Available online at www.sciencedirect.com
SCIENCE DIRECT®
PERGAMON
Vision Research 43 (2003) 2155–2171
Vision Research
www.elsevier.com/locate/visres

Self-organizing neural network model of motion processing in the visual cortex during smooth pursuit

Moran Furman, Moshe Gur *

Department of Biomedical Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel

Received 31 July 2002; received in revised form 10 January 2003

Abstract

A physiologically based neural network model was constructed to study cortical motion processing during pursuit eye movements. The model consists of three layers of computational units, simulating information processing by direction selective neurons in the primary visual cortex (V1), motion selective neurons in the middle-temporal area, and pursuit selective neurons in the middle-superior-temporal (MST) area. MST units integrate visual and eye-movement related information, and their connections develop during an unsupervised training process. The resulting MST units represent a transition from retinal to real-world reference frame. By analyzing the model connectivity, mechanisms underlying the functions performed by the network are studied.

Keywords: Pursuit eye movement; Area MST; Unsupervised learning; Neural network; Motion processing

1. Introduction

Detecting objects' motion is of crucial importance to adaptive behavior. One of the basic problems the visual system faces in this context is that the eyes themselves are constantly moving so that retinal image movements reflect both objects and eye movements.

Of the various types of eye movements, we concentrate here on smooth pursuit eye movements which serve to keep the image of a moving object on the fovea, where visual acuity is highest. During pursuit of a moving target, the image of the target on the retina is nearly stable, while the image of the background is moving. The visual system is able to "correct" the visual information coming from the eyes by compensating for eye movements such that during pursuit the environment is perceived as nearly stable and the object as moving. Early theoretical studies of eye-movement compensation (Gregory, 1958; Von Helmholtz, 1866; Von Holst, 1954), suggested a subtraction of extraretinal information about eye velocity from retinal information about target velocity.

Many physiological studies have sought to relate smooth pursuit to properties of motion selective neurons in the visual cortex. The basic elements of the cortical motion processing stream in primates are V1 direction selective neurons (Dow, 1974; Hubel & Wiesel, 1968). These neurons project both directly and through areas V2 and V3 to the middle-temporal area (MT) which is specialized for processing visual motion (Albright, 1984; Movshon & Newsome, 1984). Area MT, in turn, projects to the middle-superior-temporal area (MST, or V5a) and to the visual motor areas of the parietal lobe.

Two subdivisions in MST were observed (Berezovskii & Born, 2000; Komatsu & Wurtz, 1988a; Tanaka, Sugita, Moriya, & Saito, 1993). The dorsal part (MSTd) contains mainly cells which respond best to a large-field motion, while the ventral (or lateral) subdivision (MSTv, MSTl) responds best to small moving targets. Neurons responding during pursuit eye movements were found in the foveal region of MT (MTf) and in both subdivisions of MST (Erickson & Thier, 1991; Ferrera & Lisberger, 1997; Kawano, Shidara, Watanabe, & Yamane, 1994; Komatsu & Wurtz, 1988a; Squatrito & Maioli, 1997).

Newsome, Wurtz, and Komatsu (1988), by briefly blinking off the visual target or by stabilizing the target on the retina, showed that while MT pursuit cells response depended on the retinal movement of the target (retinal slip), some of the pursuit cells in MST continued

* Corresponding author. Tel.: +972-4-8294116; fax: +972-4-8234131.
E-mail address: mogi@biomed.technion.ac.il (M. Gur).

0042-6989/03/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0042-6989(03)00338-9

to respond during image stabilization, indicating that their response depended on an extraretinal input. For most pursuit cells, the preferred direction of movement for a relatively small target during fixation was the same as the preferred direction of pursuit (Komatsu & Wurtz, 1988b). An fMRI study by Dukelow et al. (2001) showed that part of the human homologue of the macaque motion complex, MT+, also seems to receive an extraretinal pursuit related signal.

Theoretical models studied possible roles of MST neurons in both perceptual and execution aspects of pursuit. Dicke and Thier (1999) modeled the role of MST in pursuit generation and maintenance. The model units responded to retinal image slip as well as to eye and head velocity with similar preferred directions and the authors suggested that such neurons are able to reconstruct target motion in world-centered coordinates, and to account for salient properties of visually guided pursuit. Pack, Grossberg, and Mingolla (2001) proposed a neural model dealing with pursuit related cells in the ventral and dorsal subdivisions of MST. The model explained how signals representing target velocity, eye-movement velocity, and retinal background motion, can be combined to explain behavioral data about pursuit maintenance and perceptual aspects of pursuit.

The present work focuses on the interaction between visual and extraretinal signals in MST neurons. We studied whether integration of these signals can result from an unsupervised training process of the connections to MST units. While previous models assumed a predetermined connectivity to MST units, here we show that an unsupervised training process resulted in MST units responding consistently to object motion regardless of whether the eyes were moving or stationary. For this purpose we constructed a physiologically based neural network model, simulating information processing in relevant cortical areas. The model studied, for the first time, integration of visual and extra-retinal signals in MST units in a broad context including full direction and velocity representations in cortical areas V1 and MT. By analyzing the connectivity to MST model units after training, the mechanisms underlying the functions performed by the network can be studied, thus helping to fill the gap between single-cell and system levels. A brief report of these results has appeared previously (Furman & Gur, 1999).

2. Architecture of the neural network

2.1. General structure of the network

The model was a feed-forward neural network, with three layers of computational units, simulating direction selective V1 neurons, MT neurons, and MST pursuit-selective neurons. Fig. 1 depicts the general structure of the model. The layer representing MST, received, in addition to visual information, an extra-retinal input that was a copy of the eye-movement motor signal. The retinal and extra-retinal inputs to the network are described in Section 2.2. Connections between every two network layers were separated into excitatory and inhibitory ones. The activity of each model unit was an analog, non-linear threshold function of its total input (see Appendix A.1 for details). Connections to model layers V1 and MT were predetermined to obtain specific functions (see Sections 2.3 and 2.4). Connections to layer MST developed via an unsupervised learning process (see Section 2.5). The model was implemented in Matlab, using the Neural-Network Toolbox.

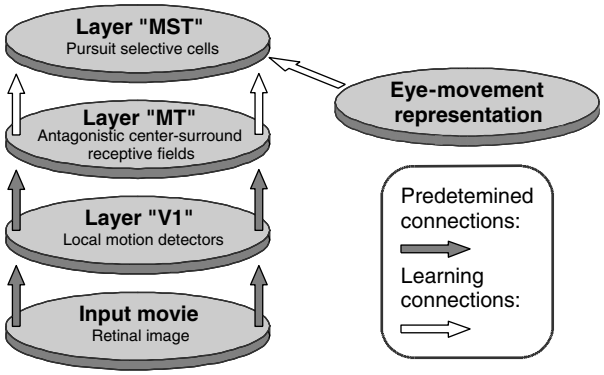

Fig. 1. General structure of the network. The input movie, representing the dynamic retinal image, is processed first by V1 units selective to local movements in a specific direction and velocity within their receptive-field (RF). The units in the next processing layer, representing MT, have antagonistic center-surround RFs. The third layer, MST, receives, in addition to the visual input arriving through MT, an extra-retinal input representing eye-movement direction and velocity.

2.2. Input—retinal movie and eye-movement representation

The input to V1 was a movie representing the retinal image. Input images were binary (each pixel was either bright or dark), with hexagonal topology consisting of 397 pixels. One pixel in the retinal image represented roughly 0.5° of the visual field, and the whole input movie covered about $12° \times 12°$ of the visual field. There were six possible directions of movement, determined by the orientation of the pixels in the retinal image, 0°, 60°, ..., 300°, where directions of movement are indicated counter-clockwise in relation to the rightward horizontal direction.

The input movies depicted an object moving either in the dark or against a textured background. The distinction between object and background was implicit by the pixels' relative movement in the series of images. Three types of input movies were used. The first simulated the retinal image during fixation. The second represented the retinal image during stabilized pursuit, where the image of the pursued target is stabilized on the

retina. The third represented the retinal image during normal pursuit. The latter was constructed from periods of continuous movement, representing smooth pursuit, separated by quick saccades. During smooth pursuit the eye's velocity was somewhat slower than the target's, causing the target's retinal image to move slowly at the same direction as the eye (retinal slip). Saccades brought back the image of the target to the center of the retinal field. The background moved at the eye velocity but in the opposite direction.

The third layer in our model, representing MST, received, in addition to visual input, an input representing eye movements. Since the exact physiological nature of this input is not known, we chose to represent eye movements by a population-vector coding, which is found in various neuronal mechanisms subserving directional variables (Georgopoulos, Taira, & Lukashin, 1993; Harris & Jenkin, 1997; Zemel, Dayan, & Pouget, 1998; Zohary, 1992). Eye movements were implemented by the activity of a set of 24 units (six preferred directions of eye movements and four preferred velocities), which can be interpreted as the neural structure generating pursuit eye movements and sending a copy of the motor commands to MST. The preferred directions were similar to the possible directions of movement in the input movie ($0°$, $60°, \ldots, 300°$), and the preferred velocities (0.5°/s, 2°/s, 8°/s, 32°/s) covered, in a log scale, most of the range of possible pursuit speed in primates (see Appendix A.2 for details).

2.3. Layer V1—modeling local motion detectors

The units in the first processing layer simulated direction selective cells in the primary visual cortex (Fig. 2). To model local motion detection we followed the delayed inhibition approach (e.g., Amthor & Grzywacz, 1991; Barlow & Levick, 1965; Borg-Graham & Grzywacz, 1992), which was shown to be one of the major determinants of directionality in the monkey primary visual cortex (Livingstone, 1998; Sato, Katsuyama, Tamura, Hata, & Tsumoto, 1995). The details of the local motion detection mechanism are not of major importance for the present model, and will be described only in general terms (see Appendix A.3 for details). To generate direction selectivity, the receptive-field (RF) of each V1 unit was divided along its long axis to two subfields, one excitatory, and the other inhibitory acting after a time delay. Movement in the preferred direction activated first the excitatory subfield, rendering the delayed inhibition ineffective. Movement in the opposite direction activated first the inhibitory subfield such that the delayed-inhibition coincided with the excitation generated by the stimulus crossing the excitatory subfield, resulting in a weak or no response.

