

Computational modelling of optic flow selectivity in MSTd neurons

Scott A Beardsley and Lucia M Vaina

Brain and Vision Research Laboratory, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, MA 02215, USA

Received 22 December 1997, in final form 6 May 1998

Abstract. In neurophysiological experiments examining the selectivity of MSTd neurons to visual motion components of optic flow stimuli in monkeys, Duffy and Wurtz (1991) reported cells with double-component (plano–radial and plano–circular) and triple-component (plano–radial–circular) selectivities, while Graziano *et al* (1994) reported neurons selective to a continuum of optic flow stimuli including spiral motion. Here, we address these reported findings under simulated experimental conditions by examining the development of optic flow selectivity in the hidden units of a two-layer back-propagation network. We also examine network motion sensitivity during simulated psychophysical tests via the addition of a competitive decision layer. Network analysis with neurophysiological stimuli identified a majority of hidden units whose position invariance and motion selectivity were consistent with MSTd responses to the visual motion components of optic flow stimuli reported by Duffy and Wurtz and Graziano *et al*. Furthermore, the hidden units developed a continuum of optic flow selectivities independent of the biases associated with the specification of the motion selectivity in the output layer. During psychophysical testing, network responses showed motion sensitivities which met or exceeded human performance. Within the limitations imposed by the learning algorithm, the psychophysical results were consistent with a model of global motion perception via local integration along complex motion trajectories.

1. Introduction

During motion, the projection of points in the visual scene onto the retina forms a motion pattern referred to as optic flow. This optic flow produces visually perceived complex motion patterns which combine self-, object and retinal motions to provide a rich source of information regarding the dynamic structure of the scene. The individual motions forming these complex motion patterns are useful in guiding navigation, computing heading, object tracking, and the recovery of three-dimensional structure from motion. In the light of this, the development of processing mechanisms which can extract individual motions in a visual scene would be a valuable component of any visual system.

Given the wide variety of motions which can occur, we might expect that such processing mechanisms should contain specialized detectors sensitive to the independent motion components (e.g. radial, rotational, translational) inherent in complex motion[†]. The

[†] During motion through a simple rigid environment, complex motion (optic flow) can be decomposed into a set of component flow fields characterized by pure radial, rotational, and translational motion. Since radial and rotational flow fields form a mutually orthogonal set of axes under a point-by-point scalar product of the flow field, they can be used to define a two-dimensional optic flow space. Within this space, off-axis points represent intermediate combinations of radial and rotational motion, (spiral flow fields) which can themselves be used as a

relative responses of these specialized detectors could be used to extract the global self-motion and more local object motions contained within the complex motion of the visual scene.

Psychophysical studies have provided support for the existence of specialized detectors sensitive to radial, rotational, and translational motion (Beverley and Regan 1979, Freeman and Harris 1992, Morrone *et al* 1995, Snowden and Milne 1996, Te Pas *et al* 1996, Burr *et al* 1998). The results of these and other psychophysical studies suggest that these specialized detectors integrate local motions to obtain a global motion percept (Watamaniuk and Sekuler 1992, Smith *et al* 1994, Morrone *et al* 1995, Burr *et al* 1998). Specifically, in experiments performed by Morrone *et al* (1995), the signal-to-noise sensitivity for motion discrimination was measured as a function of stimulus area using random dot displays which presented radial, rotational, and planar motion stimuli. For all sets of stimuli, subjects showed a linear increase in motion sensitivity as a function of stimulus area. Because these results were well approximated by an ideal linear-integrator model of motion detection, Morrone *et al* used them to postulate the existence of neural mechanisms which integrate the components of visual motion along complex trajectories.

Neurophysiological studies in the macaque have revealed cells in the dorsal division of the medial superior temporal area (MSTd) of visual cortex which exhibit preferred responses to pure optic flow fields characterized by expansion, contraction, rotation, and planar motions (Saito *et al* 1986, Tanaka and Saito 1989, Tanaka *et al* 1986, 1989, Duffy and Wurtz 1991a,b, 1995, Orban *et al* 1992, Graziano *et al* 1994, Lagae *et al* 1994). These results confirm the existence of specialized detectors in MSTd which are sensitive to complex motion and, in conjunction with psychophysical results, suggest the existence of structures in the human visual system which contain cells sensitive to the components of complex motion. However, the variability in the reported preferred motion characteristics in monkey MSTd suggests that the most suitable method for characterizing the preferred responses of cells in MSTd is not generally agreed upon.

In experiments measuring neuronal responses to dynamic random dot stimuli, Duffy and Wurtz (1991a,b, 1995) identified neurons with combinations of preferred responses to optic flow and confirmed the results of previous experiments reporting neurons in MSTd selective to radial, rotational, or planar motion. Specifically, they identified neurons exhibiting preferred responses to two or three different types of complex motion stimuli, referring to them as double- (plano-radial and plano-circular) and triple-component (plano-radial-circular) respectively. For these multi-component neurons, the degree of response to the preferred stimulus of each neuron was not constant over each group of double/triple-component neurons but extended over a continuum of quantitative values. Furthermore, the position dependence of these cells spanned a continuum of position dependence with single-component neuron responses being the least position dependent and triple-component neuron responses being the most position dependent.

Graziano *et al* (1994) reported results which suggested an alternate characterization for neurons classified as multi-component by Duffy and Wurtz. They hypothesized that double/triple-component neurons were more responsive to the motion vector combination of the preferred multi-component stimuli. In their experiments, neurons were characterized using the multi-component classification and those exhibiting radial/rotational responses were tested with radial, rotational, and intermediate spiral stimuli. The majority of multi-component radio-circular cells exhibited preferred responses characterized by Gaussian

basis set for the space. The addition of translational motions produces a four-dimensional space in which points not contained in the radial/rotational hyperplane (i.e. stimuli with translational components) represent stimuli with shifted centres of motion.

distributed tuning curves centred on the intermediate single-component spiral stimulus. In these cases, the responses were position invariant to small shifts in the centre of motion (COM) when the optimal preferred stimulus was used. However, as suboptimal multi-component radial and rotational stimuli were presented position dependence increased. Based on these results, Graziano *et al* (1994) concluded that the majority of neurons in MSTd could be most accurately classified as single component, spanning a continuum of preferred responses in the optic flow space.

In order to characterize the neurophysiological and psychophysical results of complex motion processing within a computational framework, several biologically plausible models have been developed to address specific aspects of the reported results. To detect heading direction from optic flow, Lappe and Rauschecker (1993, 1995) proposed models using the subspace algorithm introduced by Heeger and Jepson (1990), while Perrone and Stone (1994) developed a template-matching model utilizing direction- and speed-tuned motion sensors. Both models identified units with preferred responses to complex motion stimuli corresponding to specific subclasses of the multi-component classifications used by Duffy and Wurtz. To examine position invariance in MSTd, Zhang *et al* (1993) constructed a network using simple Hebbian synapses which developed position-invariant units exhibiting selectivities to a continuum of complex motions. However, unlike cells in MSTd, the network decomposed the velocity field, responding to the preferred complex motion components regardless of the magnitudes of additional complex motion components in the stimulus.

Additional models proposed by Wang (1995, 1996), Zemel and Sejnowski (1998), and Pitts *et al* (1997), examined the characteristics of complex motion tuning and the emergent properties of units in biologically plausible networks. Wang (1995, 1996) constructed a competitive network utilizing the hierarchical structure of motion processing identified between the middle temporal (MT) area and MST. The network was trained with simple schematic representations of 'cardinal' complex motion stimuli (radial, rotational, and translational motion) and the properties of units in the middle and output layers were examined. In both layers Wang identified units with a continuum of complex motion selectivities and preferred COMs. Similar results were obtained by Zemel and Sejnowski (1998) using optic flow stimuli generated from movie sequences simulating a variety of natural motions. Using an unsupervised network containing units with different receptive fields, their model identified position-invariance properties similar to those of MST and investigated the ability of MST-like units to segment individual motions in a complex motion scene. Pitts *et al* (1997) developed a model based on the organization of pattern selective units proposed by Duffy and Wurtz (1991b). Using complex motion patterns similar to those of Wang (1995, 1996), the model accounted for the complex motion properties reported by Duffy and Wurtz (1991a,b) and was extended to include preferred responses to a larger set of complex motion stimuli. Within the limitations imposed by the simplified network inputs, the model was able to develop position-independent properties via additional inhibitory connections between units.

