$ should change linearly with target vergence and not at all with target gaze angle. Figure 17a shows $\hat{h}_5$ as the target vergence is changed for different values of the gaze angle. As predicted in the analysis of Section 19 below, the slope is positive with a value of $\frac{1}{\pi}$. The dynamic range of $\hat{h}_5$ extends from .38 to .44. Figure 17b demonstrates that $\hat{h}_5$ changes little when the target gaze angle is changed and target vergence is fixed, although this is not a requirement for invariance. Together, these graphs show that a unique target vergence is mapped to a unique learned internal representation of target vergence throughout the workspace.

Figure 17

Figures 17c and 17d show that $\hat{h}_5$ is invariant when the target is stationary and the present foveation position is moved to all points in the workspace. Figure 17c shows how $\hat{h}_5$ changes with changes in the fixation radius for different values of fixation gaze angle. The curve is nearly flat and all the curves are nearly identical. The slope and differences are not significant relative to the dynamic range. These slight aberrations are due to the fact that the weights have not yet converged to their ideal values. In simulations where the network was allowed to train longer, these fluctuations disappeared. Figure 17d shows how $\hat{h}_5$ responds to changes in foveation gaze angle for different foveation radii. Ideally, the curves should not be distinguishable. The small differences between the actual weights and the ideal weights again disappear when the network is allowed to train longer.

18.3 Adaptive Weights

The system analysis predicts values to which the network should converge for perfect performance (see Section 19). The predictions specify a slope and an arbitrary offset. In this section, we examine the learned weight matrices and show that they do indeed converge to the predicted slope. Figure 18a shows the weights from the left retina to the $\hat{h}_1$ DV component. The horizontal axis is the retinal node number. Each retinal node corresponds

to a certain value of retinal angle $\Delta\theta_L$. The relationship between node number and retinal angle is linear, but it need not be. Along with the actual weights, Figure 18a shows the predicted ideal weights with zero offset; i.e., C_θ in (48) below. The slopes are identical, as described; the offsets are arbitrary and do not influence performance. The deviations from the ideal weights at the extremes are due to the fact that these locations lie beyond the specified workspace and are never sampled and so never learned. The value of the offset appears to depend upon two factors. The first factor is the average weight at the beginning of training. In this simulation, the average value was 0.0. Note that the offset is about 0.0. The second factor is the distribution of sampled retinal locations. Because the left eye is left of the center of a symmetric workspace, targets are more likely to occur in the left portion of the fovea. This causes the weight curve to shift to the right, thereby increasing the offset. Just the opposite happens for the weights form the right retina, as is shown in Figure 18b where the offset is slightly less than that of 18a. Figure 18b demonstrates that weights corresponding to the right retina converge on the ideal slope. Figures 18c and 18d show the weight matrices from the left and right retinas to the $\hat{h}_5$ DV component along with their ideal, zero-offset values. (See Section 19 for a discussion of these values.) As can be seen, the difference in slopes between the actual and ideal are nearly zero.

Figure 18

The performance of Models 1–3, 5, and 6 were also evaluated using computer simulations. The performance graphs for all models on both the vergence and gaze angle components were essentially identical to those of Model 4 shown in Figures 16 and 17. For all models, the steady-state error for both components was below .5% indicating that all the models have similar asymptotic performance. The main difference in performances was in the time it took the networks to converge. Model 1 (explicit, monocular) converged the fastest (less than 400,000 trials at $\delta = .5$). Models 2 and 3 (both explicit, binocular) converged in less than 2,000,000 trials at $\delta = .5$. The models with the implicit teacher (Models 4–6) took

slightly longer to converge than their explicit counterparts. The binocular models converged more slowly because the interpolation scheme used in the simulations caused many sites to become active at once, but each with a low activation level; thus, each location learned more slowly. In simulations with only a few locations active at a high level, the binocular models converged as quickly as the monocular models. The convergence of Models 2 (explict) and 5 (implicit) are shown in Figure 19 for both the gaze angle and vergence components. Each point represents the average absolute error at the DV stage over 1000 trials; the vertical axis is this error was divided by the total dynamic range of the component. Because all points in the workspace are being sampled randomly during the generation of the curve in Figure 19, this is a measure of the global performance of the network.

The learning rate depended upon two factors: how often a node became active, and the activity that it attained. For the monocular models, each node became active on approximately 4.8% of the trials with an average activity of .5. For the binocular models, each node became active on approximately 3.2% of the trials with an average value of .018.

Figure 19

19. Derivation of Ideal Weight Vectors

It will now be shown for Model 4 that there exists a set of weights for which the performance of the network is perfect, given a retina with infinite resolution (i.e., no quantization error) on which each target activates a single location. These results can be extended to a discrete retina in which a target activates several locations in a smooth manner. First we review the pertinent geometry and system equations, then we derive a differential equation for the ideal weights using the performance constraints. Next we solve the differential equation to obtain the ideal weights and show that the other performance constraints are also satisfied.

Figure 20

There are six basic performance constraints on the system. The basic idea is that an internal representation of a target position should not change when the target is fixed and the foveation position is changed. Also, there should be a unique mapping between the internal representation and its external analog. These constraints are mathematically defined as follows. The internal representation of target vergence is $\hat{h}_5$. Invariance of $\hat{h}_5$ over eye movements is defined by the equations

$$\frac{\partial \hat{h}_5}{\partial \gamma^P} = 0 \tag{43}$$

and

$$\frac{\partial \hat{h}_5}{\partial \theta^P} = 0. \tag{44}$$

Equations (43) and (44) require that $\hat{h}_5$ does not change for changes in fixation vergence and gaze angle. The uniqueness constraint can be fulfilled by the following equation:

$$\frac{\partial \hat{h}_5}{\partial \gamma^T} = C_\gamma, \tag{45}$$

where C_γ is a non-zero constant. Equation (45) means that the internal representation of vergence changes linearly with actual target vergence γ^T. Linearity is a more rigorous constraint than uniqueness but, as shown below, it is achieved by the network.

Figure 21

The internal representation of the gaze angle is $\hat{h}_1$. Invariance of this component over changes in fixation position is given by equations

$$\frac{\partial \hat{h}_1}{\partial \theta^P} = 0 \tag{46}$$

and

$$\frac{\partial \hat{h}_1}{\partial \gamma^P} = 0. \tag{47}$$

The uniqueness (linearity) constraint is given by

$$\frac{\partial \hat{h}_1}{\partial \gamma^T} = C_\theta. \tag{48}$$

We now describe how to define ideal weights such that all the six constraints (43)–(48) are obeyed.

The geometry of the foveation system is shown in Figure 20. The eyes are at some fixation position when a new target is presented. When the eyes foveate the fixation point, the angle of the left eye is

$$\gamma^P = \theta_L^P - \theta_R^P. \tag{49}$$

For the target position,

$$\gamma^T = \theta_L^T - \theta_R^T. \tag{50}$$

Thus, the change in vergence due to the eye movement is

$$\gamma^T - \gamma^P = (\theta_L^T - \theta_R^T) - (\theta_L^P - \theta_R^P). \tag{51}$$

Rearranging terms gives

$$\gamma^T - \gamma^P = (\theta_L^T - \theta_L^P) - (\theta_R^T - \theta_R^P), \tag{52}$$

which is just the difference between the eye angles before and after the movement:

$$\gamma^T - \gamma^P = \Delta\theta_L - \Delta\theta_R, \tag{53}$$

where $\Delta\theta_L$ and $\Delta\theta_R$ define how far the left eye and right eye need to move to foveate the target. Combining (8), (10), (16), and (22) leads to approximations for θ^P and θ^T, namely

$$\theta^P \approx \frac{\theta_L^P + \theta_R^P}{2} \tag{54}$$

and

$$\theta^T \approx \frac{\theta_L^T + \theta_R^T}{2}, \tag{55}$$

which are accurate for points whose distance from the head is sufficiently large relative to the distance between the eyes. Thus, as in (53),

$$\theta^T - \theta^P \approx \frac{\Delta\theta_L + \Delta\theta_R}{2}. \tag{56}$$

By (7) and (8), the corollary discharge of the left extraocular muscle of the left eye is

$$l_1 = \frac{1}{2} - \frac{\theta_L^P}{\pi} \tag{57}$$

and for the left extraocular muscle of the right eye is

$$r_1 = \frac{1}{2} - \frac{\theta_R^P}{\pi}. \tag{58}$$

Using (49), (57), and (58), the simulated internal representation of vergence in (36) becomes

$$h_5 = \frac{1}{2} + r_1 - l_1 = \frac{1}{2} + \frac{1}{\pi}\gamma^P \tag{59}$$

which implies that

$$\frac{\partial h_5}{\partial \gamma^P} = \frac{1}{\pi}. \tag{60}$$

If a target activates only one retinal position of each eye with a strength of 1.0, then, by (41), the internal representation of target vergence is

$$\hat{h}_5 = h_5 + z_{L5} + z_{R5}, \tag{61}$$

where z_{L5} is the weight from the active location in the left retina to the $\hat{h}_5$ component of the DV stage, and z_{R5} is the weight from the active location in the right retina. Now differentiating both sides of (61) with respect to the fixation vergence γ^P and setting the result equal to zero, as required by (43), we obtain

$$\frac{\partial \hat{h}_5}{\partial \gamma^P} = \frac{\partial h_5}{\partial \gamma^P} + \frac{\partial z_{L5}}{\partial \gamma^P} + \frac{\partial z_{R5}}{\partial \gamma^P} = 0. \tag{62}$$

Combining (60) and (62) shows that

$$\pi \frac{\partial z_{L5}}{\partial \gamma^P} + \pi \frac{\partial z_{R5}}{\partial \gamma^P} = -1, \tag{63}$$

which specifies how changes in the internal representation of vergence are balanced against changes in the vergence weights as γ^P varies. Equation (53) provides another equation of

balance for the corresponding external parameters. Here vergence changes are balanced against azimuth changes. Comparison of (53) and (63) suggests that vergence weights adapt to azimuth changes. More precisely, differentiating (53) with respect to γ^P yields

$$-1 = \frac{\partial \Delta\theta_L}{\partial \gamma^P} - \frac{\partial \Delta\theta_R}{\partial \gamma^P}. \tag{64}$$

Equating corresponding terms in (63) and (64) leads to the anzatz that

$$\frac{\partial z_{L5}}{\partial \gamma^P} = \frac{1}{\pi}\frac{\partial \Delta\theta_L}{\partial \gamma^P} \tag{65}$$

and

$$\frac{\partial z_{R5}}{\partial \gamma^P} = -\frac{1}{\pi}\frac{\partial \Delta\theta_R}{\partial \gamma^P}. \tag{66}$$

Integrating these equations suggests that the ideal vergence weights are

$$z_{L5} = \frac{1}{\pi}\Delta\theta_L + C_{L5} \tag{67}$$

and

$$z_{R5} = \frac{-1}{\pi}\Delta\theta_R + C_{R5}, \tag{68}$$

where C_{L5} and C_{R5} are constants of integration. When a target is presented, it activates locations on the left and right retinas given by $\Delta\theta_L$ and $\Delta\theta_R$, which define how far the eyes have to move to foveate the target (see Figure 2 and equation (53)). Equations (67) and (68) specify ideal weights for these locations. When these equations are substituted into (61), the change of $\hat{h}_5$ with respect to fixation gaze angle θ^P is zero, as required by (44), and the change with respect to γ^P is a positive constant $\frac{1}{\pi}$, as required by (60).

Using a similar procedure, the weights from each of the retinas to the $\hat{h}_1$ component can be derived and are given by

$$z_{L1} = \frac{-1}{2\pi}\Delta\theta_L + C_{L1} \tag{69}$$

and

$$z_{R1} = \frac{-1}{2\pi}\Delta\theta_R + C_{R1}. \tag{70}$$

These weight formulas are accurate approximations as long as the target and fixation points are far relative to the distance between the eyes. These weights provide invariance with respect to changes in fixation vergence and gaze angle, as required by (46) and (47). The internal representation of target gaze angle is also linear, with slope $C_\theta = \frac{-1}{\pi}$, which guarantees weight uniqueness by (48).

This type of analysis has also been used to derive the weights for a two-dimensional binocular look-up table, and for networks wherein a target generates a diffuse Gaussian region of activation on the retinas. The computer simulations show that all the networks actually converge to these ideal weights.