The RFs of the V1 units used in our simulations were elongated, 5 pixels long and 3 pixels wide (corresponding to approximately $2.5° \times 1.5°$ of the visual field) with preferred movement directions perpendicular to their long axis. The orientation of the pixels in the retinal image gave rise to six preferred directions: $30°$, $90°, \ldots, 330°$. V1 units were retinotopically organized, with partial overlap between adjacent units. The visual field was represented by 271 retinotopic positions. Six preferred directions of movement and four preferred velocities (0.5°/s, 2°/s, 8°/s, 32°/s) were represented at each visual field location (see Fig. 2). In total, the V1 layer contained 6504 units.

2.4. Layer MT—modeling antagonistic center-surround receptive fields

Direction selective cells in V1 are the major input to cortical area MT (Albright, 1984; Maunsell & Van Essen, 1983; Zeki, 1974), where most neurons have a

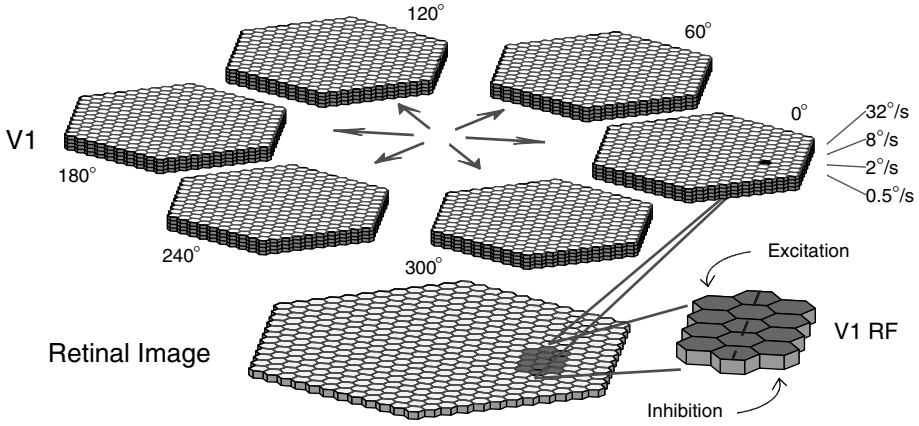

Fig. 2. Organization of the V1 layer and a schematic RF structure of a V1 unit. The retinal image (bottom layer) has a hexagonal topology. The shaded area shows the RF of a single V1 unit, which is elongated, and divided along its long axis to excitatory and inhibitory subfields. V1 units (top layer) were separated into six groups according to their preferred direction of movement, as inidicated. Each group covers the whole visual field and within each group the units are retinotopically organized. At each combination of preferred-direction and retinotopic position, there are four V1 units, having different preferred velocities. RFs of adjacent V1 units are partially overlapping.

center-surround RF organization (Allman, Miezin, & McGuinness, 1985; Born & Tootell, 1992; Raiguel, Van Hulle, Xiao, Marcar, & Orban, 1995; Tanaka et al., 1986; Xiao, Raiguel, Marcar, Koenderink, & Orban, 1995). The units in the second processing level in our network simulated neurons in the foveal representation of MT. Several authors suggested that direction selective surround inhibition results from direction-dependent interactions between motion detectors at different spatial locations (Kim & Wilson, 1997; Liu & Van Hulle, 1998; Raiguel et al., 1995; Tanaka, 1998). Our modeling of MT units was based on a similar principle, implemented by constructing appropriate connections between V1 and MT units. The details of the mechanism underlying MT model units can be found in Appendix A.4.

MT units RFs (including the inhibitory surround) size was the same as that of the input movies retinal image ($12° \times 12°$). The MT layer contained 96 units, with six preferred directions of movement (0°, 60°, ...,300°), four preferred velocities (0.5°/s, 2°/s, 8°/s, 32°/s), and four sizes of RF center (3, 5, 7 and 9 pixel diameter). The velocity selectivity of the model units was based on physiological findings by Lagae, Raiguel, and Orban (1993), which showed that low-pass velocity tuned units (0.5°/s and 2°/s in our model) dominated MT foveal region, with units tuned to mid level velocities (8°/s in our model) and high velocities (32°/s) making up the rest. Fig. 3 depicts the organization of the MT layer and its relation to V1 topology.

2.5. Layer MST

One of the major target areas of cortical area MT is the MST area (Desimone & Ungerleider, 1986; Maunsell & Van Essen, 1983) where cells discharging during smooth pursuit were found (Ferrera & Lisberger, 1997; Kawano et al., 1994; Komatsu & Wurtz, 1988a; Squatrito & Maioli, 1997). The third layer in our network contained 60 units and was aimed to model MST pursuit selective cells. MST units received both a visual input from MT units (Section 2.4) and an input representing eye movements (Section 2.2). Fig. 4 illustrates the organization of the MST layer.

Connections between MT and MST model layers were not predetermined, but developed via an unsupervised learning process, as were the connections between the eye-movement representation units and MST. At the beginning of the training period all connection weights were relatively weak and random, therefore MST units did not perform any specific function. At each training step, weights were updated, according to the network units' activities. Different training rules were used for excitatory and inhibitory connections ("synapses"), reflecting the different functionalities of these synapses (see Section 4). For the excitatory connections we used Oja's rule (Oja, 1982), while for the inhibitory synapses we constructed a learning rule favoring anti-correlation between the activity of pre- and post-synaptic neurons (see Appendix A.5 for details).

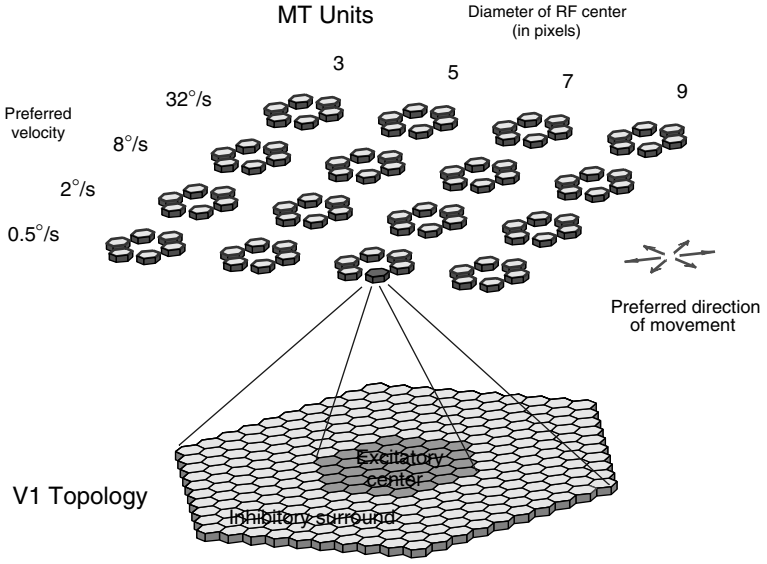

Fig. 3. Organization of the MT layer and a schematic structure of an MT RF. The bottom layer depicts V1 topology, where each hexagon represents 24 V1 units, with the same retinotopic position but different preferred directions or velocities. MT RFs have an antagonistic center-surround structure. The surround of all MT units covers the whole V1 retinotopic map, while the excitatory center (shaded area) covers a relatively small part of the V1 topology. The top layer depicts the 96 MT units, organized according to the RF center preferred direction of movement, preferred velocity, and size.

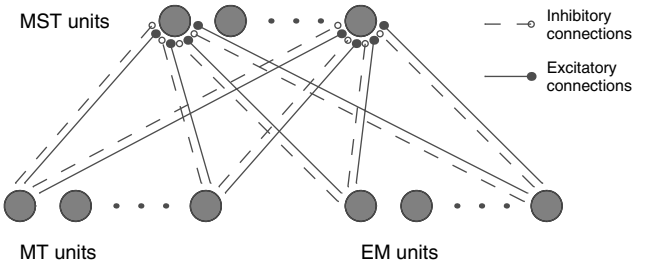

Fig. 4. Schematic structure of the MST layer. MST units receive visual input arriving through MT and an additional input representing eye movements (EM). Each input unit is connected to each MST unit by both an excitatory and an inhibitory connection, whose weights are adapted during network training.

3. Simulation results

3.1. Simulating local motion detection—layer V1

In this section we demonstrate response patterns of layer V1 to two input movies. The first example demonstrates that retinal local movements are properly represented by V1 units. Fig. 5a shows a snapshot of the input movie, depicting a 120° oriented bar moving in the 60° direction, as indicated by the arrow. Fig. 5b shows a corresponding snapshot of the time-varying V1 activity pattern. For simplicity, only V1 units with the same preferred velocity as the moving bar's are shown. The response of the other units is modulated according to the difference between their preferred velocity and the stimulus' one.

V1 activity is presented as follows: V1 units were separated into six groups according to their preferred direction of movement as indicated by the direction of the arrows pointing to them. Each group covers the whole visual field and within each group the units are depicted according to their retinotopic position. For example, all six units at the very center of the six groups have RFs at the same location—the center of the visual field. The brightness of each unit represents its activity level.

As can be seen in Fig. 5b, the most active units in this simulation have a 30° preferred direction of movement. This result demonstrates that V1 units are subject to the "aperture problem" (e.g., Movshon, Adelson, Gizzi, & Newsome, 1986; Sereno, 1993); since their RF is relatively small, the bar's direction of movement measured by them (30°) is perpendicular to the bar's orientation, while its true direction of movement is 60°. As the bar passes along the visual field, the retinotopic locations of the active V1 units shift accordingly.