This paper consists of two parts. First, we use a two-layer back-propagation network to investigate the properties of the hidden layer under conditions simulating the multi-component and a continuum of single-component responses in MSTd. Second, we test the ability of the network to integrate complex local motions into a global motion percept based on its psychophysical performance. While previous models utilized simple optic flow stimuli we have developed a more complex stimulus set which minimizes symmetry in the input. Unlike unsupervised networks which use similar MT/MST structures and more natural and complex optic flow, the supervised nature of the network presented here allows

the development of intermediate layers to be biased via the output layer. This allows us to compare the methods of cell classification proposed in MST by implementing distributions in the output consistent with each classification and examining the hidden layer of the trained networks under an optimal encoding hypothesis.

To investigate hidden-unit properties we developed a network consistent with a hierarchical structure of motion processing. Network inputs consistent with MT responses to schematic representations of optic flow stimuli were propagated through the network via weights fully connecting adjacent layers. After training under conditions simulating multi-component and a continuum of single-component classifications of cell selectivity, hidden-unit properties including their preferred responses and degree of position invariance were examined. By manipulating the selectivity of units in the output layer through the target conditions, we addressed the multi-component (Duffy and Wurtz 1991a,b) and a continuum of single-component (Graziano *et al* 1994) classifications and investigated the distribution of hidden-unit selectivities under an optimal encoding hypothesis.

We then extended the network to examine its psychophysical performance and investigated its ability to integrate local motions into a global motion percept. To obtain psychophysical responses from the trained network we added a competitive decision layer to the primary network. The decision layer was fully connected to the output layer of the primary network and trained to respond to radial and rotational stimuli. After training, schematic representations of the psychophysical stimuli used by Morrone *et al* (1995) were presented to obtain the motion sensitivity of the network. These results were then compared and contrasted with psychophysical results from human subjects and also with the motion sensitivity predicted by an ideal linear-integrator model of motion discrimination. Finally, the resulting network structure, hidden-unit characteristics, and psychophysical responses of the computational model were discussed and compared with neurophysiological and psychophysical results.

Preliminary versions of this work have been presented at the Society for Neuroscience 1996 Annual Meeting (Beardsley *et al* 1996) and the Association for Research in Vision and Ophthalmology 1997 Meeting (Beardsley and Vaina 1997).

2. Methods

To simulate the neurophysiological experiments on MSTd, we used a two-layer back-propagation network and examined the development of optic flow selectivity in the hidden layer. The physiological and computational organizations of the network were selected to remain consistent with a hierarchy of motion information processing in the visual pathway from MT to MSTd.

2.1. Primary network structure and training: investigation of hidden-unit properties

The primary network was implemented using the physiological and computational organization shown in figures 1 and 2. Using a hierarchical structure, schematic representations of complex motion stimuli were presented to directionally selective MT units whose receptive fields were contained within a set of coincident MST receptive fields 63° in diameter (mean receptive field size reported by Duffy and Wurtz (1991a,b)). Network inputs were fully connected to the hidden layer and consisted of MT directional responses to complex motion/optic flow stimuli. The hidden layer, tentatively designated as MSTd, was fully connected to an output layer explicitly defined to contain units whose responses to optic flow were consistent with the reported selectivity of MSTd neurons.

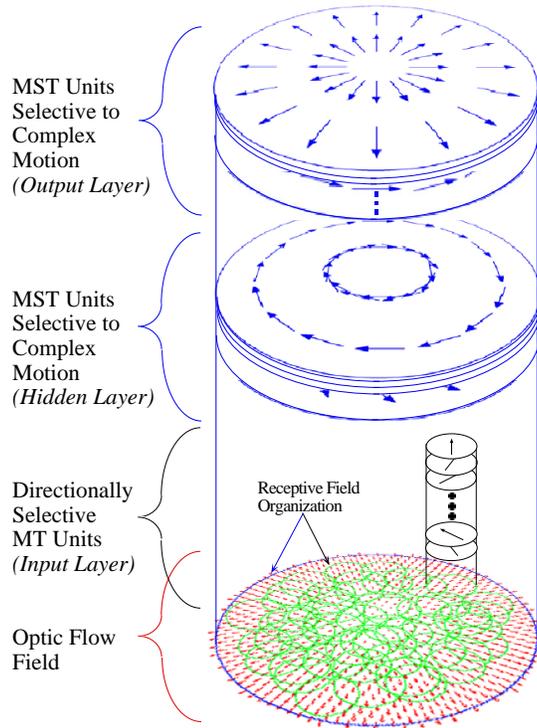


Figure 1. Primary network structure. Inputs to the network consisted of 1072 directionally selective MT units uniformly distributed across 67 overlapping receptive fields and 16 directional selectivities (22.5° increments across the 360° directional space). MT inputs were fully connected to a hidden layer, tentatively designated as MSTd, such that the ‘effective’ receptive field for each hidden unit coincided with the net receptive field of the input layer. The hidden layer was fully connected to an output layer, producing receptive fields in the output layer which were coincident with those of the hidden layer. To accommodate supervised learning in the network the output layer was explicitly defined as MSTd.

Each schematic representation of a complex motion pattern represented an optic flow field containing constant-speed vector representations of motion uniformly spaced throughout a 63° diameter ‘visual field’ with a density of 1 point/degree². The COM of each stimulus was centred in the visual field during training and testing. Network inputs consisted of a population code of 1072 Gaussian MT responses to optic flow stimuli uniformly distributed among 67 receptive fields. Each receptive field corresponded to 16 MT units whose directional selectivities were uniformly distributed in 22.5° intervals throughout the 360° directional space[†]. The net response of each MT unit represented a normalized

[†] To reduce the complexity of the network we used idealized optic flow components and did not include speed selectivity in the input layer. A more biologically plausible and physically valid implementation of the network should include speed variability in the optic flow and speed selectivity in the MT-like inputs. This would increase both the size of the stimulus set and the network while adding a second dimension to the selectivity (direction and speed tuning) of the input layer. Computationally this would allow us to investigate speed tuning in the hidden layer and permit the development of forward and inverse preferred speed gradients reported in MST (Duffy and Wurtz 1997).

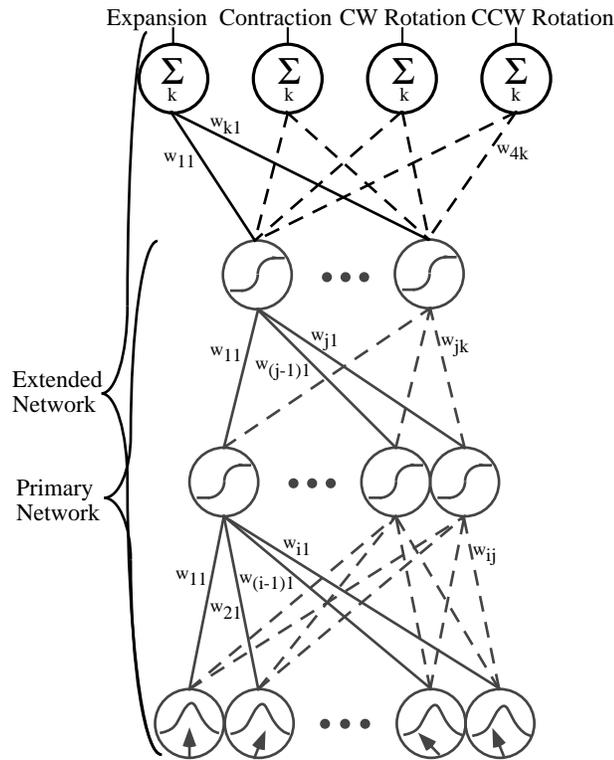


Figure 2. Computational structure of the primary and extended networks. The network input consisted of 1072 directionally selective Gaussian responses ($\sigma = 11.25^\circ$) fully connected with the hidden layer. The response of the hidden and output units was formed from the sum of weighted input to each unit passed through a logistic activation. To force an optimal encoding in the hidden layer, the size of the hidden layer ranged from 15 to 45 units as determined by the size of the output layer and the minimum number of units required to consistently achieve the output error bound. The output layer of the primary network was fully connected with the hidden layer and contained 10 or 20 units whose target responses to input stimuli were chosen to remain consistent with observed MSTd responses to complex motion stimuli. In the extended network, the output layer of the primary network was fully connected to four decision units whose responses utilized a winner-take-all rule based on the sum of weighted inputs to each unit. After training was completed with the psychophysical stimuli, the decision units were classified as expansion, contraction, clockwise rotation, and counterclockwise rotation, according to their binary responses to the respective stimuli.