20. A Sketch of Model 5

In order to clarify the key differences between the monocular and binocular models, the main features of Model 5 will now be summarized. Model 5 differs from Model 4 in its use of binocular position and disparity computations. The binocular position was computed from the equation

$$\Delta\theta_H = \frac{1}{2}(\Delta\theta_L + \Delta\theta_R) \tag{71}$$

and the binocular disparity was computed from the equation

$$\Delta D = \Delta\theta_L - \Delta\theta_R \tag{72}$$

where $\Delta\theta_L$ and $\Delta\theta_R$ are the retinal offsets of the target in the left and right retinas. The binocular spatial map index corresponding to ΔD is given by

$$T = \frac{(\Delta D + \Delta D_{\max})(T_{\max} - 1)}{\Delta D_{\max}} \tag{74}$$

where $\Delta D_{\max}$ is the maximum deviation of the disparity, set to 15° in the simulations, and $T_{\max}$ is the maximum number of positions in each map dimension. The binocular spatial index for the position was calculated as in (40) with $\Delta\theta = \Delta\theta_H$ and $\Delta\theta_{\max} = 100°$.

The ideal weights from the map to the DV stage were derived assuming that only one point becomes active in the map with an activity of one. The ideal weights to the gaze angle component $\hat{h}_1$ are:

$$z_1 = \frac{-\Delta\theta_H}{\pi} + C_1. \tag{75}$$

The ideal weights to the vergence angle component ($\hat{h}_5$) are:

$$z_5 = \frac{-\Delta D}{\pi} + C_5, \tag{76}$$

where C_1 and C_5 are constants of integration. Note that the weights from a column (constant ΔD) to $\Delta\hat{h}_5$ are the same. Likewise all the weights from a row (constant $\Delta\theta_H$) to $\Delta\hat{h}_1$ are the same.

This network was simulated using a 50×50 visual position map. The following generalization gradient was used to convert analog target position $(\Delta\theta_H, \Delta D)$ into activations of the vision vector V. Suppose that the distance from the target position to binocular lattice position (i, j) is

$$d_{ij} = \sqrt{(\Delta\theta_H - i)^2 + (\Delta D - j)^2}. \tag{77}$$

Let

$$v_{ij} = \begin{cases} 1 - \frac{d_{ij}}{3\sqrt{2}} & \text{if } d_{ij} < 3\sqrt{2} \\ 0 & \text{otherwise} \end{cases}. \tag{78}$$

Then the activity at (i, j) of the vision vector equals

$$V_{ij} = \frac{v_{ij}}{v}, \tag{79}$$

where $v = \sum_{i,j} v_{ij}$. Thus the total activity of the vision vector is normalized to equal 1.

21. Concluding Remarks: Interactions between Visual, Motor, and Spatial Representations

The present article suggests how outflow eye movement commands from each of the two eyes can be binocularly combined. Two successive stages of opponent processing convert

these commands into a cyclopean representation of head-centered azimuth, elevation, and vergence. This motor representation specifies the position in 3-D space that the two eyes are both foveating at any time.

When a nonfoveated visual target activates both retinas, the activated retinal locations, taken together with the cyclopean eye position representation, implicitly code the position of the target in 3-D space. Such a distributed representation may be transformed, via a VAM learning module, into an invariant head-centered representation of 3-D target position. The VAM model illustrates how an accurately tuned visually reactive movement system can be a source of teaching signals whereby the many-to-one transformation is learned. After VAM learning takes place, the invariant head-centered representation can control internally *planned* movements that are capable of overriding *visually reactive* movements that would otherwise occur in response to environmental fluctuations (Grossberg and Kuperstein, 1989).

VAM learning is also capable of discovering an invariant spatial representation even if an explicit teaching signal does not exist. Here, the model detects invariant structure that is hidden in a time series of environmental fluctuations. It does so by comparing previous estimates of the invariant with present data that represent the same target position, and uses DV learning to cancel inconsistent signals.

This comparison process utilizes a multiplicative gate that acts between the DV and PPC stages of the VAM. In related VAM applications, such gates can control the production of variable movement speeds (GO signal) or variable movement sizes (GRO signal). Thus the gating option is a general design constraint that enables invariant structure to be discovered for purposes of learning, while also allowing this invariant structure to be performed through variable movements whose characteristics may be flexibly modified to meet changing environmental conditions. The gates thus afford a huge reduction in memory load, by allowing a single learned invariant structure to be expressed in many different ways.

From a more cognitive perspective, these various gating signals are all different expres-

sions of the will-to-act. The VAM modules provide a unified computational format wherein the will-to-act can be expressed in several ways while invariant transformations are learned in real-time. In particular, a series of VAM modules, forming a VAM Cascade, can learn a sensory-to-spatial transformation followed by a spatial-to-motor transformation. The fact that a single type of neural circuit can be used for both types of transformation, while providing the crucial property of synchronous trajectory formation for free, clarifies how consistent perception-action cycles are organized. It also provides a new understanding of why neural vectors are computed in the several cortical areas—including parietal, frontal, and motor cortices—that contribute to spatial orientation and motor control (Bruce and Goldberg, 1984; Georgopoulos, Kalaska, Caminiti, and Massey, 1982; Georgopoulos, Schwartz, and Kettner, 1986; Gradt and Anderson, 1988).

References

Anderson, R.A., Essick, G.K., and Siegel, R.M. (1985). Enclosing of spatial location by posterior parietal neurons. *Science*, **230**, 456–458.

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, 1978.

Bruce, C.J. and Goldberg, M.E. (1984). Physiology of the frontal eye fields. *Trends in Neuro-Sciences*, **7**, 436–441.

Bullock, D. and Contreras-Vidal, J.L. (1991). How spinal neural networks reduce discrepancies between motor intention and motor realization. In K.M. Newell and D.M. Corcos (Eds.), **Variability and motor control**. Champagne, IL: Human Kinetics Press. Boston University Technical Report Series **CAS/CNS-TR-91-023**.

Bullock, D., Contreras-Vidal, J.L., and Grossberg, S. (1992). A neural network for spino-muscular generation of launching and braking forces by opponent muscles. Submitted to **Proceedings of the 1992 International Joint Conference on Neural Networks**.

Bullock, D., Greve, D, Grossberg, S., and Guenther, F.H. (1992). A head-centered representation of 3-D target location derived from opponent eye position commands. Submitted to **Proceedings of the International Joint Conference on Neural Networks**.

Bullock, D. and Grossberg, S. (1988a). Neural dynamics of planned arm movements: Emergent invariants and speed-accuracy properties during trajectory formation. *Psychological Review*, **95**, 49–90.

Bullock, D. and Grossberg, S. (1988b). The VITE model: A neural command circuit for generating arm and articulator trajectories. In J.A.S. Kelso, A.J. Mandell, and M.F. Shlesinger (Eds.), **Dynamic patterns in complex systems**. Singapore: World Scientific Publishers.

Bullock, D. and Grossberg, S. (1989). VITE and FLETE: Neural modules for trajectory

formation and postural control. In W. Hershberger (Ed.), **Volitional action.** Amsterdam: North-Holland, 253–297.

Bullock, D. and Grossberg, S. (1991). Adaptive neural networks for control of movement trajectories invariant under speed and force rescaling. *Human Movement Science*, **10**, 3–53.

Bullock, D., Grossberg, S., and Guenther, F. (1992). A self-organizing neural network model for redundant sensory-motor control, motor equivalence and tool use. Submitted to the **Proceedings of the 1992 International Joint Conference on Neural Networks.**

DeJong, R., Coles, M.G.H., Logan, G.D., and Gratton, G. (1990). In search of the point of no return: The control of response processes. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 164–182.

DeValois, R.L. and DeValois, K.K. (1975). Neural coding of color. In E.C. Carterette and M.P. Friedman (Eds.), **handbook of perception. Volume 5: Seeing.** New York: Academic Press.

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

Gaudiano, P. and Grossberg, S. (1991). Vector associative maps: Unsupervised real-time error-based learning and control of movement trajectories. *Neural Networks*, **4**, 147–183.

Gaudiano, P. and Grossberg, S. (1992). Adaptive vector integration to endpoint: Self-organizing neural circuits for control of planned movement trajectories. *Human Movement Science*, **11**, 141–155.

Gibson, J.J. (1966). **The senses considered as perceptual systems.** Boston: Houghton Mifflin.

Georgopoulos, A.P., Kalaska, J.F., Caminiti, R., and Massey, J.T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate cortex. *Journal of Neuroscience*, **2**, 1527–1537.

Georgopoulos, A.P., Kalaska, J.F., Crutcher, M.D., Caminiti, R., and Massey, J.T. (1984). The representation of movement direction in the motor cortex: Single cell and population

studies. In G.M. Edelman, W.E. Goll, and W.M. Cowan (Eds.), **Dynamic aspects of neocortical function**. Neurosciences Research Foundation, 501–524.

Georgopoulos, A.P., Schwartz, A.B., and Kettner, R.E. (1986). Neuronal population coding of movement direction. *Science*, **233**, 1416–1419.

Goodale, M.A. and Milner, D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, **15**, 20–25.

Gradt, J.W. and Anderson, R.A. (1988). Memory related motor planning activity in posterior parietal cortex of macaque. *Experimental Brain Research*, **70**, 216–220.

Greve, D., Grossberg, S., Guenther, F., and Bullock, D. (1992). Neural representations for sensory-motor control, I: Head-centered 3-D target positions from opponent eye commands. *Acta Psychologica*, in press.

Grossberg, S. (1982). **Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control**. Boston: Reidel Press.

Grossberg, S. (1992). 3-D vision and figure-ground separation by visual cortex. Technical Report CAS/CNS-TR-92-019, Boston, MA: Boston University. Submitted for publication.

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

Grossberg, S. and Kuperstein, M. (1989). **Neural dynamics of sensory-motor control: Expanded edition**. Elmsford, NY: Pergamon Press.

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

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

Hollerbach, J.M. Moore, S.P., and Atkeson, C.G. (1986). Workspace effect in arm movement kinematics derived by joint interpolation. In G. Gantchev, B. Dimitrov, and P. Gatev (Eds.),

Motor control. Plenum Press.

Horak, F.B. and Anderson, M.E. (1984a). Influence of globus pallidus on arm movements in monkeys, I: Effects of kainic acid-induced lesions. *Journal of Neurophysiology*, **52**, 290–304.

Horak, F.B. and Anderson, M.E. (1984b). Influence of globus pallidus on arm movements in monkeys, II: Effects of stimulation. *Journal of Neurophysiology*, **52**, 305–322.

Humphrey, D.R. and Reed, D.J. (1983). Separate cortical systems for control of joint movement and joint stiffness: Reciprocal activation and coactivation of antagonist muscles. In J.E. Desmedt (Ed.), **Motor control mechanisms in health and disease**. New York: Raven Press, 347–372.

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

Kaufman, L. (1974). **Sight and mind: An introduction to visual perception**. New York: Oxford University Press.

Keller, E.L. (1981). Brain stem mechanisms in saccadic control. In A.F. Fuchs and W. Becker (Eds.), **Progress in oculomotor research**. New York: Elsevier/North-Holland.

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

Nagasaki, H. (1989). Asymmetric velocity and acceleration profiles of human arm movements. *Experimental Brain Research*, **74**, 319–326.

Piaget, J. (1963). **The origins of intelligence in children**. New York: Norton.

Robinson, D.A. (1975). Oculomotor control signals. In G. Lennerstrand and P. Bach-y-Rita (Eds.), **Basic mechanisms of ocular motility and their clinical implications**. Oxford: Pergamon Press.

Schlag-Rey, M. and Schlag, J. (1983). Saccade-related pause-rebound cells in central thalamus of monkeys. *Society of Neuroscience Abstracts*, **9**, 1087.

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

von Tschermak-Seysenegg, A. (1952). **Introduction to physiological optics**. P. Boeder (Trans.). Springfield, IL: C.C. Thomas.

Werner, H. (1937). Dynamics in binocular depth perception. *Psychological Monograph*, (whole no. 218).

Figure Captions

Figure 1. The geometry of 3-D target of localization by the two eyes: Symbols L and R are the centers of the left and right eyes: (a) Left side shows how a closer target generates a larger vergence angle. Right side shows how the vergence angle is calculated from the angles of the eyes in their orbits. (b) shows the vergence as a function of target radius for a target on the sagittal plane.