Next, V1 responses during pursuit eye movements were simulated. It was assumed that the eyes followed an object moving at 8.5°/s to the right against a textured background. Fig. 6 depicts a snapshot of the input movie. During the periods of smooth pursuit, to simulate retinal slip, the retinal image of the object moved at 0.5°/s to the right. The periods of smooth pursuit were interleaved with quick saccades that brought back the image of the object to its original position. The retinal image of the background moved in the direction opposite to the eye movement. In the actual stimuli there was no explicit distinction between pixels belonging to the object or background; the different intensities used in the figure are for clarity only. Fig. 7 shows the simultaneous responses of V1 cells. Each panel represents V1 units with a specific preferred velocity. Within each panel, units are depicted as in Fig. 5b. As can be seen, at the visual field center, some units having a preferred velocity of 0.5°/s and selective to near rightward directions (±30°) responded vigorously (Fig. 7a). This activation resulted from the slow movement of the pursued target on the retina (retinal slip). At the visual field periphery, some of the V1 units having a preferred velocity of 8°/s and selective to near leftward directions ($180° \pm 30°$) were activated, due to the background movement (Fig. 7c).

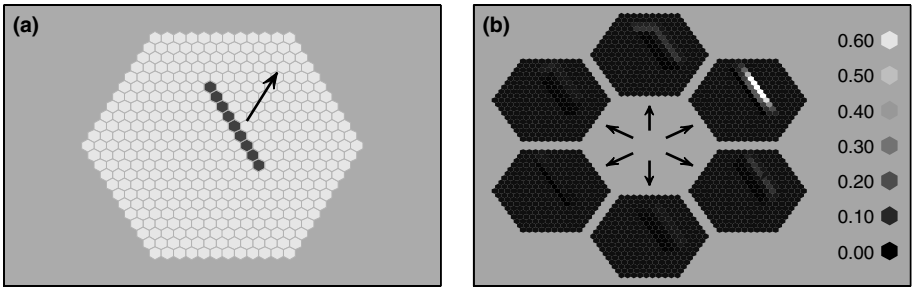

Fig. 5. (a) A snapshot of an input movie, depicting a bar moving in the direction indicated by the arrow and (b) a snapshot of the V1 layer response to the bar depicted in (a). Each pixel represents a V1 unit, its brightness corresponding to its activity level (see scale). For simplicity, only V1 units with the same preferred velocity as the moving bar's are shown. As in Fig. 2, V1 units are separated into six groups, according to their preferred direction of movement. Each group is organized retinotopically, and within a group, all units have the same preferred direction of movement, as indicated by the direction of the arrow pointing to the group.

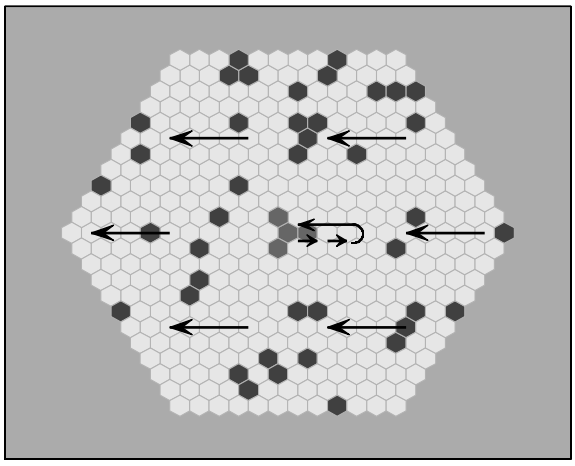

Fig. 6. A snapshot of an input movie depicting the retinal image during pursuit of a target moving at 8.5°/s to the right against a textured background. The movie is constructed from periods representing smooth pursuit, separated by quick saccades. During smooth pursuit the eye's velocity is 8°/s so that the retinal image of the object moves at 0.5°/s to the right (retinal slip). Saccades bring back the image of the target to the center of the retinal field. The retinal image of the background moves in a direction opposite to the eye movement, as indicated.

3.2. Simulating MT cells

V1 activity served as an input to the next processing stage, MT. We first show the responses of an MT unit to different directions of movement in the center and the surround of its RF.

The RF center and surround were stimulated by coherent random dot patterns, moving in the MT unit preferred velocity. Fig. 8a depicts responses to movement in different directions within the RF center. As can be seen, the RF center is directional with a preferred direction of 240°. Fig. 8b shows responses when the RF center was stimulated with movement in the preferred direction, and the surround with movement in different directions. When the movement in the surround was in the same direction as the center's, the cell's response decreased by 79%, while the cell's response was facilitated by 70% by a movement in the RF surround in a direction opposite to the center's. The results shown in Fig. 8 are comparable to the physiological results of Allman et al. (1985).

Next, the response of the MT layer during pursuit is demonstrated. Again, it was assumed that the eyes followed an object moving at 8.5°/s to the right against a textured background (Fig. 6). Fig. 9 shows the responses of the MT units. Each hexagon represents an MT unit, its brightness corresponding to its activity level. The units are organized according to their preferred velocity and RF center size (the RF total extent, including the surround, covered the whole retinal image). For each combination of preferred velocity and RF structure there are six units with different preferred directions.

As can be seen, some of the MT units were active during pursuit. Their response resulted from V1 input (Fig. 7). At the visual field center, some V1 units having a preferred velocity of 0.5°/s and selective to near

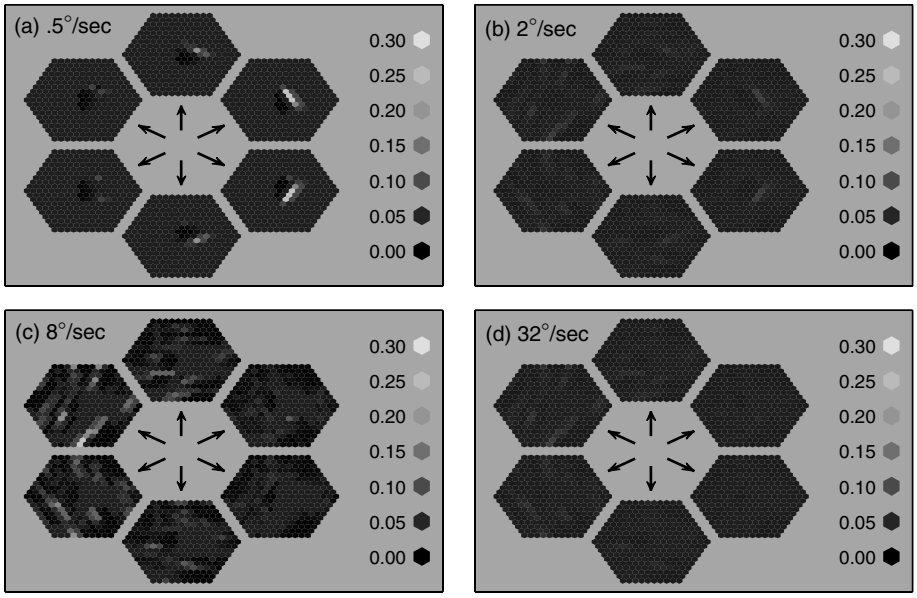

Fig. 7. A snapshot of the V1 layer response during pursuit of a target moving at 8.5°/s to the right (see Fig. 6). Each panel represents V1 units with a specific preferred velocity: (a) 0.5°/s; (b) 2°/s; (c) 8°/s and (d) 32°/s. Each preferred velocity is presented as in Fig. 5b. At the visual field center, some units having a preferred velocity of 0.5°/s and selective to near rightward directions (±30°), respond vigorously. This activation results from the retinal slip of the pursued target. At the visual field periphery, some of the V1 units having a preferred velocity of 8°/s and selective to near leftward directions (180° ± 30°) are activated by background movement.

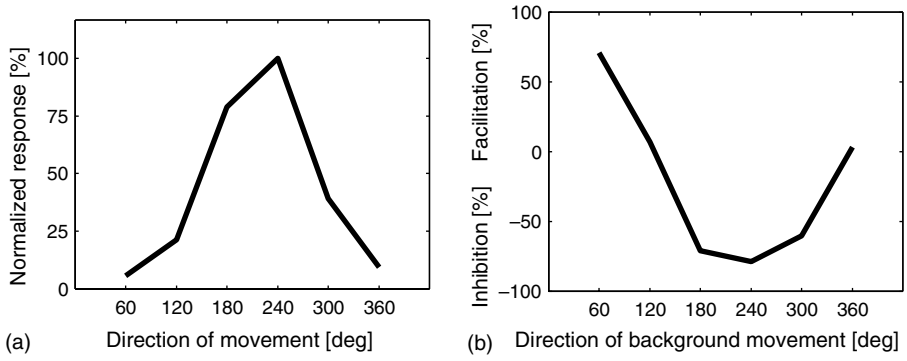

Fig. 8. Responses of an MT model unit to different directions of movement in the RF center and surround during fixation: (a) responses to movement in different directions within the RF center. The response is normalized by its maximum value; (b) response to different directions of movement in the surround while the RF center was simultaneously stimulated by movement in the preferred direction (240°). The relative modulation (inhibition/facilitation) is measured relative to the maximal response in panel (a). Thus, a value of 0% in (b) is equivalent to a response of 100% in (a), while a value of −100% in (b) indicates that movement ins the surround reduced the unit's response to 0.