spike rate calculated using the sum of responses across (N) non-zero motion points in the MT receptive field

$$o_{pi} = \frac{1}{N} \sum_{m=1}^N \exp \left[\frac{(\min[\mu_i - x_{pm}])^2}{2\sigma^2} \right]$$

where o_{pi} represents the response of the i th MT unit to pattern (p), x_{pm} is the orientation angle of the m th motion vector in pattern (p), μ_i is the directional selectivity of the i th MT unit, $\min[]$ refers to the minimum angular distance between μ_i and x_{pm} , and σ is the standard deviation of the Gaussian response for each MT unit. To reduce the effects of stimulus symmetry in the input representation, MT receptive fields (with normally distributed diameters, $\mu_d = 10^\circ$ and $\sigma_d = 0.67^\circ$) consistent with average MT receptive field sizes

(Albright and Desimone 1987, Rodman and Albright 1987, Andersen 1997), were pseudo randomly placed within the encompassing MST receptive field.

To provide above threshold responses to intermediate local motion stimuli in the 360° directional space, a standard deviation of 11.25° was chosen for the MT Gaussian responses. This corresponded to a half-amplitude direction tuning bandwidth of 43.2° such that over the input range the minimum net response to intermediate motion directions was 97% of maximum. While previous neurophysiological studies (Albright 1984, Maunsell and Van Essen 1983), indicate that cells in MT exhibit a broad range of tuning bandwidths (32–186°) with a mean of 95°, network training with larger tuning bandwidths did not significantly affect the hidden unit properties†.

In the hidden layer, the number of units varied as a function of the number of output units and ranged from 15 to 45 hidden units. Hidden units were fully connected to the input layer with the output response of the j th hidden unit to pattern (p) given by

$$o_{pj} = \frac{1}{1 + \exp\left[-(\sum_i w_{ij}o_{pi} + \theta_j)\right]}$$

where θ_j is the bias associated with the j th unit and w_{ij} corresponds to the weighted connection between the i th input and the j th hidden unit.

Similarly, the output layer, designated as MST, was fully connected to the hidden layer and responded via a logistic activation to the summed weighted input. To remain consistent with the neurophysiological findings, the output layer target responses to optic flow were obtained from Gaussian tuning curves predetermined to satisfy either the multi-component or a continuum of optic flow selectivities and characteristics reported by Duffy and Wurtz (1991a,b) and Graziano *et al* (1994) respectively.

The primary network was trained using MT responses to 32 schematic representations of optic flow spaced at 11.25° intervals throughout the 360° optic flow space illustrated in figure 3. For each distribution of preferred responses simulated in the output layer, a predetermined set of target responses was specified for the input set to facilitate supervised learning. Using the sum-squared output error (SSE), the network was minimized through a gradient descent of the error space via a modified version of the classic back-propagation with momentum learning rule (Rumelhart *et al* 1986),

$$\Delta w_{ij}(n+1) = \eta \delta_{pj}(n) o_{pi}(n) + \alpha \Delta w_{ij}(n)$$

where for the n th iteration o_{pi} is the output of the i th unit in the preceding layer to pattern (p), δ_{pj} is the output error of the j th unit in the current layer to pattern (p), and η and α correspond to the learning rate and momentum respectively. To increase the speed of learning and execution, weight adjustments were made adaptively by incrementing the learning rate as a function of the SSE

$$\eta = \begin{cases} 1.001\eta & \text{SSE}(n+1) \leq \text{SSE}(n) \\ 0.07\eta & \text{SSE}(n+1) > \text{SSE}(n) \end{cases}$$

and by performing batched matrix calculations to obtain a gradient descent for each training epoch. During training, the learning rate was set to an initial value of 0.2 and the momentum

† To investigate the effects of the tuning bandwidth on network responses, we constructed a network using broader input direction tuning bandwidths of 95° ($\sigma = 24.5^\circ$). While increasing the input tuning bandwidth provided a net input more closely approximating a uniform distribution throughout the directional space, post-training analysis of the network units showed that the development of hidden-unit properties was consistent with primary networks using narrow tuning bandwidths ($\sigma = 11.25^\circ$).

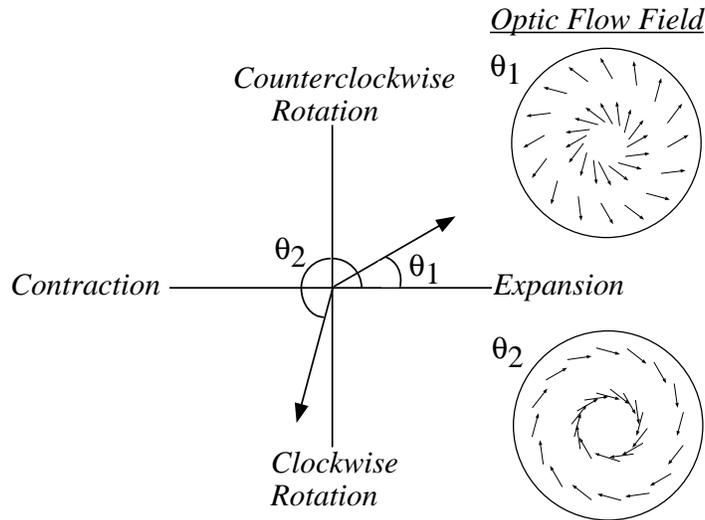


Figure 3. A graphical representation of an optic flow space can be defined by increments of the schematic motion vector angles from the baseline vector representation of a ‘pure’ expansion stimulus. Regions intermediate to the basis motions defined by expansion, contraction, clockwise rotation, and counterclockwise rotation correspond to varying degrees of spiral motion.

was held constant at 0.9. Training was completed when the average error of the output layer across the training set reached a minimum bound of 0.5%.

To constrain the hidden layer to construct an optimal encoding of the MT inputs for each configuration of output units, network structure was minimized via the hidden layer. This was accomplished by reducing the number of hidden units until the minimum error bound could not be consistently achieved within 30 000 epochs. After the network was minimized, and prior to examining the hidden layer, the target Gaussian tuning curves were verified in the output layer.

Because the network contained a minimum of 16 230 weighted connections trained on 32 schematic patterns the system was capable of grossly overfitting the training set. Under these conditions, the network could develop highly non-Gaussian tuning curves in the output layer which satisfied the training patterns but were not consistent with cells in MSTd. To confirm that the output layer correctly generalized to the targetted tuning curves, a test set of 360 patterns spanning the optic flow space formed by radial and rotational motions was generated in 1° increments of the flow angle θ (figure 3). The learned responses to the test set were correlated with the ideal target tuning curves to ensure that the network developed correct generalizations ($r \geq 0.95$). When the output layer generalized correctly to the training patterns, the hidden-unit properties were examined for each network (trained and minimized).

2.2. Extended network structure and training: examining psychophysical performance

After training was completed, the primary network was extended for psychophysical testing by the addition of a decision layer containing four linear units fully connected to the output layer of the primary network (figure 2). To obtain a psychophysical ‘decision’ consistent with the ‘2 alternative forced choice’ (2AFC) protocol of Morrone *et al* (1995), the output responses from the decision layer utilized a binary winner-take-all representation based on

the maximum weighted sum of inputs across units.