Figure 2. Illustration of relationships between spherical coordinates R_H, ϕ_H, θ_H 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_H is measured from the origin to the binocular fixation point on the object. Elevation ϕ_H 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_H is defined similarly, but with respect to the x-z plane.

Figure 3. Geometry of cyclopean position: The angles θ_L and θ_R that the left eye and right eye assume to foveate a target correspond to a cyclopean, head-centered angle θ_H.

Figure 4. Control of the extraocular muscles: The muscles are arranged in agonist-antagonist pairs. Stimulation by neuron L_2 causes a contraction of the left medial muscle, which rotates the left eyeball to the right.

Figure 5. Opponent processing architecture for the calculation of the internal representation of gaze angle (h_2) and vergence (Γ). Signals L_1, L_2, R_1, and R_2 are corollary discharges from the outflow movement cells that control eye position as in Figure 4. The activity of each pair of cells is normalized at cells l_1, l_2, r_1, and r_2.

Figure 6. A schematic diagram of the Adaptive VITE (AVITE) circuit. The Now Print

(NP) gate copies the PPC into the TPC when the arm is stationary, and the adaptive synapses (semicircles in the TPC→DV pathways) learn to transform target commands into correctly calibrated outflow signals at the PPC. (Reprinted with permission from Gaudiano and Grossberg (1991).)

Figure 7. The VITE model, adapted from Bullock and Grossberg (1988a). TPC = Target Position Command, DV = Difference Vector, PPC = Present Position Command. The GO signal acts as a nonspecific multiplicative gate that can control the overall speed of a movement, or the will to move at all. Use of a single GO signal insures synchronous activation of all muscles in the synergies involved in a coordinated movement.

Figure 8. A diagrammatic illustration of a single babbling cycle in the AVITE. (a) The Endogenous Random Generator ON channel output (ERG ON) is integrated at the PPC, giving rise to random outflow signals that move the arm. (b) When the arm stops moving at ERG ON offset, a complementary ERG OFF signal opens the Now Print (NP) gate, copying the current PPC into the TPC through an arbitrary transformation. (c) The filtered TPC activation is compared to the PPC at the DV stage. DV activation would be zero in a properly calibrated AVITE. (d) The learning law changes TPC→DV synapses to eliminate any nonzero DV activation, thus learning the reverse of the PPC→NP→TPC transformation. (Reprinted with permission from Gaudiano and Grossberg (1991).)

Figure 9. A VAM Cascade: Activation of the upper left map represents eye position, and that of the upper right map represents target position on the retina. Activation from these two maps cooperate to form a head-centered representation. A given shift in eye position can be canceled by an equal and opposite shift in retinal target position. (Reprinted with permission from Gaudiano and Grossberg (1991).)

Figure 10. Model 1: Monocular visual representations with explicit teacher. A target activates a position on each retina which is stored until after movement. The initial position of the eyes generate the cyclopean head-centered representation (h_1, h_2, h_5, h_6) which is also

stored until after movement.. The visual and head-centered representations both project to the Difference Vector (DV) stage to generate a prediction of what the head-centered representation of the target will be when foveated. After movement, the target is foveated, and the teaching vector $(\hat{h}_1, \hat{h}_2, \hat{h}_5, \hat{h}_6)$ instates the actual head-centered representation of the target at a stage analogous to the AVITE PPC stage. The Posture Gate then opens, and the actual target representation is compared with the desired target representation to generate an error DV, which changes the adaptive weights that link the visual representations to the DV stage.

Figure 11. Model 2: Binocular disparity model with explicit teacher. When a target is presented, it activates a single site in each retina, as in the monocular model; then the retinas combine to form a two-dimensional spatial map of binocular position and disparity. Such a binocular map could be used to attentively choose a single target from multiple possible target positions. This model operates in the same way as Model 1 (Figure 10).

Figure 12. Model 3: Binocular look-up model with explicit teacher. This model combines the explicit teacher of Model 1 with a binocular look-up table that directly combines the monocular visual representations into a two-dimensional spatial array.

Figure 13. Model 4: Monocular model with implicit teacher. This model operates similarly to the monocular model with the explicit teacher. The difference is that the model discovers invariant 3-D target position representations from environmental fluctuations. With this model, the eyes do not need an independent system to accurately foveate the target in order to produce accurate teaching signals. See text for details.

Figure 14. Model 5: Binocular disparity model with implicit teacher. This model combines the binocular visual map of Model 2 with the implicit teacher of Model 4.

Figure 15. Model 6: Binocular look-up model with implicit teacher. This model combines the implicit teacher of Model 4 with the binocular look-up table of Model 3.

Figure 16. Performance of Model 4 on the gaze angle component $\hat{h}_1$: The model was trained over 500,000 trials at a learning rate of $\delta = 0.5$. (a) Plot of $\hat{h}_1$ as target vergence is changed with target gaze angle equal to $0°, 15°, 30°$, and $45°$ while the eyes foveate a point 20 inches directly in front of the nose. Ideally, $\hat{h}_1$ should be independent of the target vergence and should shift for shifts in gaze angle. (b) Plot of $\hat{h}_1$ as target gaze angle is changed with target vergence equal to $15.6°, 12.1°, 7.9°$, and $5.2°$. Ideally, $\hat{h}_1$ should be linear with target gaze angle and independent of target vergence. (c) Plot of $\hat{h}_1$ as the present fixation vergence is changed while the target position remains 20 inches directly in front of the head with present gaze angle equal to $0°, 15°, 30°$, and $45°$. Ideally, $\hat{h}_1$ should not change as the eyes move as long as the target is fixed. (d) Plot of $\hat{h}_1$ as the present fixation gaze angle is moved around the workspace with present vergence equal to $15.6°, 12.1°, 7.9°$, and $5.2°$. Ideally, there should be one flat curve indicating that $\hat{h}_1$ is not changing due to movement of the eyes.

Figure 17. Performance of Model 4 on the vergence component $\hat{h}_5$: Training parameters are the same as those of Figure 15. (a) Plot of $\hat{h}_5$ as the target vergence is changed with target gaze angle equal to $0°, 15°, 30°$, and $45°$. Ideally, $\hat{h}_5$ should be linear with target vergence and independent of the target gaze angle. (b) Plot of $\hat{h}_5$ as the target gaze angle is varied with target vergence equal to $15.6°, 12.1°, 7.9°$, and $5.2°$. Ideally, there should be four distinct, flat curves. (c) Plot of $\hat{h}_5$ as the present fixation vergence is varied with present gaze angle equal to $0°, 15°, 30°$, and $45°$. Ideally, there should be one flat curve. (d) Plot of $\hat{h}_5$ as the present fixation gaze angle is varied with present fixation vergence $15.6°, 12.1°, 7.9°$, and $5.2°$. Ideally, there should be one flat curve.

Figure 18. Learned adaptive weight values for Model 4. The ideal weight values are also shown. The actual weights may differ from the ideal weights by an offset and still give ideal performance. (a) Weights from the left retina to the $\hat{h}_1$ component of the DV stage. (b) Weights from the right retina to the $\hat{h}_1$ component of the DV stage. (c) Weights from the

left retina to the $\hat{h}_5$ component of the DV stage. (d) Weights from the right retina to the $\hat{h}_5$ component of the DV stage.

Figure 19. Convergence plots for Models 2 (Explicit) and 5 (Implicit). The vertical axis is the error (averaged over 1000 trials) expressed as a fraction of the dynamic range for the given component. (a) Convergence of the gaze angle $\hat{h}_1$ component. (b) Convergence of the gaze angle $\hat{h}_5$ component.

Figure 20. Geometry of 3-D target localization: In an intitial foveated gaze position, the left eye assumes an angle of θ_L^P, and the right eye assumes an angle of θ_R^P. To foveate the target position, the eyes assume angles θ_L^T and θ_R^T. The angular change $\Delta\theta_L$ is the difference between the angle that the left eye must assume to foveate the target and the angle where it starts out. Quantity $\Delta\theta_R$ is defined similarly. Quantity γ^P is the angle formed by the intersection of the rays emanating from the eyes in their initial gaze position. Vergence γ^T is defined similarly.

TABLE 1

Scaled Quantity	Will-to-Act Signal	Brain Region
speed	GO	globus pallidus
stiffness	CO	motor cortex, spinal cord
size	GRO	parietal cortex, basal ganglia

TABLE 2

	Monocular Visual Signals	Binocular Visual Signals
Explicit Teacher	Model 1	Models 2 and 3
Self-Organized Teacher	Model 4	Models 5 and 6