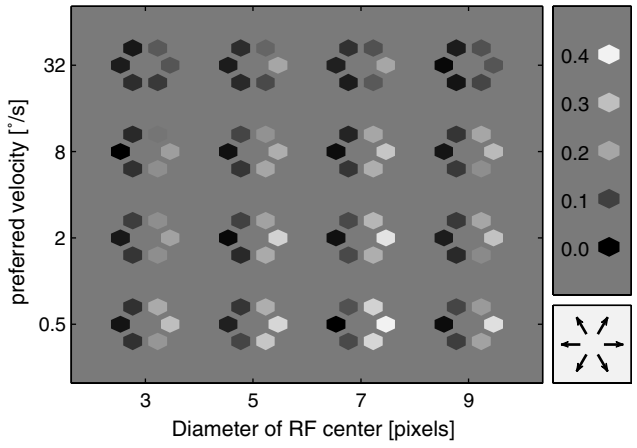

Fig. 9. Simulated responses of MT cells during pursuit of a target moving to the right (see Fig. 6). Each hexagon represents an MT unit, its brightness corresponding to its activity level (see scale). The units are organized according to their preferred velocity (ordinate) and RF center size (abscissa). For each combination of preferred velocity and RF center size there are six units with different preferred directions (as indicated by the arrows diagram). Some of the MT units are active due to an excitatory input generated by the slowly moving target at the center of the retinal field. The retinal movement of the background modulate the response of these MT units.

rightward directions (±30°), responded vigorously, due to the retinal slip of the pursued target (Fig. 7a). The activity of these V1 units stimulated, in turn, the excitatory center of some MT units. The MT units receiving the strongest excitatory input were the ones having a 0° preferred direction of movement, a 0.5°/s preferred velocity and a 5 or 7 pixels diameter RF center (matching the region of activity in the V1 layer due to target movement). However, units with other direction preferences, velocity preferences, and RF sizes were also activated due to retinal slip.

The retinal movement of the background modulated the response of MT units. Fig. 7c showed that at the visual field periphery, background movement activated some of the V1 units having a preferred velocity of 8°/s and selective to near leftward directions ($180° \pm 30°$). This leftward background movement facilitated, by surround disinhibition, the response of MT units selective to movement to the right or near directions. Facilitation was maximal for MT units having a 8°/s preferred velocity and a 5 or 7 pixels diameter RF center, but responses of units with other preferred velocities and RF sizes were also facilitated.

The differences between MT responses (Fig. 9) and V1 responses (Fig. 7) reflects a transition from local to pattern motion representation. At the visual field center, V1 units responded to the slow rightward movement of the target's retinal slip, while at the periphery, V1 units responded to the fast leftward movements of the background. MT units, on the other hand, responded only to the slow rightward movement since V1 units active at the periphery inhibited MT units selective to leftward movements and facilitated units selective to rightward movements.

3.3. Simulation results of the MST layer

3.3.1. Connections development

The network was trained on a set of input patterns representing periods of pursuit eye movements. An input pattern consisted of two elements: (1) a movie representing the retinal-image during pursuit which, prior to reaching MST, was processed by V1 and MT, and (2) the activity of the set of units representing eye-movement direction and velocity.

The connections to MST were chosen from a uniform random distribution on the $[0, 0.01]$ interval; consequently, MST units did not perform any specific function before training.

A pool of input movies was constructed for the network to be trained on. In each, a target, assumed to be followed by a pursuit eye movement, was moving

against a textured background. The target was an object 3 or 5 pixel in diameter (corresponding roughly to 1.5° or 2.5° of the visual field), moving in one of six possible directions (0°, 60°, ..., 300°) and one of two possible velocities (8.5°/s, 34°/s), for a total of 24 input movies. The eye velocity was assumed to be approximately 94% of the target velocity, resulting in 0.5°/s and 2°/s retinal slip for target velocities 8.5°/s and 34°/s, respectively. The network was trained on a sequence of input movies, chosen randomly from this pool.

The total-weight-change-rate during training, $\dot{w}_{\text{total}}$, (see Appendix A.5) was used to assess convergence and terminate the training process. Fig. 10 shows a typical evolution of $\dot{w}_{\text{total}}$. When $\dot{w}_{\text{total}}$ crossed the 10^{-6} level for the second time, training was terminated. In most simulations the training process lasted for, roughly, 1500 presentations of input movies. Each movie lasted 20 time-steps, so total training duration was about 30,000 time steps.

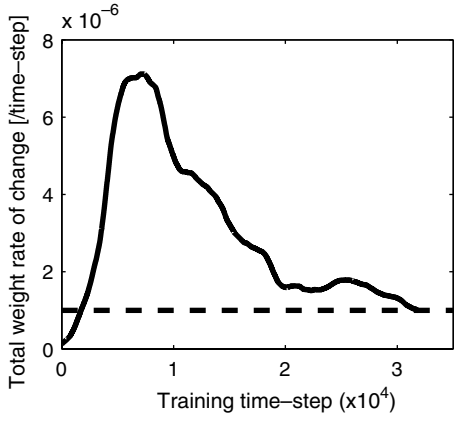

Fig. 10. An example of the evolution of the total-weight-change-rate ($\dot{w}_{\text{total}}$) during training. The graph shows an increasing change rate of the connection weights during the initial stages of training, followed by a gradual decrease of $\dot{w}_{\text{total}}$, indicating convergence. When the value of $\dot{w}_{\text{total}}$ crossed the 10^{-6} level for the second time, training was terminated.

As described below, MST units' responses after training showed clear functional characteristics. The specific parameters of each individual unit (e.g., its preferred direction of movement) depended on the initial (random) connections weights. However, the general functions performed by MST units after training, as well as MST population characteristics, were not sensitive to the choice of initial conditions.

The training rules included two parameters, α^{exc} and α^{inh}, that determined the learning rates of the excitatory and inhibitory synapses, correspondingly. The results in the following sections were obtained with learning rates yielding approximately balanced growth rates of the excitatory and inhibitory synapses. The influence of α^{exc} and α^{inh} on the training process is discussed in Section 3.3.6.

3.3.2. MST units responses during fixation

After training the network, we simulated the responses of the resulting MST units to different stimuli. First we simulated the response to a target moving in the dark in different directions during fixation. Maximal activity during each single simulation was used as a measure of MST units' response. The polar representation of the response was used to determine the unit's preferred direction of movement, and its degree of selectivity, measured by the selectivity index (Orban, 1994; see Appendix A.6). An MST unit was considered directional if its selectivity index exceeded 0.5. Fig. 11a shows the responses of three MST units to a target moving at 8°/s in different directions during fixation. The units are clearly selective to the direction of movement; 52 out of the 60 MST units (~86%) were directional. In the following, we concentrate only on the directional units of the MST population. Fig. 12 shows MST units preferred directions distribution.

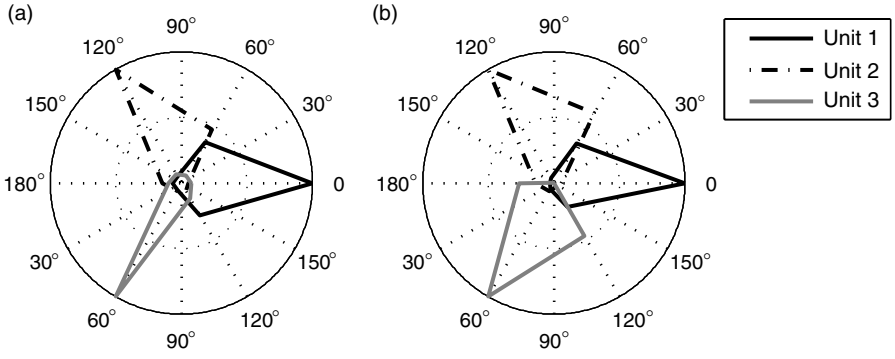

Fig. 11. Polar plots of three MST units responses to different directions of movement. The stimulus consisted of a target moving in the dark at 8°/s during fixation (a), and during stabilized pursuit (b). Responses are normalized relative to response in the preferred direction of movement. The three units are clearly directional, both during fixation and during pursuit. Moreover, for each of the three units, the preferred directions during fixation and during pursuit are closely related.

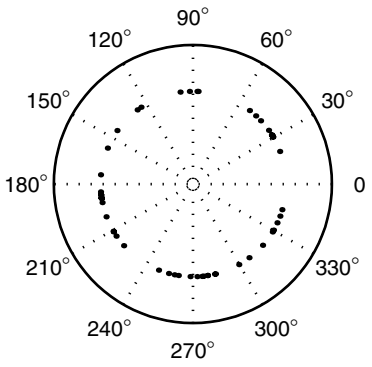

Fig. 12. A Polar plot of the preferred directions of movement of all directional MST units. Each point indicates the preferred direction of a single MST unit during fixation.

3.3.3. MST units responses during pursuit

Next we studied the responses of the directional MST units during stabilized pursuit. Stabilized rather than normal pursuit was used here since it enables us to separate the eye-movement input from the retinal one, thus greatly simplifying the analysis of MST units responses and connectivity. For the MST units, response properties during normal and stabilized pursuit were qualitatively similar.

Again, a target moving in the dark in different directions was used. Fig. 11b shows the response of the same three MST units depicted in Fig. 11a, during pursuit of a moving target. The units are clearly selective to direction of pursuit, with the preferred direction of pursuit closely related to the preferred direction of movement during fixation. We tested directionality during pursuit by the same method used during fixation; of the 52 units that were directional during fixation, 51 were also directional during pursuit.

The difference between preferred direction during pursuit and during fixation had a mean value of $-2.8° \pm 8.7°$. It can be concluded that after training, most MST units were selective to direction of objects' movement, and their preferred direction was nearly the same whether the eyes fixated or followed the object (see Section 4).

3.3.4. Velocity tuning of the MST units

Next, the velocity tuning of the directional MST units was tested. We simulated responses to movements at different velocities (0.5°/s, 2°/s, 8°/s, 32°/s) in each unit's preferred direction. Fig. 13a and b show the velocity tuning of three MST units during fixation and during stabilized pursuit, respectively. As can be seen, the velocity responses during fixation and during pursuit were correlated, although MST units responded better to slower movement velocities during fixation than during pursuit (see Section 4). MST units were classified to three groups according to their velocity preferences: (1) units with low-pass velocity tuning, responding preferentially to slow movements (0.5°/s–2°/s; Fig. 13, unit 1); (2) band-pass units, preferring mid-range velocities (2°/s–8°/s; Fig. 13, unit 2); and high-pass units, preferring high velocities (8°/s–32°/s; Fig. 13, unit 3). Most MST units had either low-pass or high-pass velocity tuning (11 and 38 units, respectively), while the remaining 3 units had band-pass tuning.