Since the primary network responses were deterministic while psychophysical performance is probabilistic, we included internal Gaussian noise ($\sigma = 0.05$) in the extended network to simulate human psychophysical responses more accurately. The Gaussian noise used to represent the internal noise of an observer was added to the output of the decision layer prior to choosing the winning decision unit and provided a decision uncertainty in the psychophysical responses of the network.

The decision layer was trained using a winner-take-all competitive learning algorithm

$$\Delta w_{m^*k}(n) = \eta(o_{pk} - w_{m^*k})$$

where o_{pk} is the output of the k th output unit of the primary network to pattern (p), w_{m^*k} is the weight connection between the winning unit (m^*) in the decision layer and the k th output unit of the primary network, and η corresponds to the learning rate. To prevent domination of the decision response by a single unit and to promote the classification of stimuli into one of four ‘cardinal’ decisions (expansion, contraction, clockwise rotation, and counterclockwise rotation), modifiable biases (b_k) were applied to each decision unit. To distribute winning across units, the biases for consistent losers were increased and the biases of consistent winners decreased such that for the k th decision unit during the $(n+1)$ th iteration

$$b_k(n+1) = e/z_k(n+1)$$

where

$$z_k(0) = 0.25 \quad z_k(n+1) = 0.999z_k(n) + 0.001o_{pk}(n).$$

Each network was trained for a minimum of 3000 cycles with $\eta = 0.05$ using 400 full-field (63°) schematic representations of cardinal optic flow distributed across 10 levels of motion noise (0, 10, ..., 90%). Motion noise was incorporated into the static stimuli by probabilistically designating each motion vector as noise with a probability equivalent to the level of motion noise. When the motion was designated as noise, the velocity vector was randomly reoriented to a new position in the 360° directional space. Since each stimulus used a single schematic image to represent a time series of motion, the dynamic nature of the motion noise was conveyed to the network by presenting 10 separate stimuli per noise level for each of the four cardinal stimuli.

3. Simulations

3.1. Examining the development of optic flow selectivity in the hidden layer

The development of optic flow selectivity in the hidden layer was examined under two conditions for the primary network. The first, referred to as the uniform condition, simulated a ‘continuum of patterns to which MSTd cells are selective’ (Graziano *et al* 1994) by selecting target tuning curves whose Gaussian means were uniformly distributed throughout the optic flow space. To remain consistent with neurophysiological results, the standard deviation of each tuning curve was normally distributed with $\mu = 61^\circ$ and $\sigma = 15^\circ$, (Graziano *et al* 1994).

The second condition, referred to as the cardinal condition, simulated the decomposition of optic flow into the orthogonal ‘basis’ selectivities proposed by Duffy and Wurtz (1991a,b). As for the uniform condition, a normal distribution of standard deviations with $\mu = 61^\circ$ and $\sigma = 15^\circ$ was used to create the target tuning curves. However, in the cardinal condition the means of the target tuning curves were distributed uniformly throughout four cardinal

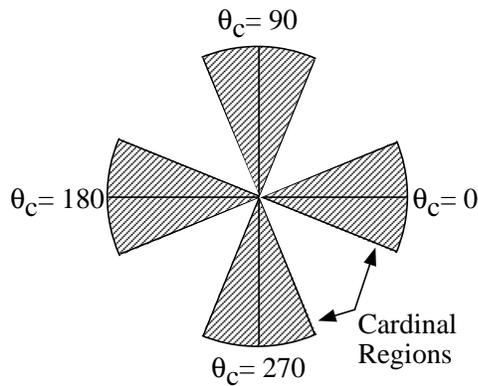


Figure 4. Cardinal regions in the optic flow space centred around expansion, counterclockwise rotation, contraction, and clockwise rotation at $\theta_c = 0, 90^\circ, 180^\circ,$ and $270^\circ,$ respectively. Each region encompassed an angular width of 45° such that intermediate areas form an equivalent set of spiral regions.

regions formed by dividing the optic flow space equally into ‘cardinal’ and ‘spiral’ regions (figure 4). As a result, each cardinal region was centred at its cardinal angle (θ_c) with a range of $\pm 22.5^\circ$.

Under both simulated conditions, two network sizes were implemented. The first used 10 output units in conjunction with a minimized hidden layer of 15, 16, or 17 hidden units. Although the use of 16 and 17 hidden units did not strictly minimize the network, the resulting convergence and hidden-unit properties were consistent with the minimized 15-hidden-unit network. Their inclusion in the simulation analysis increased the sample set of hidden units (reducing biases in the hidden-unit selectivities of individual networks) and indicates that the system need not be strictly minimized. The second network size used 20 output units in conjunction with a minimized hidden layer of 45 hidden units. The increase in network size provided additional variation in the hidden layer and extended the results to larger networks.

3.2. Testing for position invariance in the hidden layer

To examine the position-invariant characteristics of units in the hidden layer, two sets of shifted stimuli were presented to the primary network. The first set simulated the position-invariance test outlined by Graziano *et al* (1994) and extended it to tests with larger stimulus apertures. Hidden-unit responses to optic flow stimuli were examined at five positions within the network receptive field using stimulus apertures of $10^\circ, 20^\circ,$ and 63° . An overlapping cloverleaf arrangement was used such that the COM of shifted stimuli were separated by 5° axially for 10° and 20° aperture stimuli and 10° axially for 63° aperture stimuli (figure 5).

To measure the position invariance quantitatively under the cardinal and uniform conditions, the hidden units and their preferred/anti-preferred optic flow stimuli were classified into one of four extended cardinal regions ($\theta_c \pm 45^\circ$). Hidden unit responses to the preferred and anti-preferred cardinal stimuli (θ_c) at each of the five positions in the receptive field were used to calculate a directional selectivity (DS) index

$$DS = 1 - \frac{\text{response to anti-preferred stimulus}}{\text{response to preferred stimulus}}.$$

Using the DS indices, a position invariance (PI) index was calculated at each of the four surrounding positions via the ratio of shifted and centred DS indices

$$PI = \frac{DS_{\text{surround}}}{DS_{\text{centre}}}$$

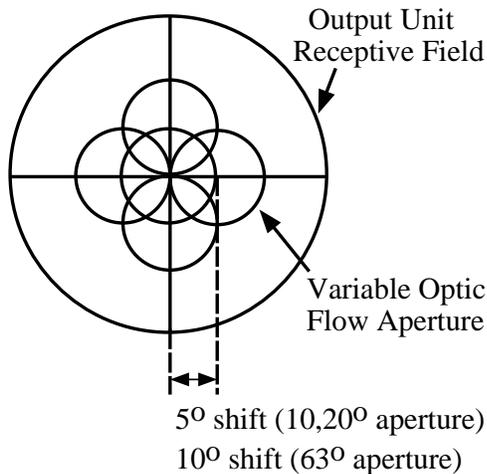


Figure 5. Cloverleaf arrangement of the stimulus apertures used to test for position invariance in the hidden layer. Cardinal stimuli with apertures of 10° , 20° , or 63° were shifted to four positions in the hidden-unit receptive field with axial shifts of 5° , 5° , and 10° , respectively.

and was used to quantify the degree of position invariance observed in the hidden layer.

The second set simulated tests of the spatial distribution of COM specificity similar to those performed by Duffy and Wurtz (1995). Unlike the protocol of Duffy and Wurtz, the network units did not contain a view-centred reference frame which did not coincide with the receptive field of the units in the network. As a result, the network tests were modified from the original neurophysiological tests to present stimuli in the receptive field reference frame of the network.

As with the first set of position-invariance tests, each hidden unit was classified into one of four extended cardinal regions and tested with the corresponding cardinal stimulus. Hidden-unit responses to full-field cardinal stimuli with shifted COMs were examined for 25 COM positions arranged in three concentric rings of eight combined with a centred baseline stimulus. As illustrated in figure 6, each ring corresponded to an angular shift of 22° , 45° , or 90° from centre and contained the appropriate cardinal stimulus shifted in one of eight different directions. To remain consistent with neurophysiological results, COM shifts were labelled as pericentric, eccentric and planar, respectively.