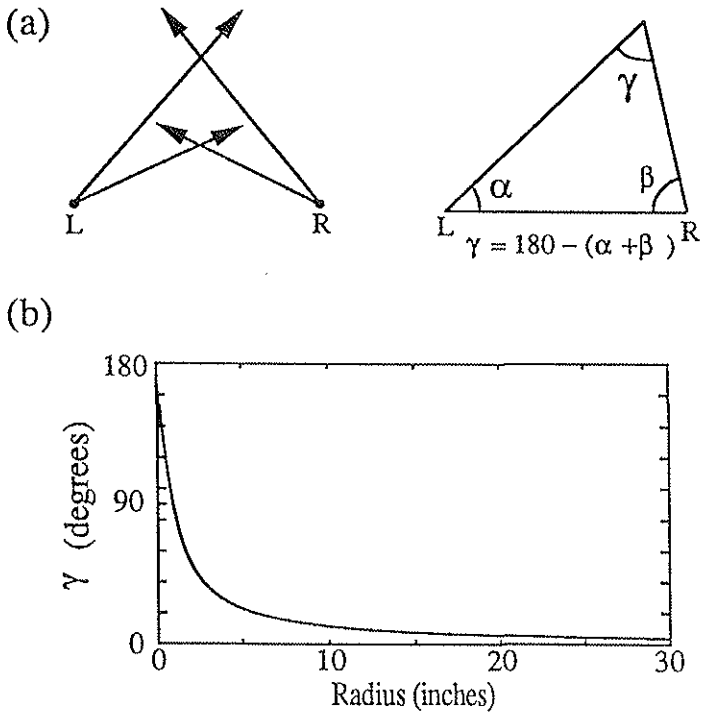


FIGURE 1

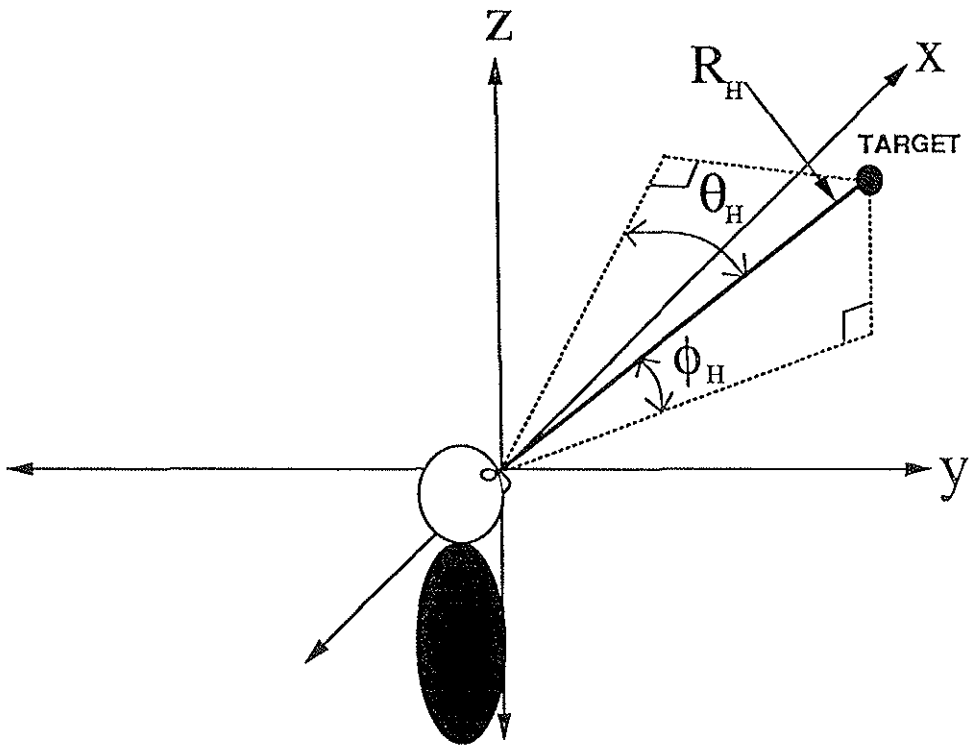


FIGURE 2

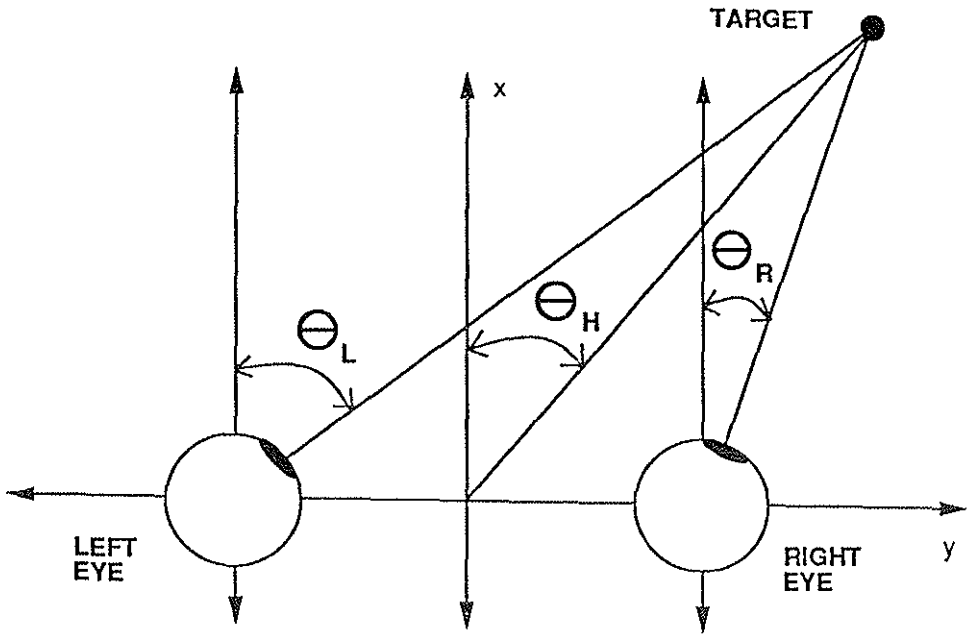


FIGURE 3

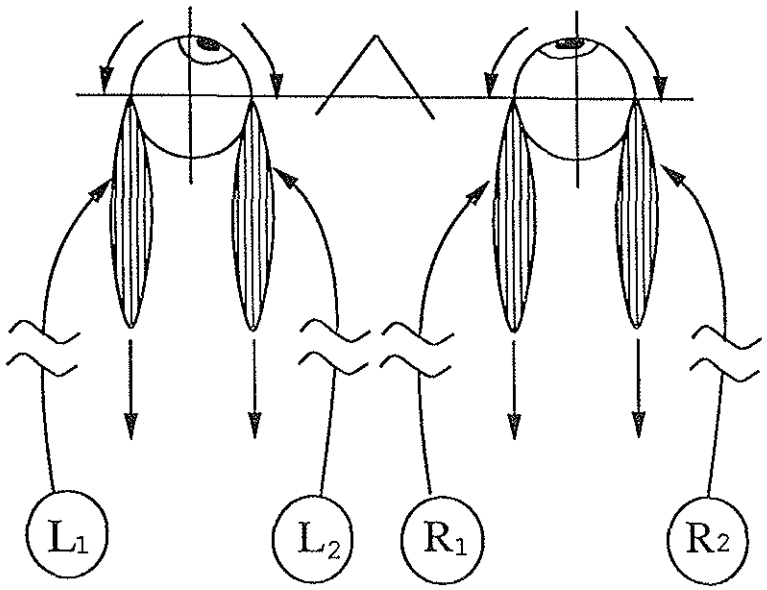


FIGURE 4

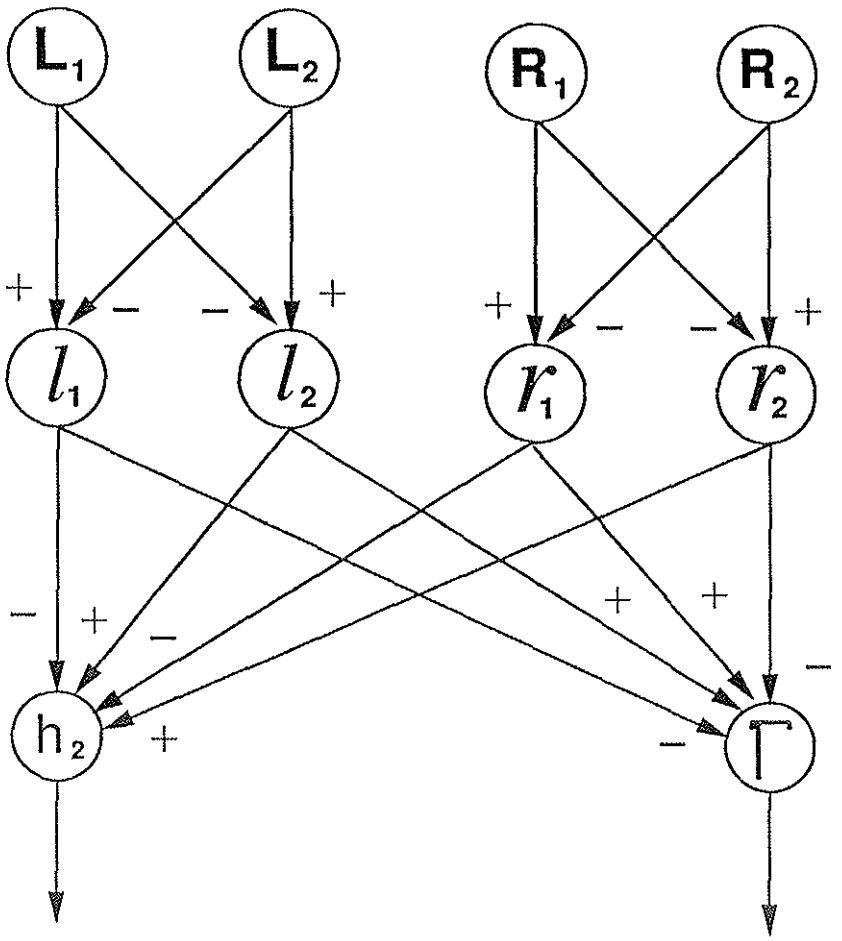


FIGURE 5

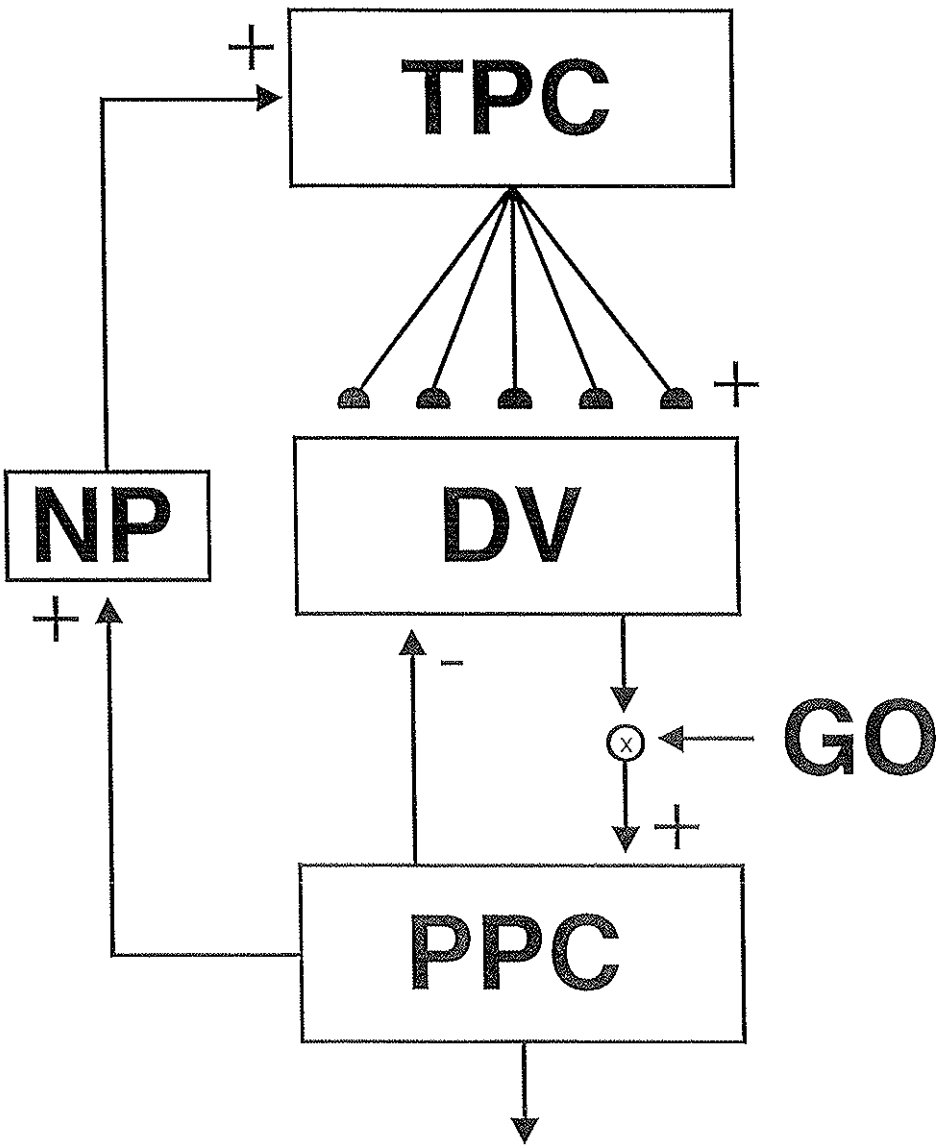


FIGURE 6

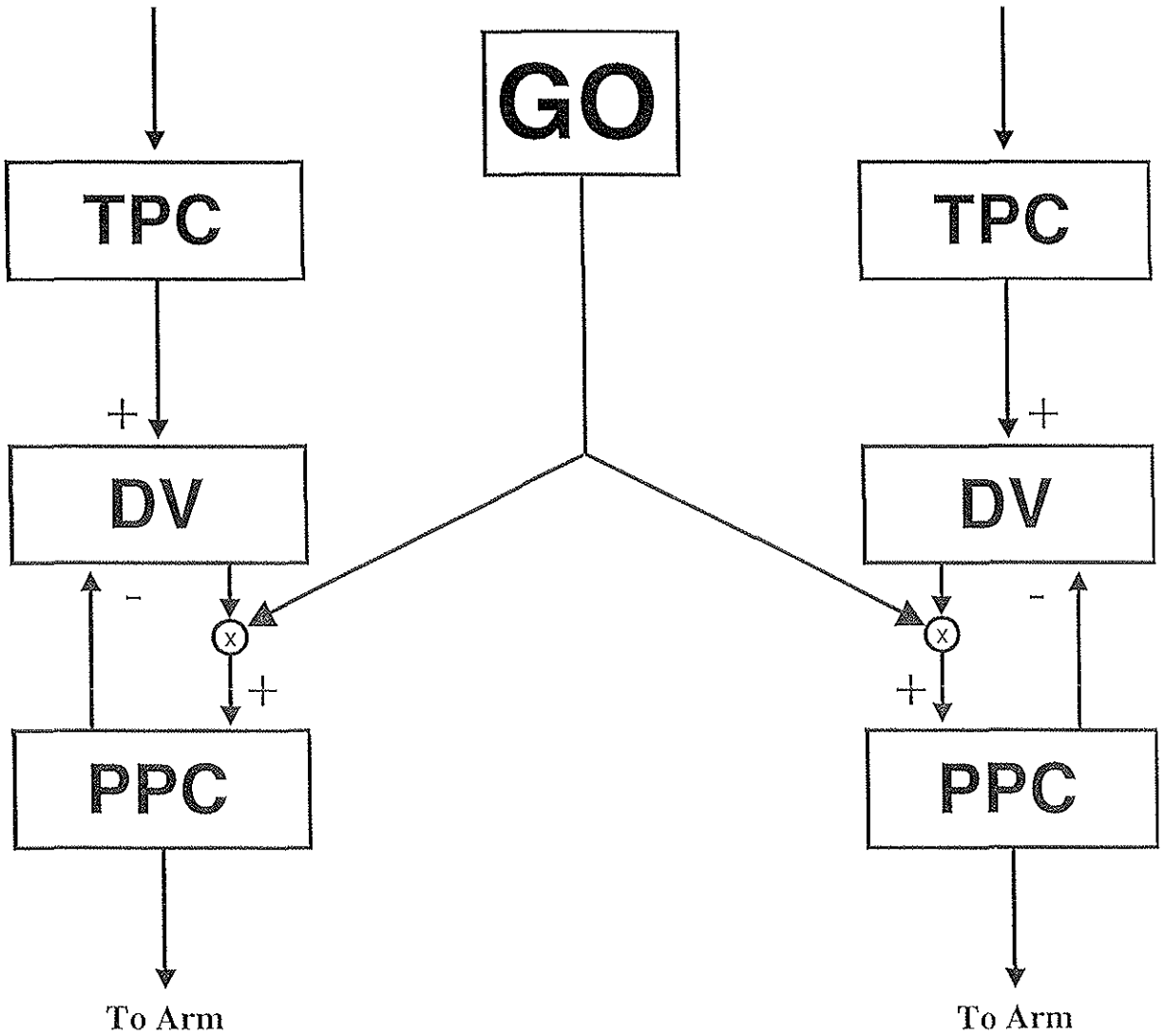


FIGURE 7

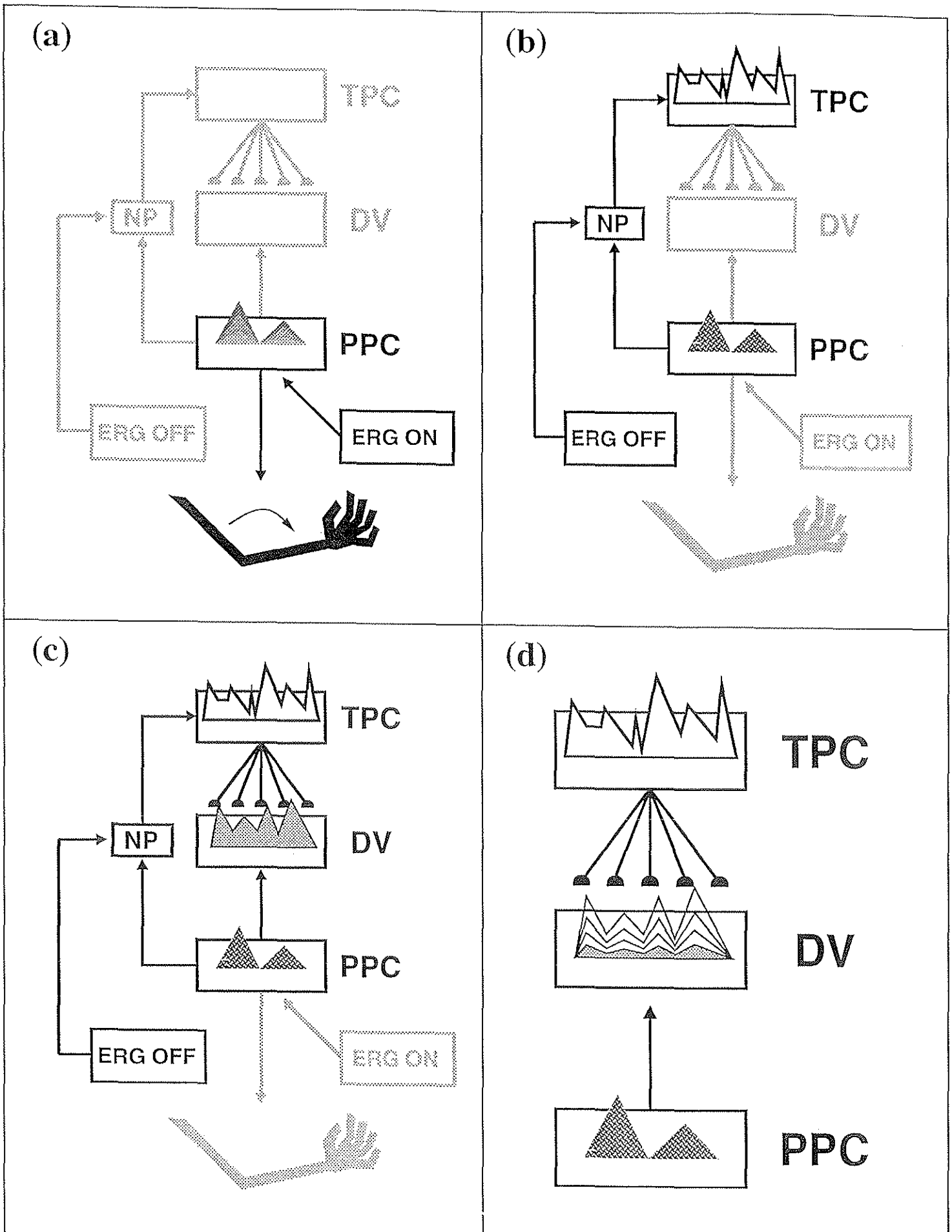

(a)
TPC
NP
DV
PPC
ERG OFF
ERG ON
(b)
TPC
NP
DV
PPC
ERG OFF
ERG ON
(c)
TPC
NP
DV
PPC
ERG OFF
ERG ON