3.3.5. MST connections analysis

We turn now to a description of the input connections to the MST layer that were formed after training. For clarity, we concentrate on connections to the three MST units whose responses were shown in Fig. 11 and restrict the discussion to input units having a single preferred velocity of movement: 8°/s for MT, and 32°/s for the eye movement representation. For these preferred velocities the strongest input connections were formed, since the three units had a high-pass velocity tuning (see Section 3.3.4). Four classes of connections to MST were formed. Fig. 14a and b show the excitatory and inhibitory MT-to-MST connections, respectively. MT units are labeled according to their preferred direction of movement at their RF center. Fig. 14c and d show the excitatory and inhibitory connections between the eye-movement representation units and the three MST units, respectively. The eye movement units are labeled according to their preferred eye movement direction. The size of each rectangle represents the connection strength between an input unit and an MST one.

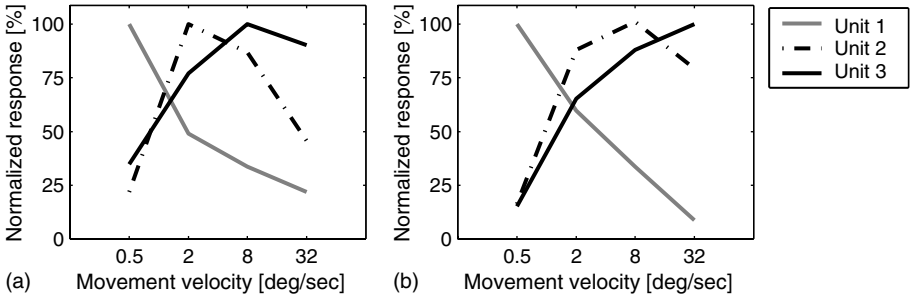

Fig. 13. Responses of three MST units to different movement velocities during fixation (a) and during stabilized pursuit (b). Each unit was stimulated by a target moving in its preferred direction. Responses were normalized relative to responses in the preferred velocity.

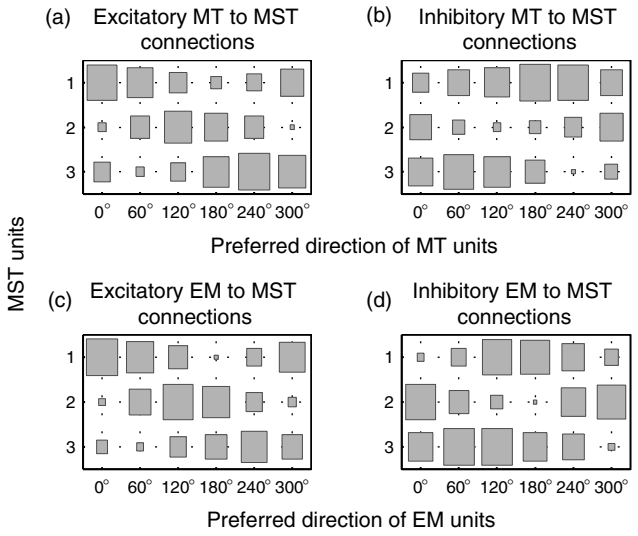

Fig. 14. Connection patterns to the three MST units depicted in Fig. 11, after training. Only connections to input units with a preferred velocity forming the strongest connections are shown: MT units with a 8°/s preferred velocity, and eye-movement units with 32°/s preferred velocity. The area of each rectangle represents the connection weight between an input unit and an MST one: (a) excitatory connections between MT and MST units. MT units are labeled according to their preferred direction of movement in the RF center; (b) inhibitory MT to MST connections; (c) excitatory connections between eye-movement representation units and MST units. Eye movement units are labeled according to their preferred direction of eye movements; (d) inhibitory connections between eye-movement units and MST units.

The direction selectivity of MST units during pursuit and during fixation can be explained by the nature of the underlying connections. Fig. 11a showed the responses of the three MST units to different directions of movement during fixation. Unit 1, for example, had a preferred direction of ~0°. According to Fig. 14a, it had strong excitatory connections to MT units selective to movement about 0°, while its connections with MT units selective to the opposite direction were the weakest. This explains its direction preference during fixation. The inhibitory MT-to-MST connections shown in Fig. 14b are complementary to the excitatory ones; the strongest of these connections to MST unit 1 were with MT units selective to movements about 180°. By a kind of a push-pull mechanism the inhibitory connections also contributed to direction selectivity. The same principle holds for MST units 2 and 3, resulting in different preferred directions. Fig. 11b showed the responses of the three MST units during stabilized pursuit in the dark. Here the retinal image was stable, so MST-units stimulation came from the units representing eye movements. According to Fig. 14c, MST unit 1, for example, was strongly connected to eye-movement units selective to eye movements about 0°. As a result, MST unit 1 was selective to a direction of pursuit about 0°. Again, the inhibitory connections, shown in Fig. 14d are complementary to the excitatory connections of Fig. 14c.

The velocity-dependent responses of MST units (Section 3.3.4) are also directly related to the connection patterns with the input units. MST units that preferred high movement velocity, for example, were more strongly connected to input units having higher preferred velocities.

In conclusion, post-training connection patterns between MST units and their input units show clear regularities. The response properties of MST units during fixation and pursuit were directly related to the connection patterns with visual and non-visual inputs.

3.3.6. Sensitivity to excitatory and inhibitory learning rates

The training process was repeated with various combinations of excitatory and inhibitory learning rates (α^{exc} and α^{inh}). For classification of the post-training MST population responses, we used two criteria. First, we calculated the percentage of directional units (see Section 3.3.2). Second, we defined a distribution index, DI, to measure the distribution of preferred directions. $DI \simeq 1$ when the direction distribution is uniform, while $DI = 0$ when all preferred directions are the same (see Appendix A.7).

Different α^{exc} to α^{inh} combinations determined the future of population responses. When $\alpha^{inh} \gg \alpha^{exc}$, inhibitory connections predominated after training. Therefore, most units were inhibited by all input directions, and as a result less than 50% of the units were directional (Fig. 15, squares). When $\alpha^{inh} \simeq \alpha^{exc}$, the excitatory and inhibitory connections grew in a balanced way. In this case, more than 50% of the units were di-

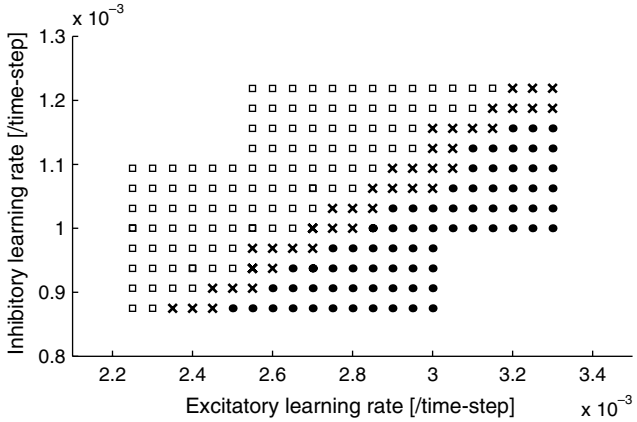

Fig. 15. MST population characteristics as a function of the excitatory and inhibitory learning rates. Training of MST connections was repeated with various combinations of excitatory and inhibitory learning rates. After training, three types of population response were observed. In one (squares), less than 50% of the units were directional. In the second one (crosses), more than 50% of the units were directional, and the distribution of preferred directions was well spread. In the third (filled circles), more than 50% of the units were directional, but preferred directions tended to concentrate around 1–3 principal directions.

rectional, and the distribution of preferred directions was well spread (DI > 0.5; Fig. 15, crosses). Finally, when $\alpha^{inh} \ll \alpha^{exc}$, the excitatory connections grew significantly faster than the inhibitory ones, and most units were directional. Preferred directions, however, tended to concentrate around 1–3 principal directions (DI < 0.5; Fig. 15, filled circles). These results indicate the importance of the interplay between inhibitory and excitatory connections in the training process (see Section 4).

4. Discussion

4.1. Modeling motion analysis during pursuit eye movements

Motion processing during pursuit eye movements involves transformation from a retinal reference frame to world-centered coordinates. Psychophysical and physiological evidence indicates that this process is achieved via integration of retinal motion signals and an internal signal related to the execution of eye movements. While neurons in the first cortical stages of motion processing respond only to retinal image movements, cells in area MST receive, in addition to direct visual input, an extra-retinal one related to pursuit eye movements (Newsome et al., 1988). The focus of the present paper is this interaction between retinal and extraretinal signals in MST pursuit cells.

Dicke and Thier (1999) studied the role of area MST in a model of combined smooth eye–head pursuit. Their work deals mainly with the execution of pursuit eye and head movements, so that the model does not go into details of motion representation in V1 and MT. Although the model focuses on aspects of pursuit generation, some of its results are relevant to perceptual consequences of pursuit as well. The model units are responsive to retinal image slip as well as to eye and head velocity with similar preferred directions. The authors suggest that a population of such neurons is able to reconstruct target motion in world-centered coordinates, and account for salient properties of visually guided pursuit.

The model by Pack et al. (2001) investigates interactions between cells in the ventral and dorsal subdivisions of MST, hypothesized to process target velocity and background motion. Similar to early studies, the model assumes a subtraction of extraretinal information about the velocity of eye rotation from retinal information about target velocity. The model addresses a number of behavioral phenomena related to velocity of pursuit eye movements and perceptual estimation of target and background velocities. Their model assumes, for simplicity, that movements are one-dimensional (leftward and rightward) and focuses on velocity as the central parameter.