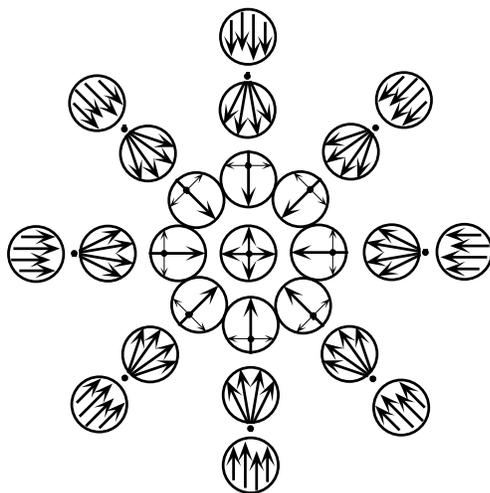


Figure 6. Schematic representation of 25 stimuli containing examples of the shifted COMs used during testing with an expanding cardinal stimulus. In each representation, the circular boundary represents the 63° receptive field of the network, filled circles represent the COM of the stimulus, and the arrows illustrate the relative directions of visual motion within the stimulus. The stimulus set consisted of a centred baseline stimulus (no shift) and three concentric rings of eight shifted COM stimuli each. The inner and middle rings labelled as pericentric and eccentric represented stimuli whose COM was shifted by 22° and 45° , respectively. The outer ring represented planar motion corresponding to stimuli whose COM was shifted to infinity (Duffy and Wurtz 1995).

For each set of shifted COMs, hidden-unit responses to the preferred shifted cardinal stimuli were examined as a function of the direction of shift. The position invariance of units in the hidden layer to large shifts in the COM was examined by comparing the hidden-unit responses for pericentric, eccentric, and planar shifts with each other and with the centred baseline response.

3.3. Psychophysical tests of motion sensitivity

To compare network responses with recent psychophysical results, the extended network was tested for motion sensitivity using schematic patterns of cardinal optic flow whose structure was consistent with psychophysical stimuli used by Morrone *et al* (1995). In the psychophysical experiments, stimuli were represented by movie sequences of 360 black-and-white Gaussian patches (space constant = 0.5°) whose motion was confined to a 10° aperture with the central 1.5° removed. Cardinal and translational stimuli were presented and the subject's ability to discriminate between opposing motions (e.g. expansion or contraction) was tested as a function of stimulus area and motion noise under mask and no-mask conditions. For each test, stimulus area was decreased by dividing the stimulus aperture into 16 sectors and symmetrically masking sectors. Under the mask condition the masked sectors were filled with random dot motion while under the no-mask condition the masked sectors remained blank.

In the network simulations, psychophysical stimuli were represented by schematic motion vectors in a 21° aperture with the central 3.15° removed. The aperture contained 340 motion vectors and provided motion information which was similar to the motion of 360 dots used in the psychophysical experiments. Each stimulus was divided equally into 16 sectors such that the signal stimulus could be confined to 16, 8, 4, 2, or 1 maximally separated sectors. The mask condition was simulated by filling the masked sectors with 100% motion noise (figure 7(b)) and the no-mask condition was simulated by leaving the masked sectors blank. To approximate the dynamic nature of the motion noise during testing under mask/no-mask conditions, 10 patterns were presented per cardinal stimulus per noise level (0–90%) resulting in 400 test stimuli for each increment in stimulus area. Finally, to average the effects of the internal noise of the observer each testing sequence of 400 stimuli was averaged over 100 presentations.

As in the psychophysical 2AFC tasks, the network discriminated between expansion and contraction, or between clockwise and counterclockwise rotation. Correct responses were classified and tallied across noise levels and stimulus area and the average number of correct radial and rotational responses was calculated as a function of noise and signal area by combining the percentage of correct responses from opposing decision units.

To determine the motion sensitivity (MS) of the network

$$MS = 1 + \frac{N}{S}$$

where S and N correspond to the signal and noise components of the signal sectors respectively, a least-squares Weibull fit was applied to the average percentage of correct responses as a function of noise level. For each of the five signal areas, the 75% threshold was compared with psychophysical results under mask/no-mask conditions and analysed based on the linear-integrator model proposed by Morrone *et al* (1995) (see also Burr *et al* 1998).

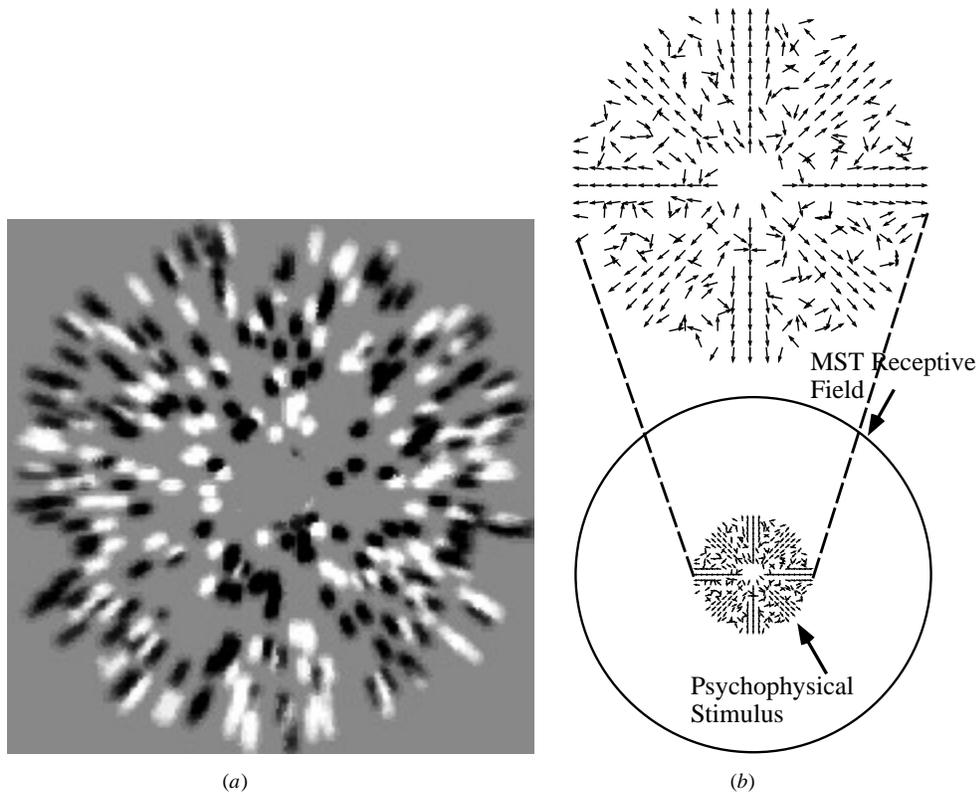


Figure 7. (a) Scaled and time-lapsed example of stimuli used during psychophysical testing. The stimuli contained 360 small black-and-white Gaussian patches (with space constant 0.5°) whose four-frame time course moved in an expanding pattern within a region subtending a visual angle of 10° (central 1.5° removed). (b) Example of a static representation of an expanding motion stimulus (containing eight signal sectors under a masked condition with $MS = 1$) used to simulate the four-frame motion sequence during psychophysical testing of the network. Each stimulus contained 340 motion vectors and was presented in a 21° aperture centred in the 63° receptive field of the hidden layer.

4. Results

4.1. The development of optic flow selectivity in the hidden layer

For each set of simulations, a series of eight-point least-squares Gaussian fits was applied to the hidden-unit responses to optic flow to estimate the mean and standard deviation of each unit's tuning curve, figure 8. For each unit, the eight responses used in the fitting procedure corresponded to the presentation of eight pure cardinal and spiral optic flow stimuli located at 45° intervals in the optic flow space. As in the neurophysiological experiments conducted by Graziano *et al* (1994), Gaussian fits with correlations below 0.9 and/or units exhibiting double-lobed tuning curves were excluded from further analysis. In addition, Gaussian fits with $\sigma < 15^\circ$ were excluded from analysis because of inadequate sampling using the eight-point set. Throughout the hidden-layer analysis, these restrictions excluded $28 \pm 10\%$ of the hidden units in each network.