(d)
TPC
DV
PPC

FIGURE 8

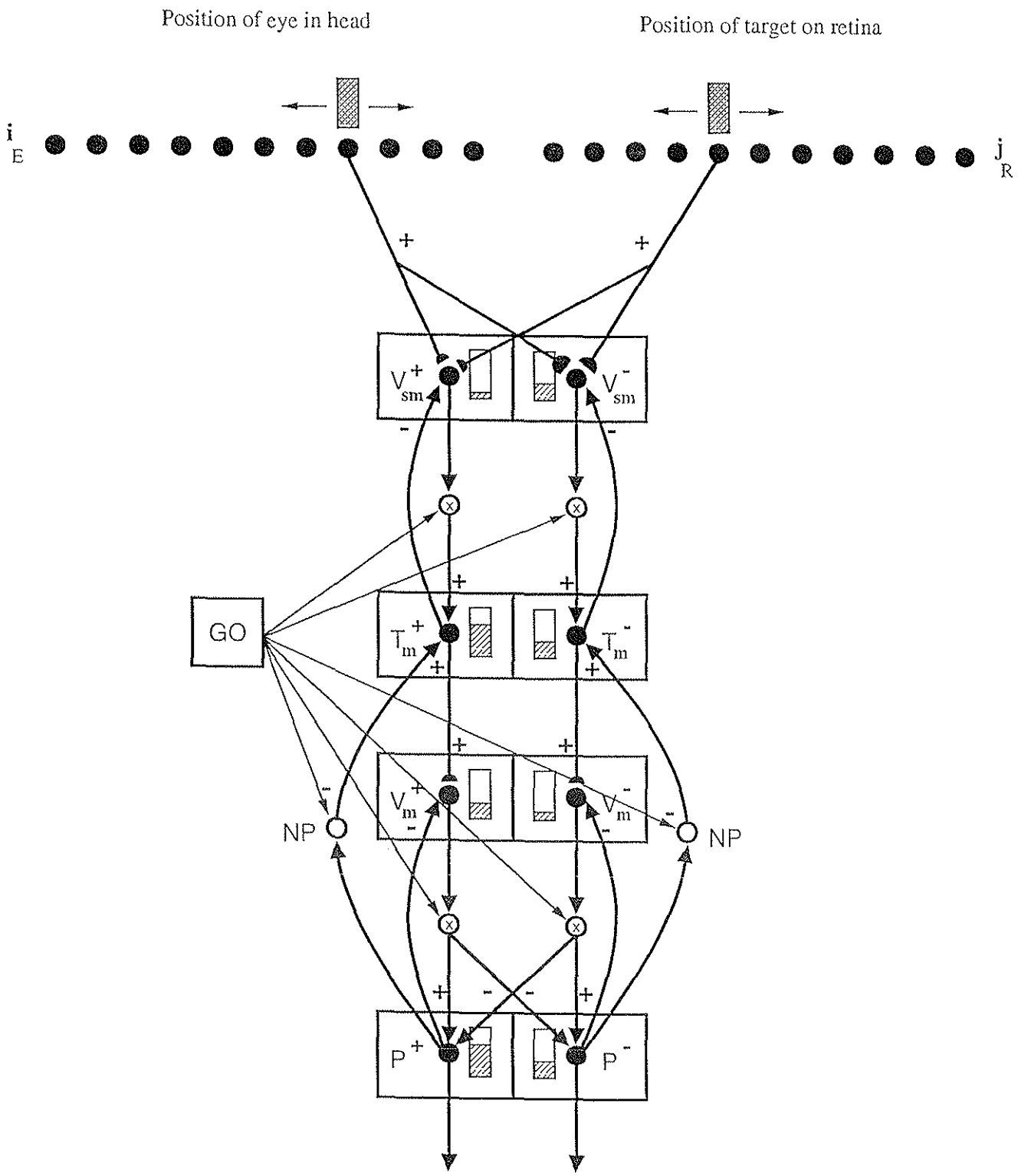

Position of eye in head
Position of target on retina
i
E
j
R
GO
NP
NP

FIGURE 9

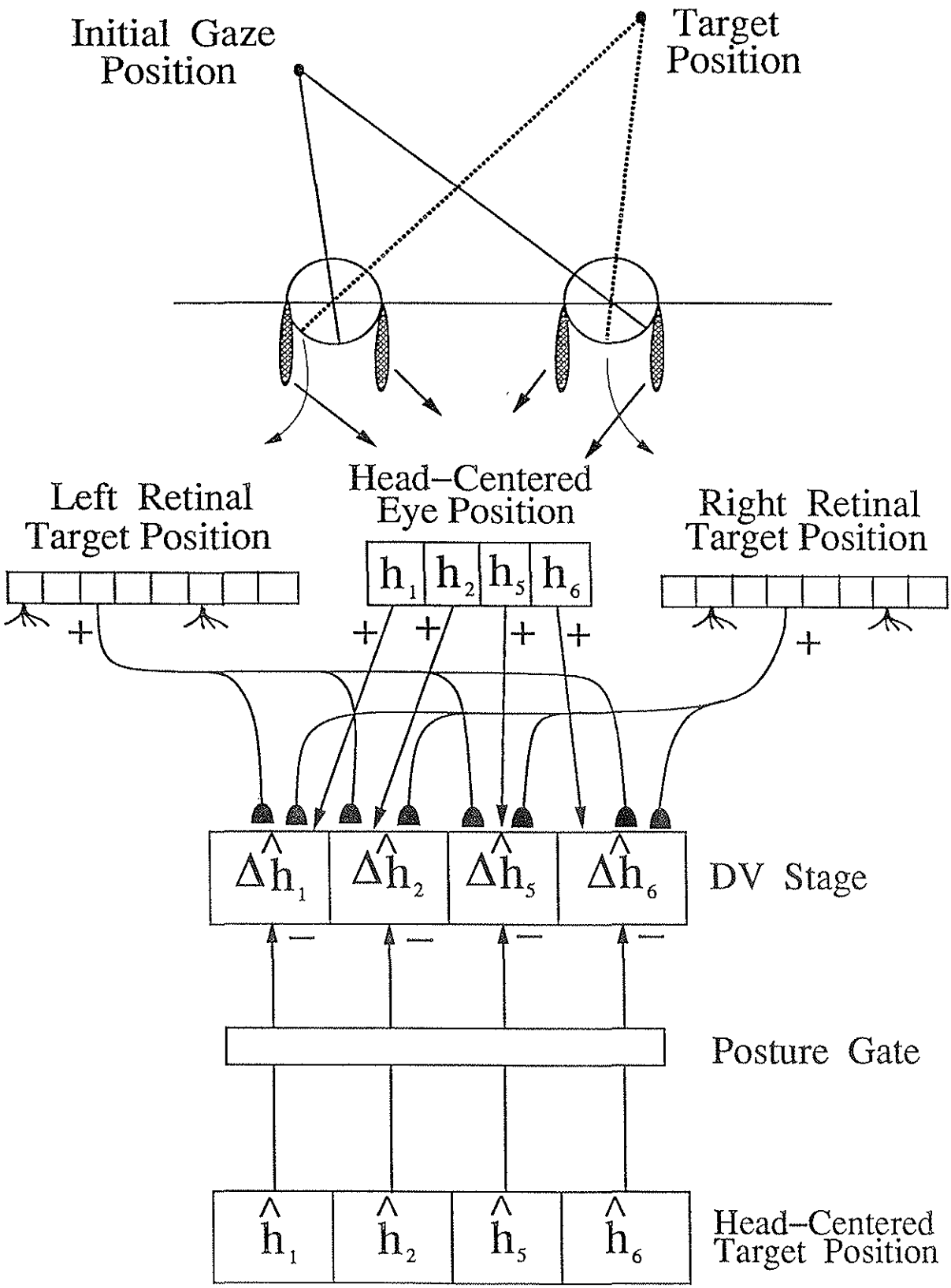

Initial Gaze Position
Target Position
Head–Centered Eye Position
Left Retinal Target Position
Right Retinal Target Position
h_1 h_2 h_5 h_6
+
$\Delta\hat{h}_1$ $\Delta\hat{h}_2$ $\Delta\hat{h}_5$ $\Delta\hat{h}_6$
DV Stage
Posture Gate
$\hat{h}_1$ $\hat{h}_2$ $\hat{h}_5$ $\hat{h}_6$
Head–Centered Target Position