While previous models assumed predetermined connectivity and functioning of MST units, the present work studied whether integration of retinal and extra-retinal signals can result from an unsupervised training process of the connections to MST units. For that purpose we constructed a neural network model with three layers of computational units, simulating properties of cortical neurons at different stages of the motion analysis process. Our model analyzes, for the first time, pursuit related neurons in area MST in a broad context including a full representation of direction and velocity of motion in V1 and MT.

The principal results of our model relate to the formation and response properties of the units in the third processing layer, simulating MST pursuit-selective cells. These units integrate retinal motion signals represented by MT units, and an extra-retinal signal indicating eye movements. We showed that an unsupervised training process of the connections to the model MST layer can generate pursuit-related units with response properties in accordance with physiological findings. These units are selective for the direction of objects' movement, and their preferred direction is the same whether the eyes fixate or follow the object, i.e., the units represent a transition from retinocentric motion analysis to real-world motion detection. By inspecting the connectivity patterns between the different units, the MST units response properties were related to the underlying neuronal mechanisms.

4.2. Motion representation in V1 and MT—simulation and physiology

The first two layers in our network model simulate motion representation by neurons in cortical areas V1 and MT. The first layer in the model simulated, by using delayed inhibition, direction selective cells in the primary visual cortex (Dow, 1974; Hubel & Wiesel, 1968; Snodderly & Gur, 1995). The model units responded selectively to a movement in the preferred direction and velocity within their RF. We demonstrated how the population of the model V1 units represented retinal movements in a manner that resembles many aspects of motion representation in cortical area V1.

The units in the second processing layer received their input from the V1 layer, and simulated antagonistic center-surround organization of MT RFs (Allman et al., 1985; Born & Tootell, 1992; Raiguel et al., 1995; Tanaka et al., 1986; Xiao et al., 1995). Following previous models, our modeling of MT units was based on direction-dependent connections between motion detectors at different spatial locations (see Section 2.4). We demonstrated how movement in different directions in the surround modulated the responses to movement in the

RF center. When the center was stimulated by movement in the preferred direction, stimulating the surround by movement in the same direction as the center's decreased the unit's response, while surround movement in the opposite direction facilitated it. These results are consistent with response properties of most MT neurons with center-surround RFs (c.f., Allman et al., 1985).

We also showed that some MT units responded during pursuit of a target moving in the preferred direction of their RF center. The retinal slip of the target, which is in the same direction as that of the eye movement, stimulated the excitatory center of the RF, while image background movements modulated the response by surround inhibition and disinhibition. The response of the model MT units is in accordance with physiological findings. Newsome et al. (1988) showed that some foveal MT cells are active during pursuit, and that their response stems mainly from the slow retinal slip of the pursued target. Lagae et al. (1993) found that, indeed, in the foveal region of MT, most cells preferred low velocities.

4.3. Unsupervised training of excitatory and inhibitory connections to MST

The third processing layer in our model represented MST pursuit cells. Following physiological findings, these units received both visual input arriving through MT and an additional input representing eye movements. The connections to the model MST units developed during an unsupervised training process where the weight modification was based solely on the network response to the input presented at each learning stage, and did not depend on an external feedback to shape the network performance.

One of our basic assumptions was the normal functioning of the pursuit system during the training processes, i.e., while perceptual mechanisms are being shaped. This assumption can be supported by the fact that mechanisms responsible for generating eye movements are present at early stages of neural development (Aslin, 1981; Dayton & Jones, 1964; Shea & Aslin, 1990).

Connections to MST units were separated to excitatory and inhibitory ones (''synapses''). The training rules used for the two kinds of connections were different, reflecting the different functions of these synapses. For training the excitatory connections, we used Oja's rule (Oja, 1982), a modification of Hebbian learning (Hebb, 1949), that favors correlated activity between pre- and post-synaptic neurons. While in basic Hebbian learning the connection weights keep on growing without bound, Oja's rule makes the weights approach a constant limit. In contrast to excitatory synapses, inhibitory synapses reduce, on average, the correlation between pre- and post-synaptic neurons' activities. Therefore, for the inhibitory synapses we constructed a learning rule favoring anti-correlation between the activity of pre- and post-synaptic neurons (see Appendix A.5 for details).

The network was trained on a series of input movies representing periods of pursuit of a target moving in different directions and velocities. At the beginning of the training period connections to MST were random and relatively weak. Therefore pre-training MST units did not perform any specific function. After training onset, a transient period was observed where organization of the synaptic weights occurred. Convergence was assessed by a monotonic decrease in the total weight rate of change.

An interesting point observed during simulations is the effect of interplay between excitatory and inhibitory connections development (Section 3.3.6). We showed that when the excitatory and inhibitory synapses grew in an approximately balanced rate, most of the resulting MST units were directional, and the distribution of preferred directions of movement was well spread. From a mathematical point of view, this result indicates a relatively large number of possible stable states for the post-training network units. Different results were obtained for unbalanced development rates for the excitatory and inhibitory synapses. If excitatory connections predominated at the end of training, many MST units were directional, but preferred directions tended to be concentrated around two or three principal directions. That is, the number of stable states for the network was dramatically reduced. On the other hand, if the inhibitory synapses grew significantly faster than the excitatory ones, training resulted in MST units that would be inhibited by all input patterns.

We notice that the growth of inhibitory connections effectively reduced the growth rate of the excitatory connections. However, maximal spread of preferred directions was obtained only with combination of excitatory and inhibitory connections, and was not observed if only excitatory connections with lower learning rates were included. Although there were no explicit inhibitory connections between MST units (lateral inhibition), the anti-correlation inhibitory learning rule implicitly favored differentiation of MST preferred directions, since it depends on the relative activity between a given MST unit and its neighbors. We see that the development rate of the inhibitory synapses determines the characteristics of the post-training units, so that different types of population responses may result from the same basic mechanisms.

4.4. Response properties of MST units after training

After training the connections to the MST layer, we investigated the response properties of the resulting units. First we checked the units' selectivity to move-

ments in different directions, both during fixation and during stabilized pursuit of a moving object. As was described in Section 3.3, most MST units were selective for direction of movement both during fixation and during pursuit. Moreover, for each unit, the preferred direction of objects' movement was nearly the same whether the eyes fixated or followed the object. These response properties are consistent with physiological findings (Komatsu & Wurtz, 1988b). We see that while MT units represent information in the retinal coordinate frame, MST cells respond according to the "real" direction of movement, independent of the retinal event. In other words, the MST units perform a coordinate transformation from the retinal reference frame to that of the environment.

We have also tested MST units' responses to different target velocities. The shapes of velocity tuning during fixation and during pursuit were correlated, although MST units responded better to slower movement velocities during fixation than during pursuit. The reason for this is the difference in the velocity input to MT and the motor units during training: MT units responded to the slow retinal slip, while the motor command represented the velocity of the eye movement itself. However, background movement during training facilitates responses of MT units with preferred velocities 8°/s and 32°/s (cf. Fig. 9). Also, MT cells respond to a range of velocities so that during training cells responding to velocities higher than 0.5°/s or 2°/s were activated. Consequently, during training these MT units develop relatively strong connections with MST, and post-training differences in velocity responses during fixation and during pursuit are moderate.

Most MST units belonged to one of two categories in their velocity preferences; some units showed low-pass responses, preferring low velocities, while others preferred fast moving objects. Units preferring high velocities are consistent with physiological results of MST neurons (Kawano et al., 1994; Tanaka et al., 1993) showing a graded response to velocity. Units having low-pass type of velocity preference are infrequently observed physiologically. The preponderance of the high-pass velocity response may be a result of developmental preferences or constraints, eliminating the second type of response.

After studying the response properties of the MST units, the model was used to explore how visual and non-visual inputs interact to generate the special functional properties of the units. While at the beginning of training connections between MST units and their inputs were random, training resulted in clear regularities of the connection patterns. The connections to the visual input (coming through MT) explain the direction and velocity selectivity of MST cells during fixation, while the connections to the eye-movement representation units explain responses during pursuit of a target moving in the dark. Connection patterns between MST units and the eye-movement representation units were correlated with MT–MST connectivity, and this is the basis for the correlation between preferred directions of pursuit and visual motion.

Post-training connection patterns are closely related to the input patterns the model was trained on. During learning, input units with correlated activity tended to strengthen their excitatory connections with the same MST units (although the learning rule does not depend explicitly on correlation between input units). Here, the input samples were periods of pursuit. During each pursuit period the activity of MT units and units representing eye movements were correlated and this correlation was the basis for the final connection patterns to MST. During training, connection patterns organized, and each MST unit ended in a stable connection pattern. Our results suggest the importance of experiencing periods of pursuit for appropriate development of the perceptual mechanism related to pursuit, and that deprivation of pursuit periods may result in a deficient cortical mechanism that is unable to fully compensate for eye movements during pursuit.

Acknowledgements

The authors thank the two anonymous reviewers for helpful suggestions and comments. This research was supported by the R.L. Kohns Eye Research Fund #130372, and the Fund for the Promotion of Research at the Technion #130358.

Appendix A. Mathematical formulations

A.1. Transfer function of a model unit

The activity $y_j(t)$ of a model unit j at a time step t is calculated as

$$y_j(t) = \sigma\left(\sum_{\tau=0}^{\tau_m}\sum_{i=1}^{n}[w_{ij}^{\mathrm{exc}}(\tau)x_i(t-\tau) - w_{ij}^{\mathrm{inh}}(\tau)x_i(t-\tau)]\right) \tag{A.1}$$

where $x_i(t)$ are the activities of its input units; $w_{ij}^{\mathrm{exc}}(\tau)$ and $w_{ij}^{\mathrm{inh}}(\tau)$ are the excitatory and inhibitory delay-dependent connection weights, respectively, and σ is a non-linear threshold function:

$$\sigma(u) = \frac{1}{1 + k_1\exp(k_2 u)} \tag{A.2}$$

This formulation keeps the activity of a model unit in the $[0, 1]$ range. The parameters k_1 and k_2 were chosen to give (1) a "spontaneous activity" level (that is, $\sigma(0)$, the activity when there is no input) of 0.04, and (2) a 0.99

response for an input value of 2 (the latter characterizes of the active range of the neuron).