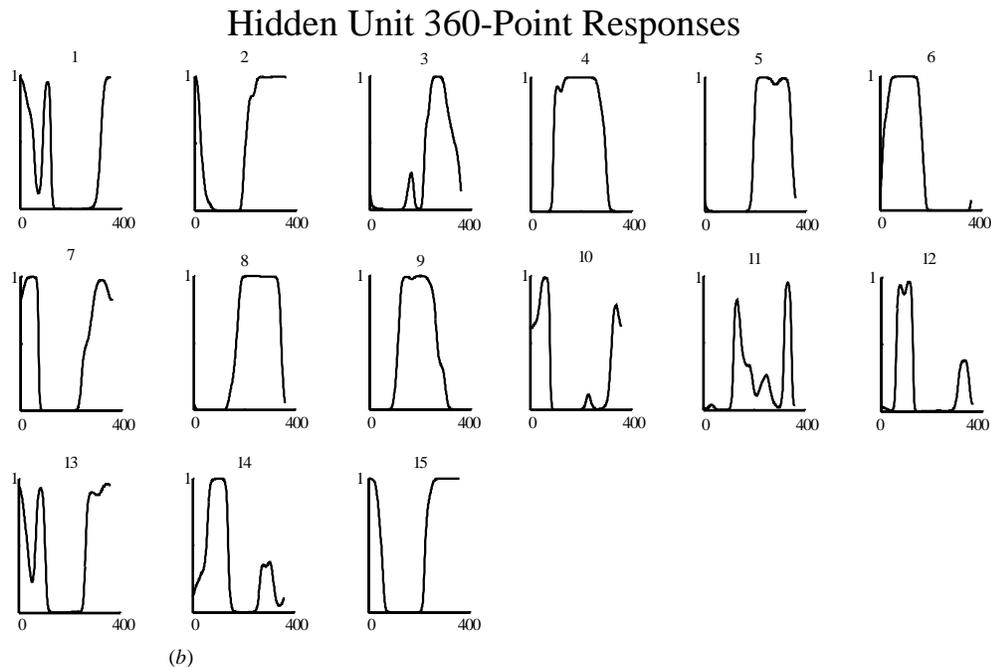
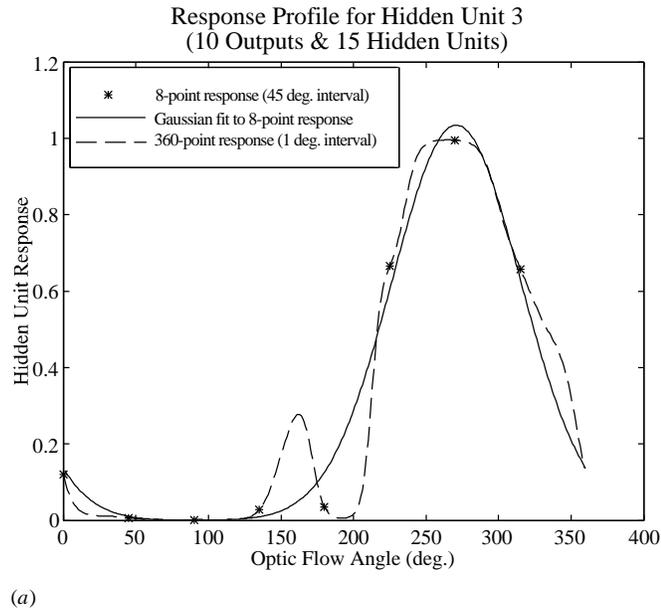


Figure 8. (a) Response profile for hidden unit 3 in a 15 hidden unit/10 output unit network with a uniform distribution of optic flow selectivity in the output layer. Asterisks denote hidden-unit responses to neurophysiological stimuli used to obtain an eight-point least-squares Gaussian fit (dashed line). The solid line represents the actual hidden-unit tuning curve obtained using the 360-pattern test set. (b) Actual hidden-unit tuning curves for the hidden layer of the 15 hidden unit/10 output unit network (note: Gaussian profiles appear rectangular owing to uneven axis scaling). Owing to double-lobed tuning curves and/or Gaussian fits with $r < 0.9$, units 11 and 12 were excluded from additional analysis.

4.1.1. *Optic flow selectivity in the hidden layer using a uniform distribution of optic flow selectivity in the output layer.* Figure 8(a) illustrates a typical eight-point least-squares Gaussian fit for a hidden unit in a 15 hidden/10 output unit network trained under the uniform condition. Across network sizes, approximately 72% of the hidden units developed single-peaked tuning curves which were well approximated by an eight-point Gaussian fit ($r \geq 0.9$). Figure 8(b) illustrates the actual hidden-unit tuning curves of the same network obtained using the 360-pattern test set. Units 11 and 12 were excluded from analysis owing to double-lobed profiles and/or Gaussian fits with $r < 0.9$. Of the remaining units, most exhibited tuning curves which were Gaussian in shape (rectangular in figure 8(b) owing to uneven scaling) while some exhibited complex tuning curves which were well approximated by an eight-point Gaussian fit. Throughout the simulations under the uniform condition, the predominance of tuning curves with eight-point Gaussian structure was consistent over multiple training sessions and across network sizes.

The selectivity of each hidden unit was defined as the mean of the corresponding eight-point fit and a distribution of hidden unit selectivity was plotted in the optic flow space for all units passing the minimum specifications, figure 9. The composite distribution in figure 9(a) was constructed using the individual network distributions in figure 9(b). For the units included in the composite distribution, the average standard deviation and correlation of the fitted Gaussian profiles were $60 \pm 38^\circ$ and 0.96 ± 0.03 , respectively. These results were consistent with those of Graziano *et al* (1994) who reported spiral tuned cells with average σ and r of $60 \pm 34^\circ$ and 0.96 ± 0.03 and cardinal tuned cells with average σ and r of $62 \pm 25^\circ$ and 0.97 ± 0.02 , respectively.

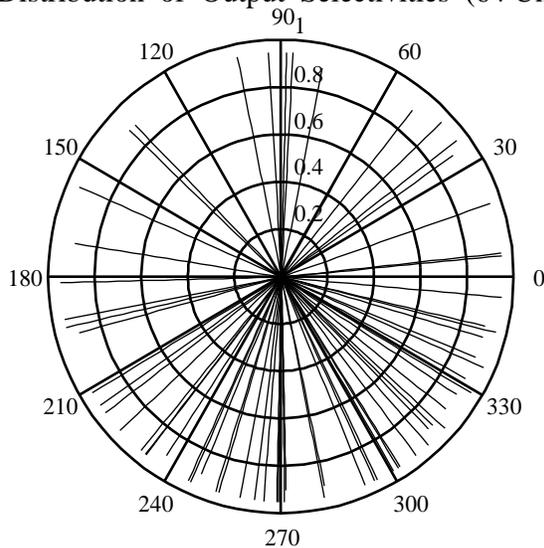
An examination of both the individual and composite distributions (figure 9) reveals units selective to spiral motion. Under the cardinal/spiral segmentation of the optic flow space (figure 4), the selectivity of these units occurs within the spiral regions centred at 45° , 135° , 225° , and 315° in the optic flow space. The distribution of selectivities in the composite plot also indicates a continuum of optic flow selectivity with a slight bias toward clockwise rotation. The presence of the bias is associated with the strong preference for counterclockwise selectivities observed in the (45 hidden unit)/(20 output unit) network (figure 9(b)). Because of the limited population used in the composite plot, it is not clear, based on the networks examined here, whether this bias is an emergent property of the network or a statistical variation from a uniform distribution.