FIGURE 10

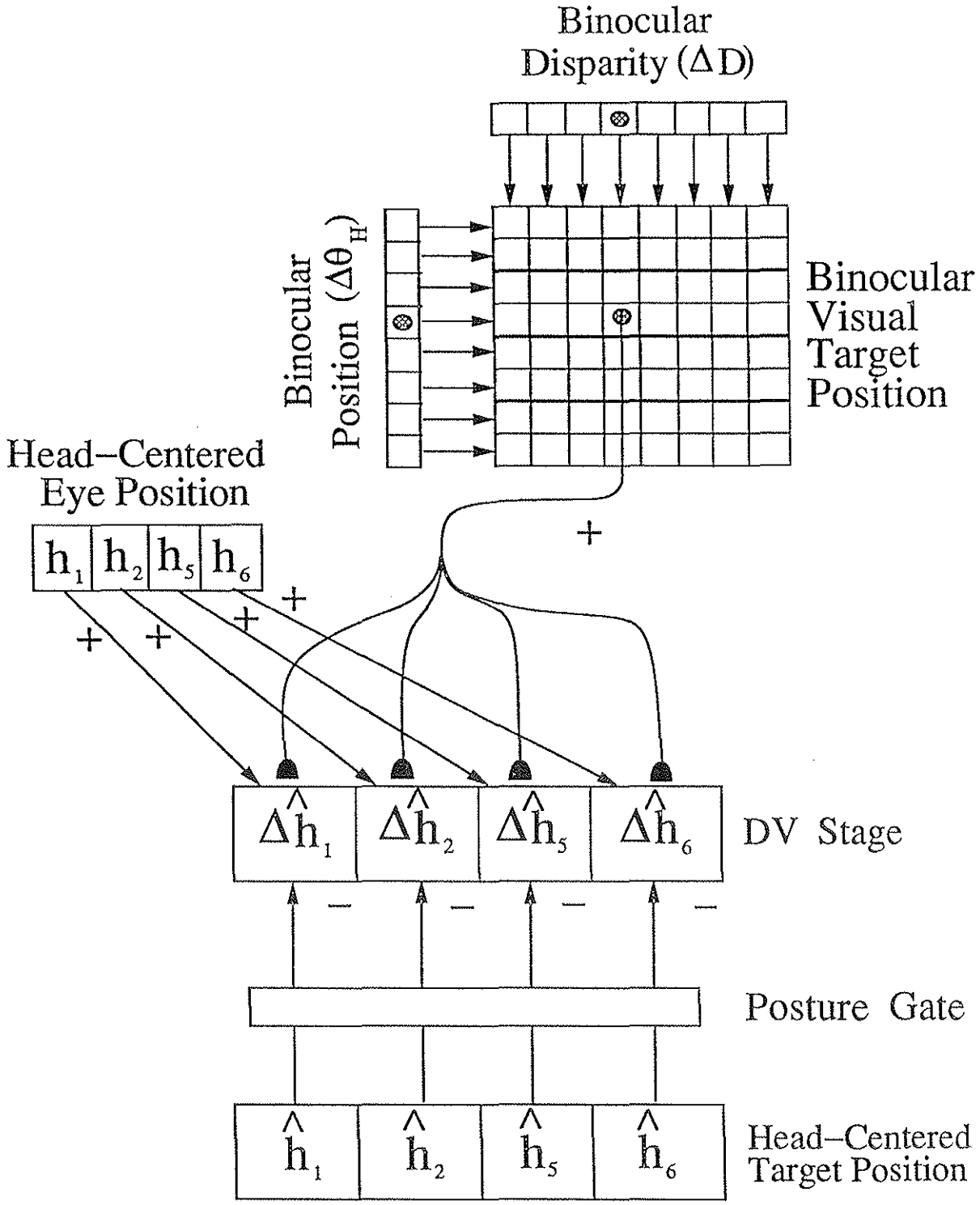

Binocular
Disparity (ΔD)
Binocular
Position ($\Delta\theta_H$)
Binocular
Visual
Target
Position
Head–Centered
Eye Position
h_1 h_2 h_5 h_6
+
$\Delta\hat{h}_1$ $\Delta\hat{h}_2$ $\Delta\hat{h}_5$ $\Delta\hat{h}_6$
DV Stage
−
Posture Gate
$\hat{h}_1$ $\hat{h}_2$ $\hat{h}_5$ $\hat{h}_6$
Head–Centered
Target Position

FIGURE 11

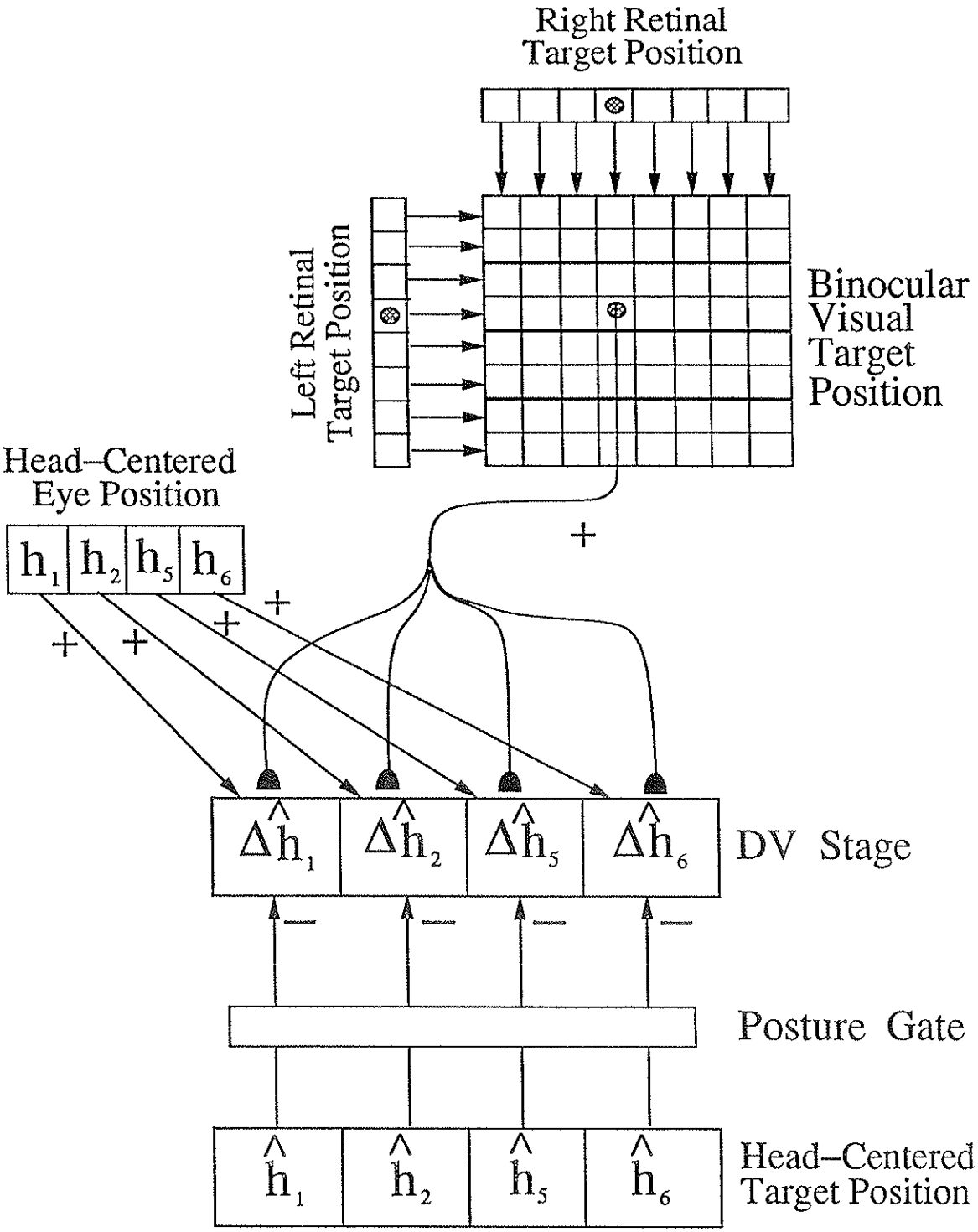

Right Retinal
Target Position
Left Retinal
Target Position
Binocular
Visual
Target
Position
Head–Centered
Eye Position
h_1 h_2 h_5 h_6
+
+
+
+
+
$\Delta\hat{h}_1$ $\Delta\hat{h}_2$ $\Delta\hat{h}_5$ $\Delta\hat{h}_6$
DV Stage
–
–
–
–
Posture Gate
$\hat{h}_1$ $\hat{h}_2$ $\hat{h}_5$ $\hat{h}_6$
Head–Centered
Target Position