A.2. Population coding of eye-movements

Eye movements are represented by the population activity of a set of 24 units, having six preferred directions of eye movement (0°, 60°, ..., 300°), and four preferred velocities (0.5°/s, 2°/s, 8°/s, 32°/s). The dynamic activity of an eye-movement unit having a preferred direction ϕ_k and a preferred velocity v_l is

$$y_{kl}^{\mathrm{EM}}(t)=\begin{cases}\frac{1}{2}[1+\cos(\phi_k-\phi^{\mathrm{EM}}(t))] \\ \quad\times\exp(-4(v_l-v^{\mathrm{EM}}(t))^2/\Delta_{\mathrm{v}}) & \text{during pursuit}\\ 0.04 & \text{during fixation}\end{cases} \tag{A.3}$$

where $\phi^{\mathrm{EM}}(t)$ and $v^{\mathrm{EM}}(t)$ are the direction and velocity of pursuit, respectively, and Δ_{v} is the characteristic width of the velocity tuning.

A.3. Delayed-inhibition mechanism for local motion detection

As mentioned in Section 2.3, the RF of a V1 unit is divided in two by its long axis. The spatio-temporal structure of the RF can be described by a two dimensional $x-\tau$ map, where x is a coordinate running perpendicular to the long axis of the RF, indicating the distance of a retinal pixel from the long axis.

We formulate now the delay-dependent connection weight $w_{ij}(\tau)$ between a retinal pixel i positioned at $\vec{r}_i$, and a V1 unit j with a preferred direction θ and a retinotopic position $\vec{r}_j$. Let $\hat{n}$ designate a unit vector pointing at the preferred direction of movement of the V1 unit, that is, $\hat{n}=(\cos\theta,\sin\theta)$. This unit vector is perpendicular to the long axis of the RF. Therefore, the scalar product $(\vec{r}_j-\vec{r}_i)\cdot\hat{n}$ indicates the distance of the pixel i from the long axis of the RF. The connection weight can be written as

$$w_{ij}(\tau)=\begin{cases}1 & ((\vec{r}_j-\vec{r}_i)\cdot\hat{n},\tau)\in\mathscr{D}^{\mathrm{E}}\\ -1 & ((\vec{r}_j-\vec{r}_i)\cdot\hat{n},\tau)\in\mathscr{D}^{\mathrm{I}}\\ 0 & \text{otherwise}\end{cases} \tag{A.4}$$

where $\mathscr{D}^{\mathrm{E}}$ and $\mathscr{D}^{\mathrm{I}}$ are the excitatory and inhibitory domains in the $x-\tau$ map, described as follows: the value $x=0$ refers to the long axis of the RF, and the positive x direction points to the preferred direction of movement. Let x_{m} designate the maximal value of x within the RF (that is, half the width of the RF). The excitatory region in the $x-\tau$ map is

$$\mathscr{D}^{\mathrm{E}}=\{0\leqslant x\leqslant x_{\mathrm{m}},x\leqslant\tau\leqslant x_{\mathrm{m}}\} \tag{A.5}$$

and the delayed inhibition region is defined by

$$\mathscr{D}^{\mathrm{I}}=\{-x_{\mathrm{m}}\leqslant x\leqslant 0,x+2x_{\mathrm{m}}\leqslant\tau\leqslant 2x_{\mathrm{m}}\} \tag{A.6}$$

Finally, V1 units responses are modulated according to retinal movement velocities within their RFs. The representation of velocity in the retinal image was explicit; at each time step t, a scalar $v_i(t)$ was attached to each retinal pixel i to indicate movement velocity. At each time step t the pool of velocities among the pixels belonging to the V1 unit j RF is taken. Let $\bar{v}_{ij}(t)$ denote the mode (in the statistical sense) of this velocities population. The total response of the V1 unit was

$$y_j^{\mathrm{V1}}(t)=\sigma_0+\left(\sigma\left(\sum_{i=1}^{n}[w_{ij}(\tau)p_i(t-\tau)]\right)-\sigma_0\right)\times\exp\left(-4(v_j-\bar{v}_{ij}(t))^2/\Delta_{\mathrm{v}}\right) \tag{A.7}$$

where σ_0 is the spontaneous activity level (0.04), v_j the preferred velocity of V1 unit j, Δ_{v} the width of the velocity tuning, and $p_i(t)$ the value of the retinal pixel i at time t.

A.4. Connection weights between V1 and MT layers

The delay-dependent connection weight $w_{ij}(\tau)$ between an MT unit j and a V1 unit i is:

$$w_{ij}(\tau)=\sum_{k=1}^{2}\frac{(-1)^k d(\tau)}{4\pi^2 D_k^2}\exp\left(-\frac{|\vec{r}_j-\vec{r}_i|^2}{D_k^2}-\frac{|\theta_j-\theta_i|^2}{\Theta^2}\right)\times V(v_j,v_i) \tag{A.8}$$

where $k=1$ refers to inhibitory connections, and $k=2$ to excitatory ones. The delay-dependence $d(\tau)$ of the connection weights was chosen to perform an averaging of the input activity during a few time steps

$$d(\tau)=\begin{cases}\mathrm{e}^{-\tau} & 0\leqslant\tau\leqslant 4\\ 0 & 4<\tau\end{cases} \tag{A.9}$$

$\vec{r}_i$ and $\vec{r}_j$ are the retinotopic positions of the V1 and MT units, v_i, v_j their preferred velocities of movement, and θ_i,θ_j their preferred directions, respectively. The difference between preferred directions is always taken in the $[-180°,+180°]$ range. D_k are the characteristic ranges of excitation and inhibition. Their values implicitly determine the sizes of the excitatory center of the RF and its inhibitory surround. Θ is the characteristic width of the direction-dependent term, and its value was chosen as 90°. Finally, the velocity-dependent term $V(v_j,v_i)$ defines the velocity tuning of the MT units. It was constructed to yield velocity response curves characteristic of foveal MT units (see Section 2.4). Let the index values 1, 2, 3 and 4 represent the preferred velocities 0.5°/s, 2°/s, 8°/s, 32°/s respectively. Then $V(v_j,v_i)$ is represented by the 4×4 matrix

$$V = \begin{pmatrix} 1 & 0.85 & 0.5 & 0 \\ 0.6 & 1 & 0.85 & 0.25 \\ 0.4 & 0.7 & 1 & 0.6 \\ 0.4 & 0.6 & 0.8 & 1 \end{pmatrix} \tag{A.10}$$

A.5. Training of connections to the MST layer

Connections to the MST layer were adapted according to the following rules. For the excitatory synapses, Oja's rule (Oja, 1982) was used. The momentary connection-weight change $\Delta w_{ij}^{\text{exc}}$ between a neuron with activity level a_i, and a neuron with activity level a_j, is

$$\Delta w_{ij}^{\text{exc}} = \varepsilon_{\text{exc}} a_j (a_i - w_{ij} a_j) \tag{A.11}$$

The weight decay factor which is proportional to a_j^2, makes the weights approach a constant limit, while in classical Hebbian learning, connection weights increase without bound. The parameter ε_{exc} determines the learning rate. For the inhibitory synapses we constructed a learning rule favoring anti-correlation in the activity of the pre- and post-synaptic neurons:

$$\Delta w_{ij}^{\text{inh}} = \varepsilon_{\text{inh}}[(1-\lambda)(\widehat{a_i} - \widehat{a_j})^2 - \lambda(\widehat{a_i} + \widehat{a_j} - 1)^2 + \lambda] \tag{A.12}$$

where $\widehat{a_i}$ and $\widehat{a_j}$ are the normalized firing rates of the two neurons (their value is between 0 and 1). The normalization is performed over the population of the neurons in the layer they belong to. ε_{inh} determines the learning rate of the inhibitory synapses. This inhibitory learning rule was a hyperbolic-paraboloid function of the normalized values $\widehat{a_i}$ and $\widehat{a_j}$, with a range of [0,1] and a saddle point value of λ, which was set to 0.7.

To assess convergence we defined the total weight change rate:

$$\dot{w}_{\text{total}}(t) = \frac{1}{T} \sum_{\tau=t-T}^{t} \sum_i \sum_j \sum_{k=1}^{2} [(w_{ij}^k(\tau) - w_{ij}^k(\tau-1))^2] \tag{A.13}$$

where i runs over all MST units; j runs over all input units to the MST layer; $k = 1, 2$ refers to excitatory and inhibitory connections, respectively; and the total weight change is averaged over $T = 500$ time steps.