While the continuum of optic flow selectivity observed in the composite and individual plots is consistent with neurophysiological results, the distribution in the hidden layer differs from neurophysiological results which indicate a strong bias toward MST selectivities centred around expanding flow fields. This is not unexpected since the network was trained with stimuli chosen uniformly from the optic flow space. However, based on the results of Zemel and Sejnowski (1998) we expect that had the network been trained using a stimulus set biased for expanding motions, the hidden layer would have developed an expansion bias.

4.1.2. *Optic flow selectivity in the hidden layer using a cardinal distribution of optic flow selectivity in the output layer.* In networks trained under the cardinal condition with a cardinal/spiral segmentation of the flow space, the output layer did not contain units selective to spiral optic flows. This biased the hidden layer against the development of spiral selective units since the simplest solution in the hidden layer under an input/output mapping would be to develop units whose tuning curves clustered in the cardinal regions represented in the output layer.

Analysis of the hidden layer revealed units whose tuning curves were accurately

Gaussian Means and Correlations for Uniform Distribution of Output Selectivities (64 Units)

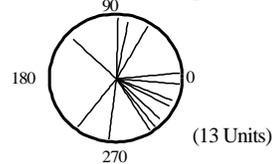


$$\sigma_{av} = 60^\circ \pm 38^\circ$$

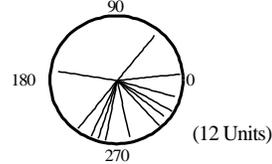
$$r_{av} = 0.96 \pm 0.03$$

(a)

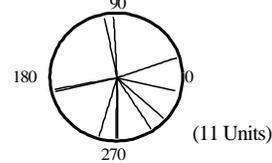
17 Hidden Units w/ 10 Outputs



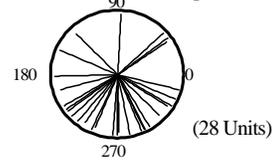
16 Hidden Units w/ 10 Outputs



15 Hidden Units w/ 10 Outputs



45 Hidden Units w/ 20 Outputs



(b)

Figure 9. (a) Composite plot of the distribution of hidden unit selectivities projected onto the optic flow space under the uniform condition. The angular position of each radial line represents the selectivity of a hidden unit in the optic flow space and the radial length represents the correlation of the eight-point Gaussian fit. Based on a cardinal/spiral segmentation of the optic flow space, the networks contain hidden units selective to spiral motion located in the 45° regions centred at 45° , 135° , 225° , 315° . The distribution of hidden-unit selectivity was consistent with a continuum of optic flow selectivity in the hidden layer and exhibited a slight bias toward clockwise rotation resulting from a numerical bias associated with the presence of the 20-output-unit network in the composite plot. (b) The distribution of hidden-unit selectivity for each network under the uniform condition where values in parenthesis indicate the number of hidden units included in each analysis. As in the composite plot, spiral selectivity and a continuum of optic flow selectivity are present across networks.

approximated by an eight-point Gaussian fit and whose structure was consistent with the units illustrated in figure 8. As with simulations under the uniform condition, the Gaussian structure observed in the hidden-unit responses was consistent over multiple training sessions and across network sizes.

Figure 10 shows the distribution of optic flow selectivity obtained with networks trained under the cardinal condition. For the units included in the composite distribution, the resulting average standard deviation and correlation of $52 \pm 36^\circ$ and 0.96 ± 0.03 were generally consistent with neurophysiological results. Despite the bias associated with the cardinal condition, the hidden layers continued to develop units selective to spiral motion. In addition to spiral selectivity, both the composite and individual distributions developed

Gaussian Means and Correlations for Cardinal Distribution of Output Selectivities (71 Units)

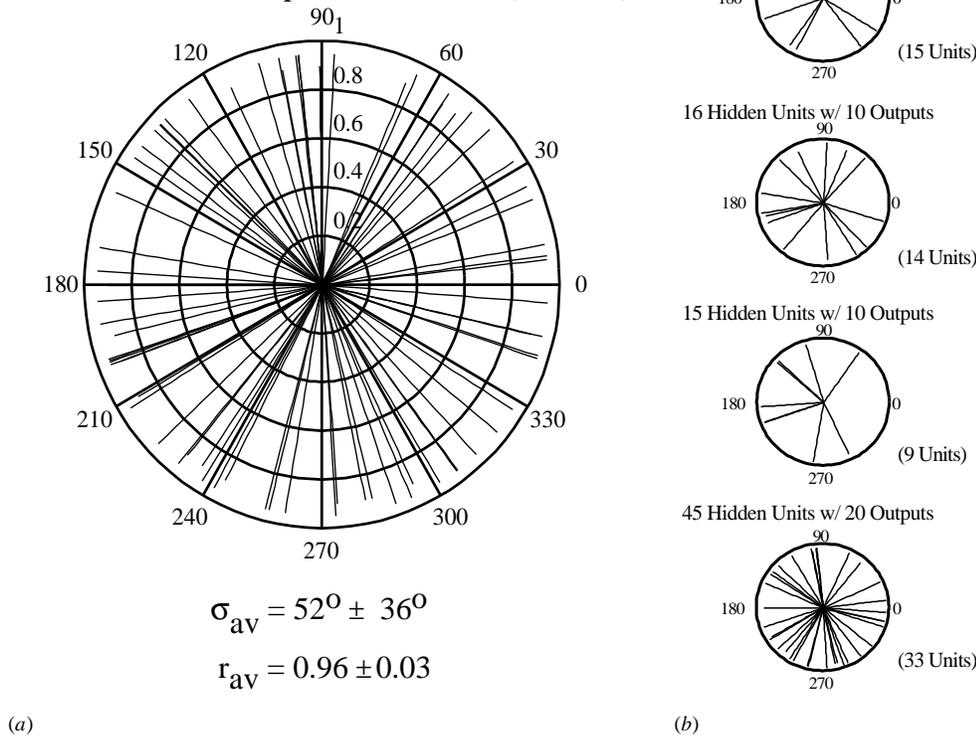


Figure 10. (a) Composite plot of the distribution of hidden-unit selectivities projected onto the optic flow space under the cardinal condition. In spite of the bias against spiral selectivity implicit in the output layer, all networks developed hidden units selective to spiral motion (located in the 45° regions centred at 45° , 135° , 225° , 315° based on a cardinal/spiral segmentation of the optic flow space). Furthermore, the distribution of hidden-unit selectivity is consistent with an unbiased continuum of optic flow selectivity in the hidden layer. (b) The distribution of hidden-unit selectivity for each network under the cardinal condition. Values in parenthesis indicate the number of hidden units included in each analysis. As in the composite plot, spiral selectivity and a continuum of optic flow selectivity were present across networks.

unbiased uniform distributions consistent with a continuum of optic flow selectivity in the hidden layer. Taken together with the minimized nature of the hidden layer, these results suggest that a continuum of optic flow selectivity may be an efficient method of encoding optic flow input from directionally selective units.

4.2. Testing for position invariance in the hidden units

For each set of position-invariance tests conducted with small shifts in the COM ($< 10^{\circ}$), PI indices were calculated at the four shifted positions and examined in a composite histogram format (bin width = 0.1). Hidden units containing PI indices of 1 represented units which were position invariant within the scope of the test. As the PI indices deviated from 1, units responded more (> 1) or less (< 1) to the shifted stimulus than to the centred stimulus. In cases where a unit flipped its selectivity at a shifted position, the PI index became negative.