FIGURE 12

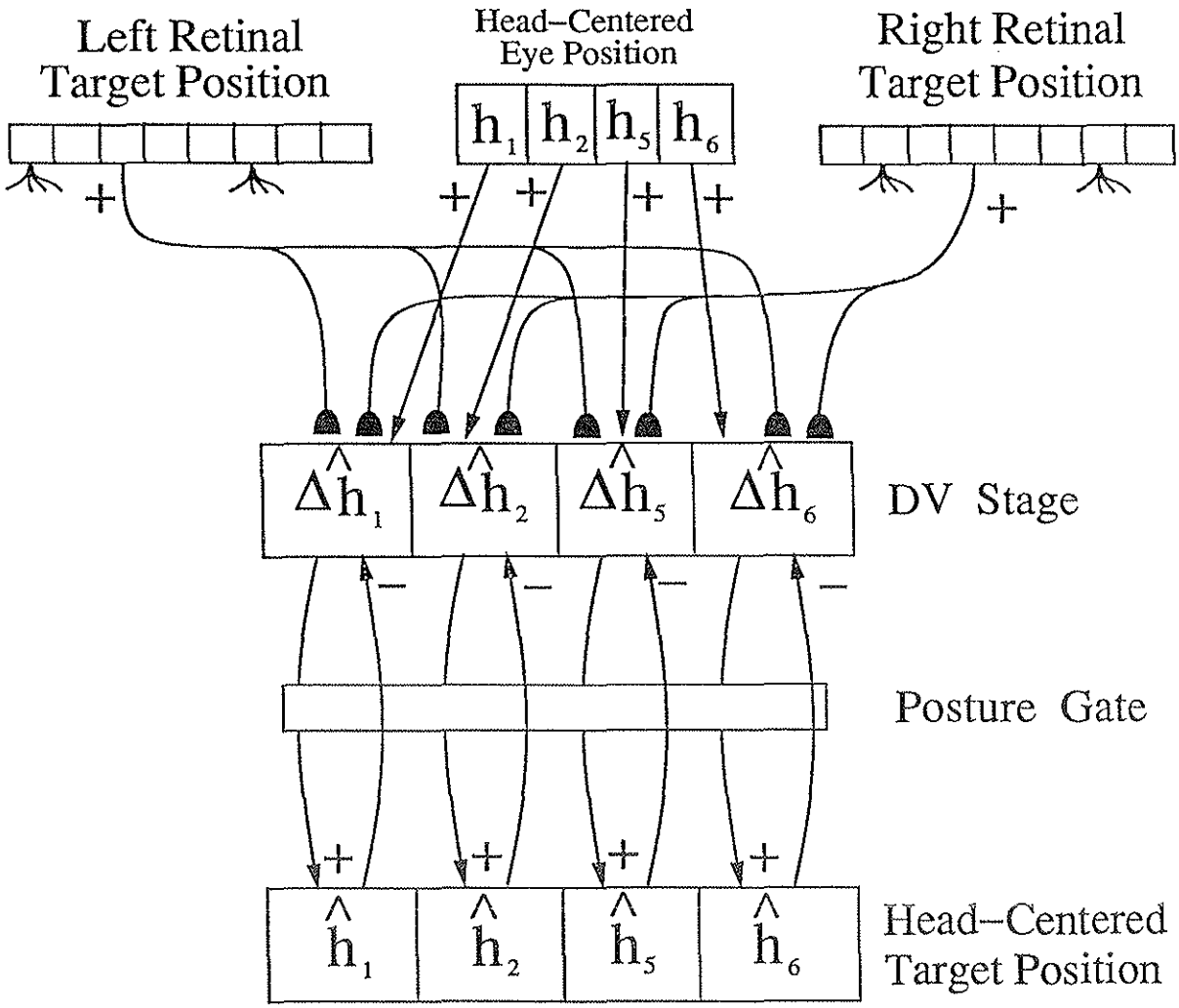

Left Retinal
Target Position
Head–Centered
Eye Position
Right Retinal
Target Position
h_1 h_2 h_5 h_6
$\Delta\hat{h}_1$ $\Delta\hat{h}_2$ $\Delta\hat{h}_5$ $\Delta\hat{h}_6$
DV Stage
Posture Gate
$\hat{h}_1$ $\hat{h}_2$ $\hat{h}_5$ $\hat{h}_6$
Head–Centered
Target Position

FIGURE 13

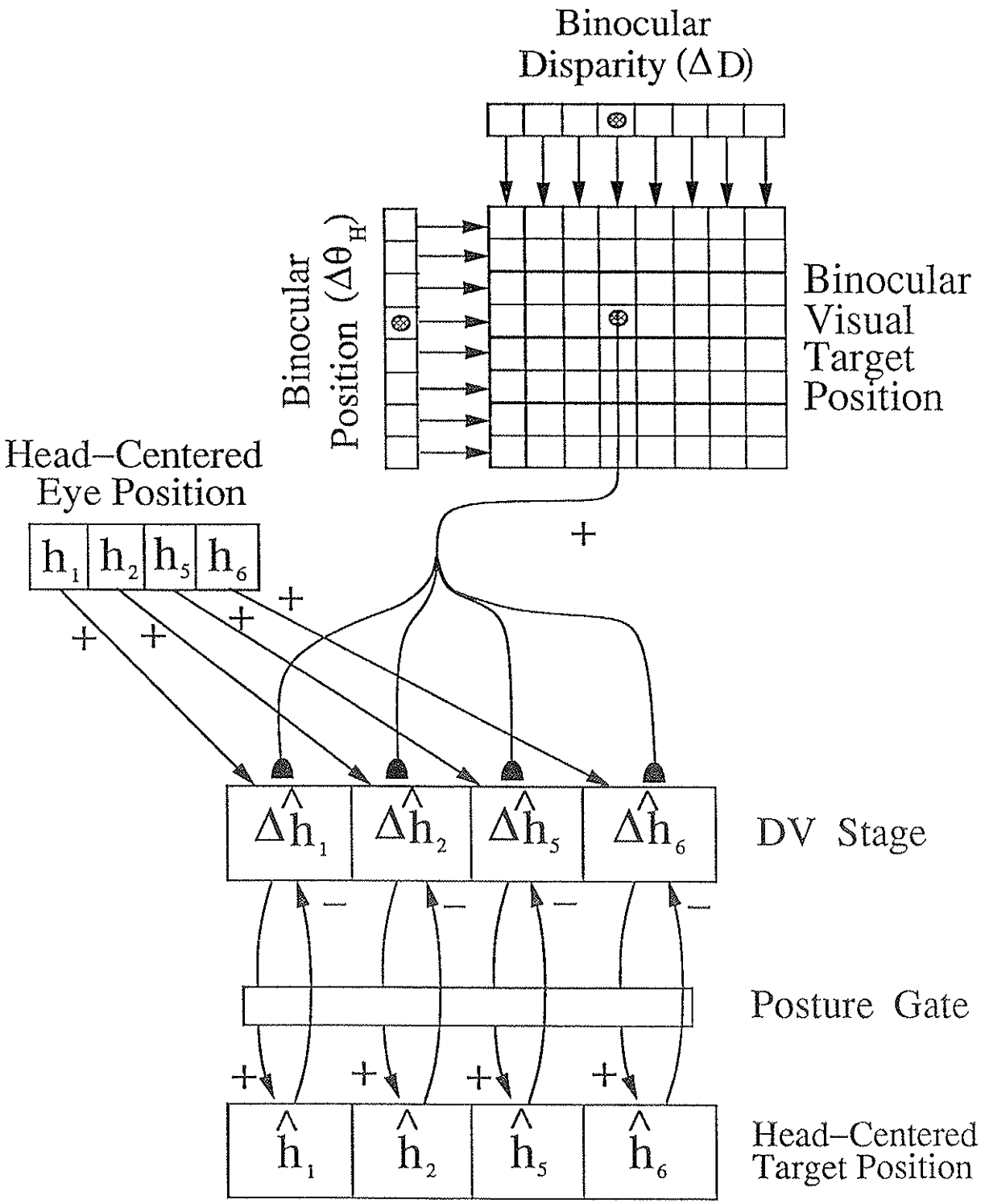

Binocular
Disparity (ΔD)
Binocular
Position (Δθ_H)
Binocular
Visual
Target
Position
Head–Centered
Eye Position
h1 h2 h5 h6
+
+
+
+
+
+
Δĥ1 Δĥ2 Δĥ5 Δĥ6
DV Stage
–
–
–
–
Posture Gate
+
+
+
+
ĥ1 ĥ2 ĥ5 ĥ6
Head–Centered
Target Position

Figure 14

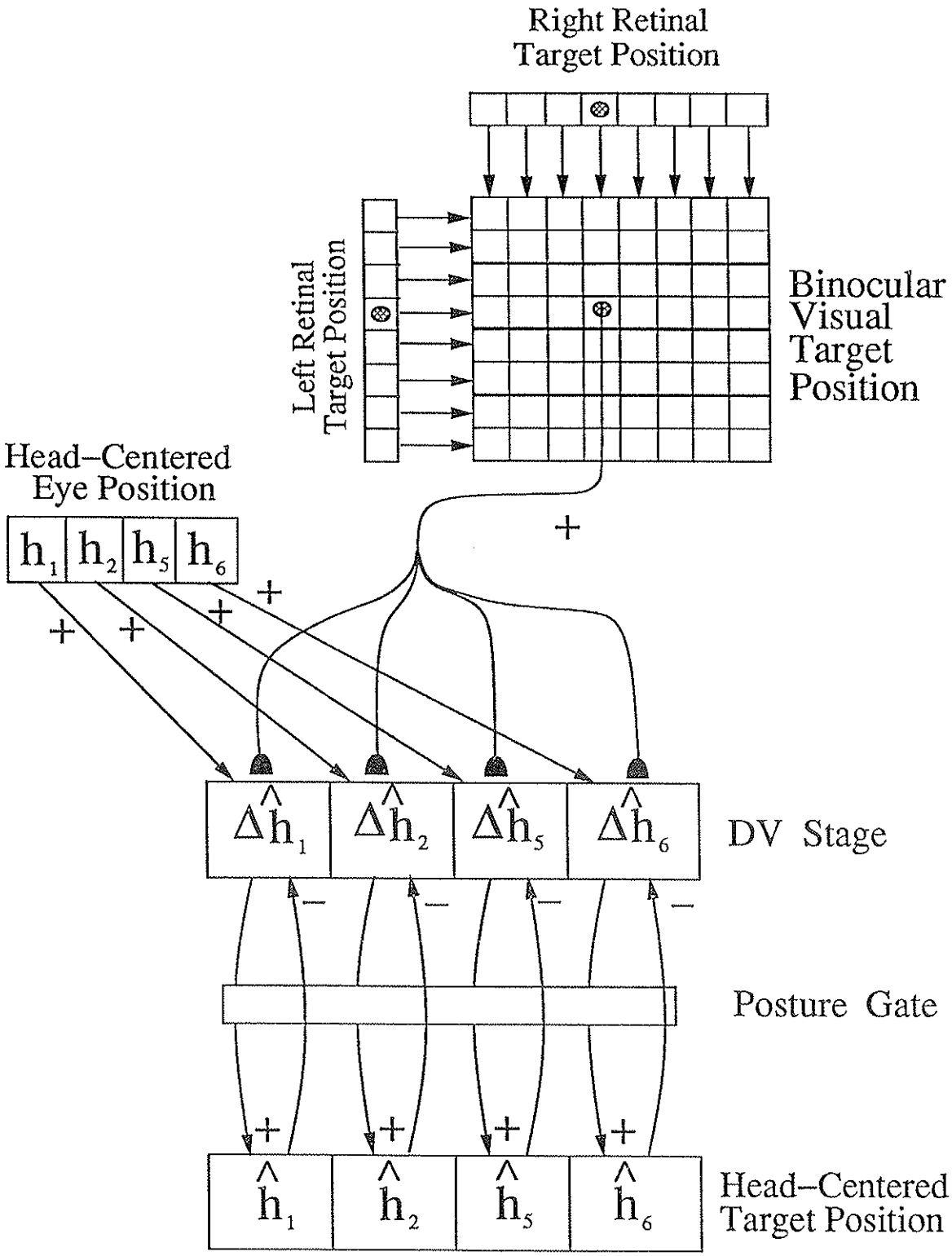

Right Retinal
Target Position
Left Retinal
Target Position
Binocular
Visual
Target
Position
Head–Centered
Eye Position
h_1 h_2 h_5 h_6
+
$\Delta\hat{h}_1$ $\Delta\hat{h}_2$ $\Delta\hat{h}_5$ $\Delta\hat{h}_6$
DV Stage
–
Posture Gate
$\hat{h}_1$ $\hat{h}_2$ $\hat{h}_5$ $\hat{h}_6$
Head–Centered
Target Position