A.6. Selectivity index and preferred direction of movement

In this section we assume that a unit was tested on stimulus directions θ_k $(k = 1, \ldots, n)$ and yielded corresponding responses r_k. Following Orban (1994) we used the selectivity index SI to measure the units' degree of selectivity

$$\text{SI} = \frac{\sqrt{\left(\sum_{k=1}^n r_k \sin\theta_k\right)^2 + \left(\sum_{k=1}^n r_k \cos\theta_k\right)^2}}{\sum_{k=1}^n r_k} \tag{A.14}$$

We have also determined the preferred direction Θ of the unit using a polar (ρ, θ) plot of its responses. The preferred direction was taken as the direction of the center-of-mass of the domain $\mathscr{D}$ enclosed by the response points, with density $1/\rho$

$$\Theta = \text{arctg}\left(\frac{y_c}{x_c}\right) \tag{A.15}$$

where

$$(x_c, y_c) = \left(\iint_{\mathscr{D}} \frac{x}{\rho}\,\mathrm{d}s, \iint_{\mathscr{D}} \frac{y}{\rho}\,\mathrm{d}s\right) \tag{A.16}$$

We assume, for simplicity, a linear correspondence $r(\theta)$ between every two measurements

$$r(\theta) = r_{k-1} + \left(\frac{\theta - \theta_{k-1}}{\theta_k - \theta_{k-1}}\right)(r_k - r_{k-1}), \quad \theta_{k-1} \leqslant \theta \leqslant \theta_k \tag{A.17}$$

and then the two integrals can be solved analytically, yielding

$$x_c = \sum_{k=1}^{n} \left(\frac{(r_k - r_{k-1})(\cos\theta_k - \cos\theta_{k-1})}{\theta_k - \theta_{k-1}} + r_k \sin\theta_k - r_{k-1}\sin\theta_{k-1} \right) \tag{A.18}$$

$$y_c = \sum_{k=1}^{n} \left(\frac{(r_k - r_{k-1})(\sin\theta_k - \sin\theta_{k-1})}{\theta_k - \theta_{k-1}} - r_k \cos\theta_k + r_{k-1}\cos\theta_{k-1} \right) \tag{A.19}$$

where $r_0 \equiv r_n$ and $\theta_0 \equiv \theta_n$.

A.7. Distribution index

To measure the distribution of preferred directions we defined a distribution index, as follows:

$$\text{DI} = \frac{2}{\pi n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |\theta_j - \theta_i| \tag{A.20}$$

where θ_k indicates the preferred direction of unit k $(k = 1, 2, \ldots, n)$, and the difference between preferred directions is always taken in the $[0°, 180°]$ range.

References

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology, 52*, 1106–1130.

Allman, J., Miezin, F., & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field

in the middle temporal visual area (MT). *Perception, 14*, 105–126.

Amthor, F. R., & Grzywacz, N. M. (1991). Nonlinearity of the inhibition underlying retinal directional selectivity. *Visual Neuroscience, 6*, 197–206.

Aslin, R. N. (1981). Development of smooth pursuit in human infants. In D. J. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 31–51). Hillsdale, New Jersey: Erlbaum.

Barlow, H. B., & Levick, W. R. (1965). The mechanism of directionally selective units in rabbit's retina. *Journal of Physiology, 178*, 477–504.

Berezovskii, V. K., & Born, R. T. (2000). Specificity of projections from wide-field and local motion-processing regions within the middle temporal visual area of the owl monkey. *Journal of Neuroscience, 20*, 1157–1169.

Borg-Graham, L. J., & Grzywacz, N. M. (1992). A model of the directional selectivity circuit in the retina: transformations by neurons singly and in concert. In T. McKenna, J. Davis, & S. F. Zornetzer (Eds.), *Single neuron computation* (pp. 347–375). Orlando, FL: Academic Press.

Born, R. T., & Tootell, R. B. (1992). Segregation of global and local motion processing in primate middle temporal visual area. *Nature, 357*, 497–499.

Dayton, G. O., & Jones, M. H. (1964). Analysis of characteristics of the fixation reflexes in infants by use of direct current electrooculography. *Neurology, 14*, 1152–1156.

Desimone, R., & Ungerleider, L. G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the macaque. *Journal of Computational Neurology, 248*, 164–189.

Dicke, P. W., & Thier, P. (1999). The role of cortical area MST in a model of combined smooth eye-head pursuit. *Biological Cybernetics, 80*, 71–84.

Dow, B. M. (1974). Functional classes of cells and their laminar distribution in monkey visual cortex. *Journal of Neurophysiology, 37*, 927–946.

Dukelow, S. P., DeSouza, J. F. X., Culham, J. C., van den Berg, A. V., Menon, R. S., & Vilis, T. (2001). Distinguishing subregions of the human MT+ complex using visual fields and pursuit eye movements. *Journal of Neurophysiology, 86*, 1991–2000.

Erickson, R. G., & Thier, P. (1991). A neuronal correlate of spatial stability during periods of self-induced visual motion. *Experimental Brain Research, 86*, 608–616.

Ferrera, V. P., & Lisberger, S. G. (1997). Neuronal responses in visual areas MT and MST during smooth pursuit target selection. *Journal of Neurophysiology, 78*, 1433–1436.

Furman, M., & Gur, M. (1999). A neural model for motion processing in the visual cortex during pursuit eye movements. *Neuroscience Letters, 54*(Suppl.), S22 (Abstract).

Georgopoulos, A. P., Taira, M., & Lukashin, A. (1993). Cognitive neurophysiology of the motor cortex. *Science, 260*, 47–52.

Gregory, R. L. (1958). Eye movements and the stability of the visual world. *Nature, 182*, 1214–1216.

Harris, L. R., & Jenkin, M. (1997). Computational and psychophysical mechanism of visual coding. In M. Jenkin, & L. R. Harris (Eds.), *Computational and Psychophysical Mechanisms of Visual Coding* (pp. 1–19). Cambridge, UK: Cambridge University Press.

Hebb, D. O. (1949). *The organization of Behavior*. New York: Wiley.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London), 195*, 215–243.

Kawano, K., Shidara, M., Watanabe, Y., & Yamane, S. (1994). Neural activity in cortical area MST of alert monkey during ocular following responses. *Journal of Neurophysiology, 71*, 2305–2324.

Kim, J., & Wilson, H. R. (1997). Motion integration over space: Interaction of the center and surround motion. *Vision Research, 37*, 991–1005.

Komatsu, H., & Wurtz, R. H. (1988a). Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. *Journal of Neurophysiology, 60*, 580–603.

Komatsu, H., & Wurtz, R. H. (1988b). Relation of cortical areas MT and MST to pursuit eye movements. III. Interaction with full-field stimulation. *Journal of Neurophysiology, 60*, 621–644.

Lagae, L., Raiguel, S., & Orban, G. A. (1993). Speed and direction selectivity of macaque middle temporal neurons. *Journal of Neurophysiology, 69*, 19–39.

Liu, L., & Van Hulle, M. M. (1998). Modeling the surround of MT cells and their selectivity for surface orientation in depth specified by motion. *Neural Computation, 10*, 295–312.

Livingstone, M. S. (1998). Mechanisms of direction selectivity in macaque VI. *Neuron, 20*, 509–526.

Maunsell, J. H. R., & Van Essen, D. C. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience, 3*, 2563–2586.

Movshon, J. A, Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1986). The analysis of moving visual patterns. In C. Chagas, R. Gattass, & C. Gross (Eds.), *Pattern Recognition Mechanisms*. New York: Springer-Verlag.

Movshon, J. A., & Newsome, W. T. (1984). Functional characteristics of striate cortical neurons projecting to MT in the macaque. *Society for Neuroscience Abstract, 10*, 993.

Newsome, W. T., Wurtz, R. H., & Komatsu, H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. II. Differentiation of retinal from extraretinal inputs. *Journal of Neurophysiology, 60*, 604–620.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology, 15*, 267–273.

Orban, G. A. (1994). Motion processing in monkey striate cortex. In A. Peters, & K. Rockland (Eds.), *Cerebral cortex—volume 10: Primary visual cortex in primates* (pp. 413–442). New York: Plenum Press.

Pack, C., Grossberg, S., & Mingolla, E. (2001). A neural model of smooth pursuit control and motion perception by cortical area MST. *Journal of Cognitive Neuroscience, 13*, 102–120.

Raiguel, S., Van Hulle, M. M., Xiao, D. K., Marcar, V. L., & Orban, G. A. (1995). Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5) of the macaque. *European Journal of Neuroscience, 7*, 2064–2082.

Sato, H., Katsuyama, N., Tamura, H., Hata, Y., & Tsumoto, T. (1995). Mechanism underlying direction selectivity of neurons in the primary visual cortex of the macaque. *Journal of Neurophysiology, 74*, 1382–1394.

Sereno, M. E. (1993). *Neural computation of pattern motion: Modeling stages of motion analysis in the primate visual cortex*. Cambridge, MA: The MIT Press.

Shea, S. L., & Aslin, R. N. (1990). Oculomotor responses to step-ramp targets by young human infants. *Vision Research, 30*, 1077–1092.

Snodderly, D. M., & Gur, M. (1995). Organization of striate cortex of alert, trained monkeys (macaca fascicularis) ongoing activity, stimulus activity, and widths of receptive field activating regions. *Journal of Neurophysiology, 74*, 2100–2125.

Squatrito, S., & Maioli, M. G. (1997). Encoding of smooth pursuit direction and eye position by neurons of area MSTd of macaque monkey. *Journal of Neuroscience, 17*, 3847–3860.

Tanaka, K. (1998). Representation of visual motion in the extrastriate visual cortex. In T. Watanabe (Ed.), *High-level motion processing* (pp. 295–313). Cambridge, MA: MIT Press.

Tanaka, K., Hikosaka, K., Saito, H., Yukie, M., Fukada, Y., & Iwai, E. (1986). Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *Journal of Neuroscience, 6*, 134–144.

Tanaka, K., Sugita, Y., Moriya, M., & Saito, H. A. (1993). Analysis of object motion in the ventral part of the medial superior temporal area of the macaque visual cortex. *Journal of Neurophysiology, 69*, 128–142.

Von Helmholtz, H. (1866). *Handbuch der physiologischen Optik*. Leipzig: Voss.

Von Holst, E. (1954). Relations between the central nervous system and the peripheral organs. *British Journal of Animal Behaviour, 2*, 89–94.

Xiao, D. K., Raiguel, S., Marcar, V., Koenderink, J., & Orban, G. A. (1995). Spatial heterogeneity of inhibitory surrounds in the middle temporal visual area. *Proceedings of the National Academy of Sciences USA A, 92*, 11303–11306.

Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of Physiology (London), 236*, 549–573.

Zemel, R. S., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Computation, 10*, 403–430.

Zohary, E. (1992). Population coding of visual stimuli by cortical neurons tuned to more than one dimension. *Biological Cybernetics, 66*, 265–272.