FIGURE 15

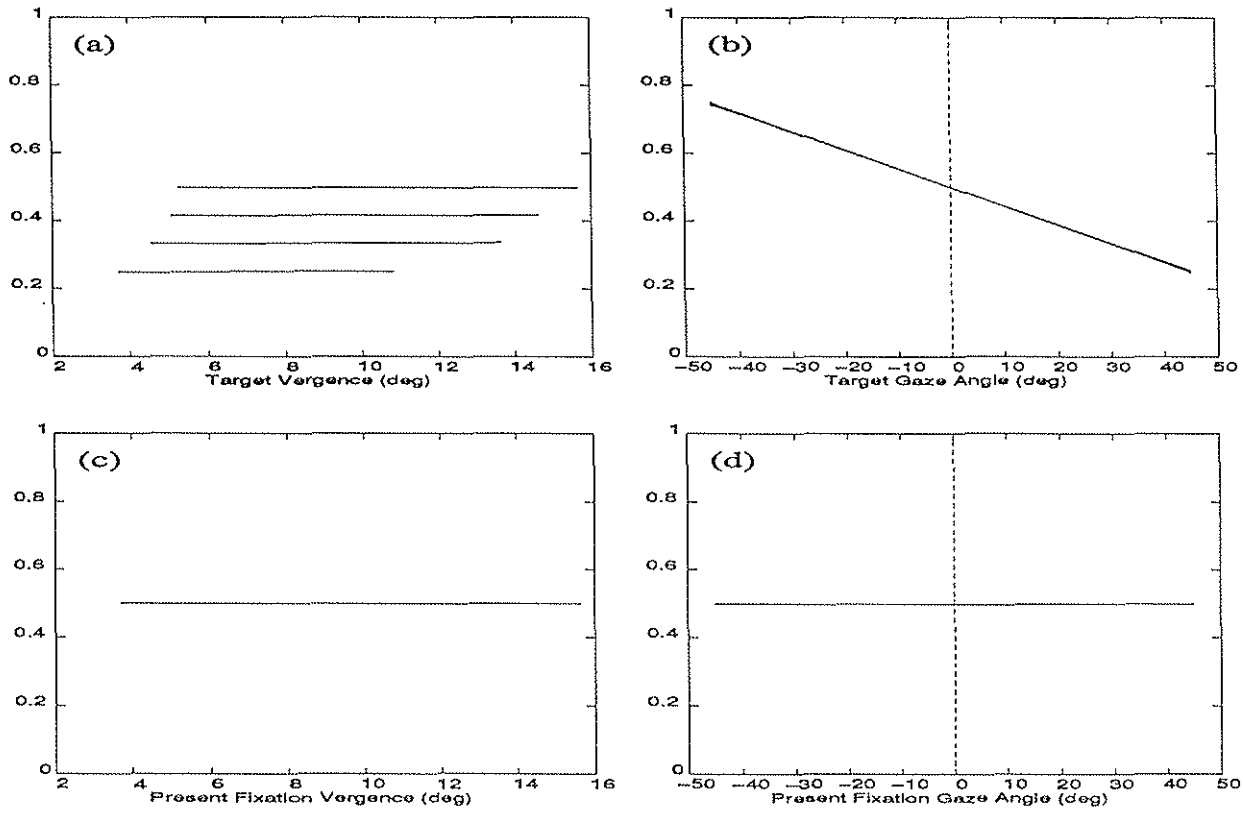

(a)
Target Vergence (deg)
(b)
Target Gaze Angle (deg)
(c)
Present Fixation Vergence (deg)
(d)
Present Fixation Gaze Angle (deg)

FIGURE 16

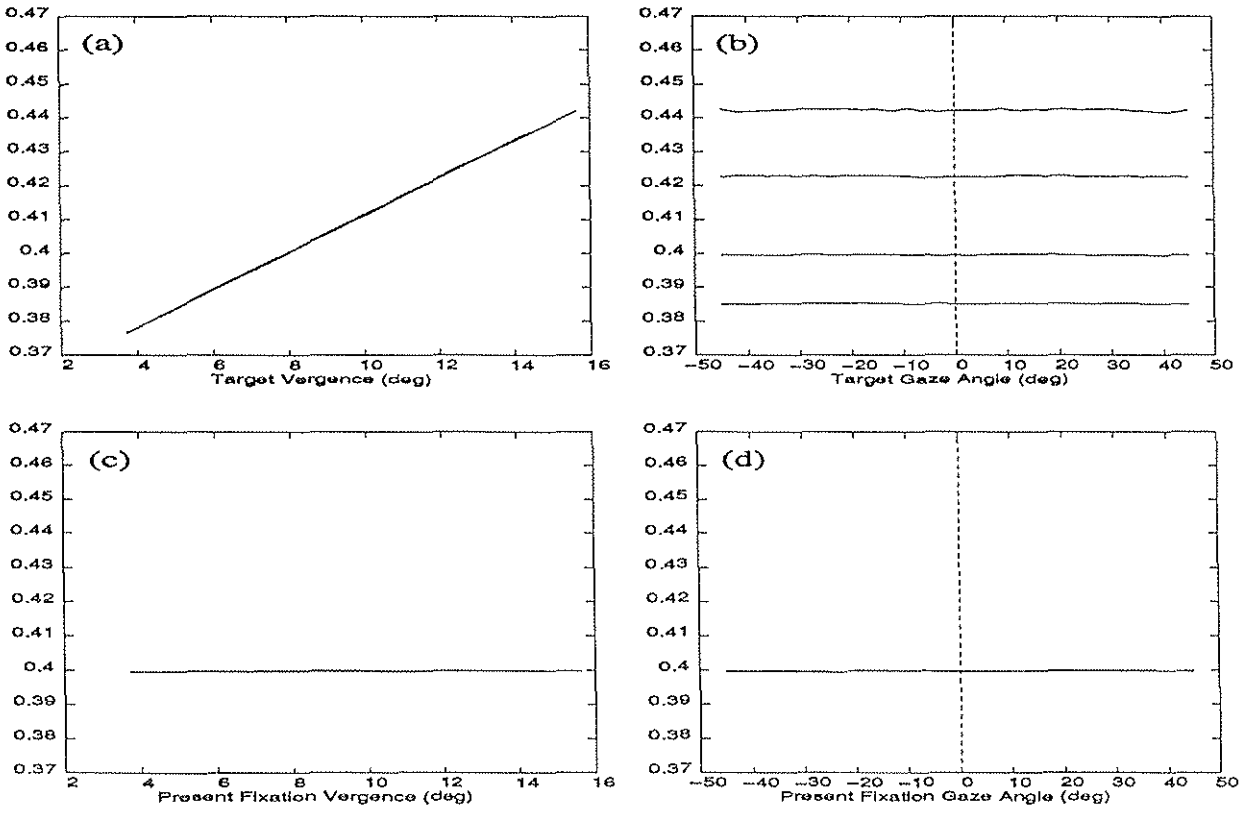

(a)
Target Vergence (deg)
(b)
Target Gaze Angle (deg)
(c)
Present Fixation Vergence (deg)
(d)
Present Fixation Gaze Angle (deg)

FIGURE 17

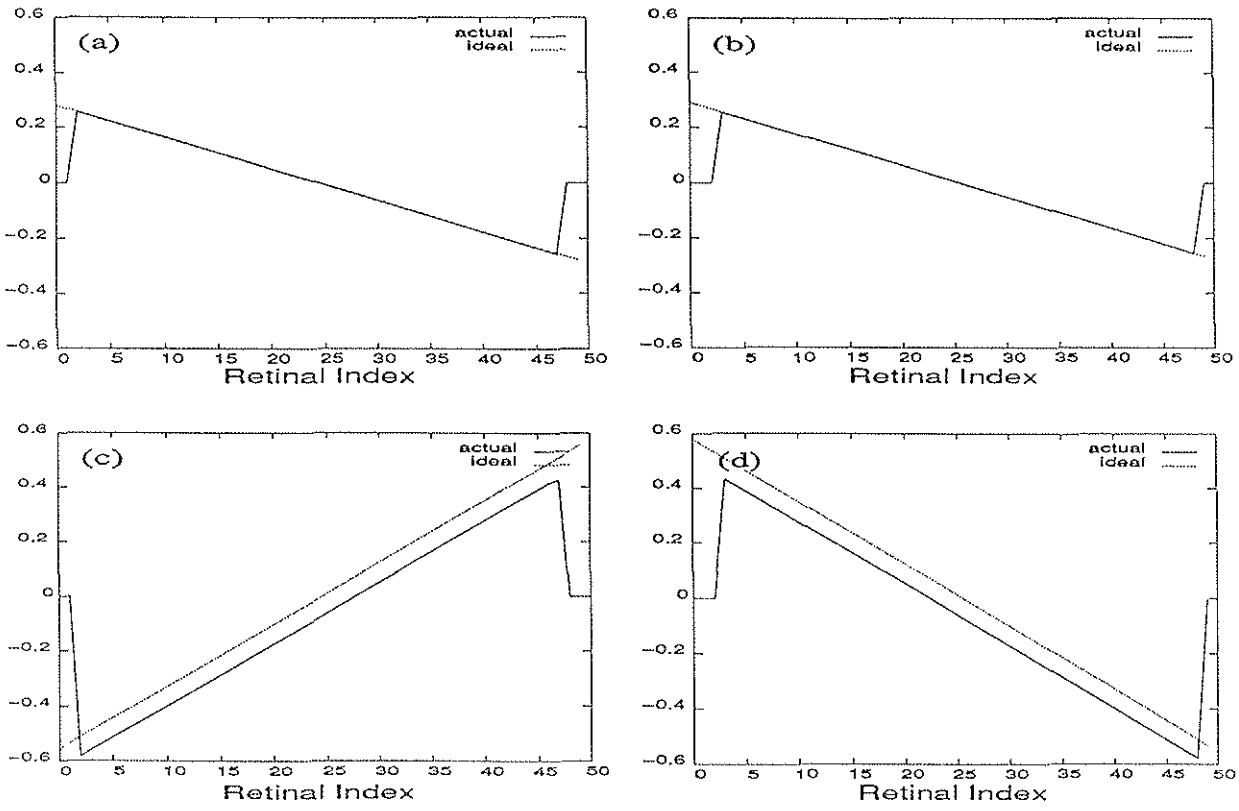
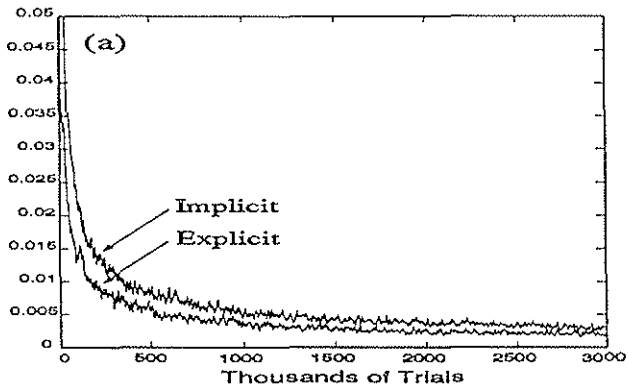

(a)
actual
ideal
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
0 5 10 15 20 25 30 35 40 45 50
Retinal Index
(b)
actual
ideal
Retinal Index
(c)
actual
ideal
Retinal Index
(d)
actual
ideal
Retinal Index

Figure 18


(a)
Implicit
Explicit
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0
500
1000
1500
2000
2500
3000
Thousands of Trials

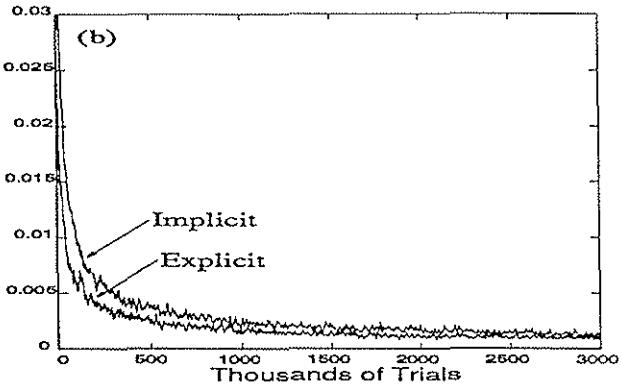

(b)
Implicit
Explicit
0.03
0.025
0.02
0.015
0.01
0.005
0
0
500
1000
1500
2000
2500
3000
Thousands of Trials

FIGURE 19

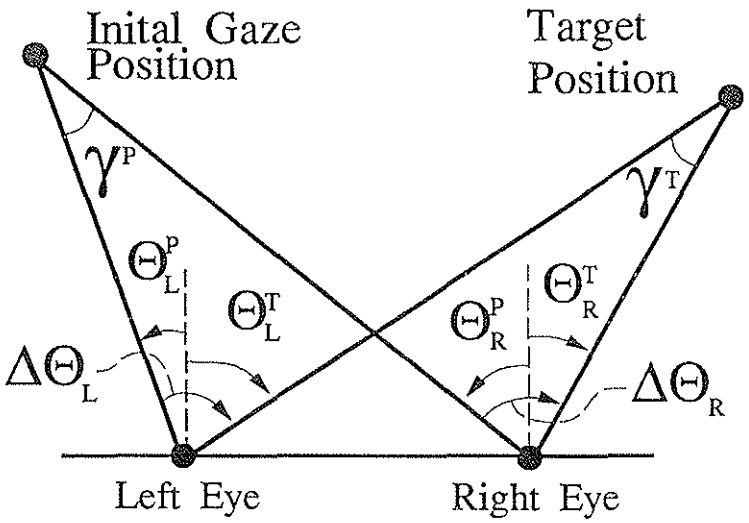

Inital Gaze Position
Target Position
γ^P
γ_T
Θ_L^P
Θ_L^T
Θ_R^T
Θ_R^P
$\Delta\Theta_L$
$\Delta\Theta_R$
Left Eye
Right Eye

FIGURE 20