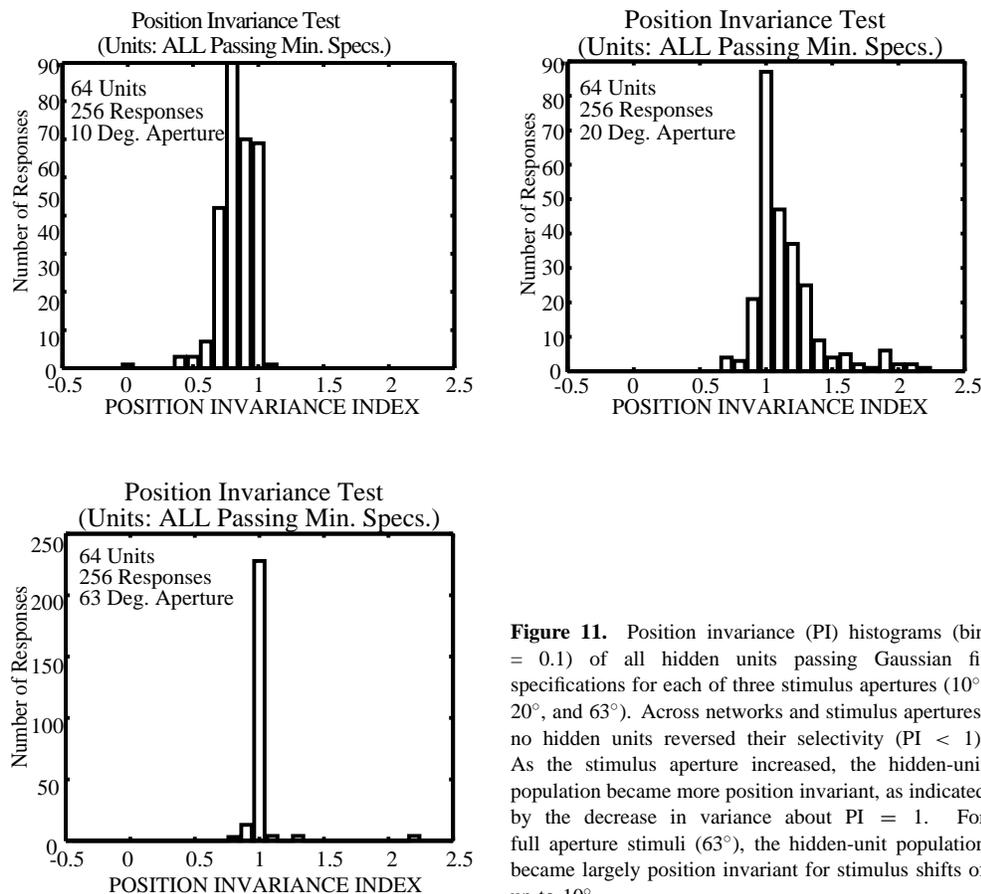


Figure 11. Position invariance (PI) histograms (bin = 0.1) of all hidden units passing Gaussian fit specifications for each of three stimulus apertures (10° , 20° , and 63°). Across networks and stimulus apertures, no hidden units reversed their selectivity ($PI < 1$). As the stimulus aperture increased, the hidden-unit population became more position invariant, as indicated by the decrease in variance about $PI = 1$. For full aperture stimuli (63°), the hidden-unit population became largely position invariant for stimulus shifts of up to 10° .

Figure 11 shows the composite PI histogram results across stimulus aperture for all hidden units passing the Gaussian fit specifications under the uniform condition. As the stimulus aperture increased, the position invariance of the units increased. Under full-aperture (63°) conditions the hidden-unit population became largely position invariant to stimulus shifts of up to 10° . As in neurophysiological results, none of the units reversed their selectivity ($PI < 0$). Under the cardinal condition, PI histograms similar to figure 11 in both distribution and bin amplitude were obtained, but these are not shown here.

When hidden units were tested with larger shifts in the COM ($> 22^\circ$), they exhibited a continuum of position-invariant responses. Figure 12(a)–(c) illustrates several examples, ranging from units with position-invariant responses (12(a)) to those with position-dependent responses (12(c)). Those units which were position invariant to large shifts in the COM exhibited significant planar responses which were often indistinguishable from responses to stimuli with COM shifts of 22° or 45° . While these units did not exhibit the preferred planar motion responses identified in multi-component neurons, their presence is intriguing given that the networks were not trained with planar stimuli. Finally, the average hidden-unit responses to pericentric, eccentric, and planar stimuli (figure 12(d)), showed the hidden layer to be predominately position dependent with respect to COM shifts greater than 22° .

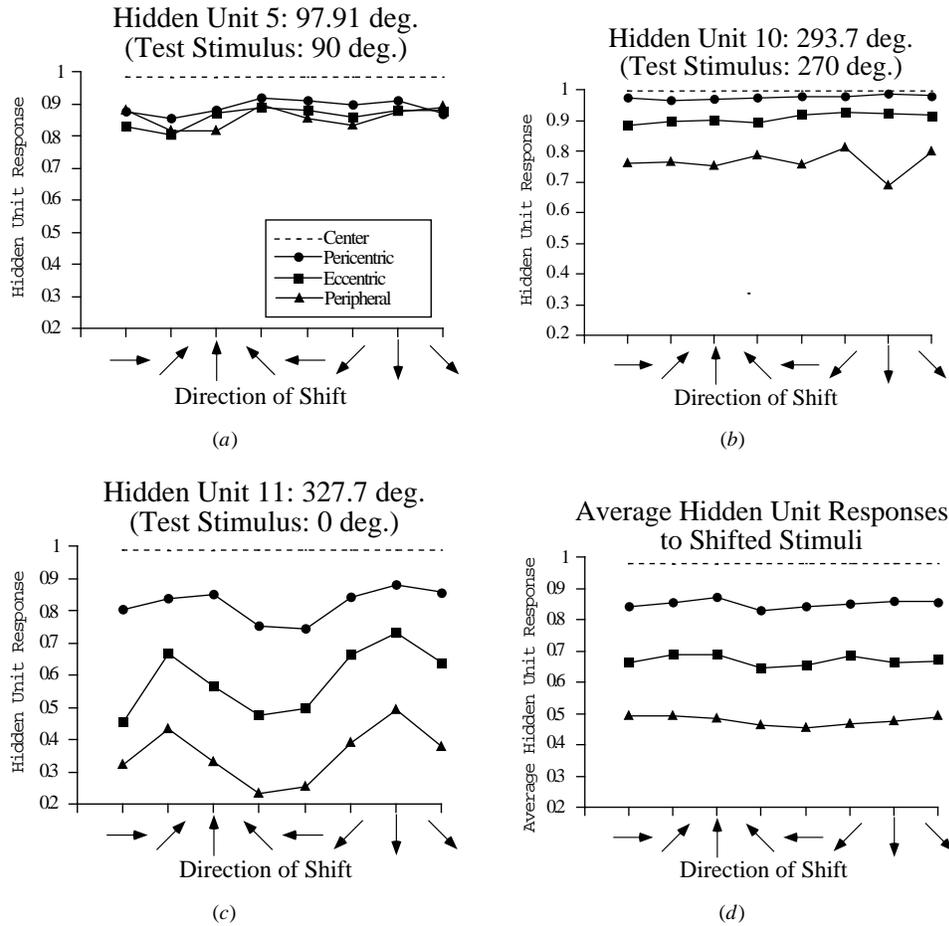


Figure 12. (a)–(c) Response amplitudes of three hidden units spanning the observed range of position invariance in the hidden layer. Each hidden unit was presented with the ‘most preferred’ cardinal set containing 25 stimuli with shifted COMs (figure 6). Each shifted set of eight stimuli was plotted according to the directional shift of the COM (shown on the abscissa) for pericentric (circles), eccentric (squares), and planar (triangles) shifts. Dashed lines represent the hidden unit’s response to the centred stimulus and were always the best response in the set of shifted stimuli. Hidden units were identified which encompassed a range of position invariance from highly position-invariant units (a) to highly position-dependent units (c). Across hidden units, little spatial specificity was observed in each shifted set of eight stimuli. (d) Average hidden-unit responses to the 25-pattern set of shifted stimuli. The average responses in the hidden layer indicate a significant degree of position dependence for COM shifts > 22°.

4.3. Weight analysis

In addition to the examination of optic flow selectivity in the hidden layer, the primary network structure was examined for functional and topographic structure in the weight connections (hidden–output and input–hidden respectively). In figure 13, the weight connections between the hidden and output units were examined for functional structure, under the cardinal condition, using a weighted projection of hidden-unit classifications to output units. Hidden units were grouped into eight classifications according to the cardinal/spiral segmentation of the optic flow space. Classifications were designated by

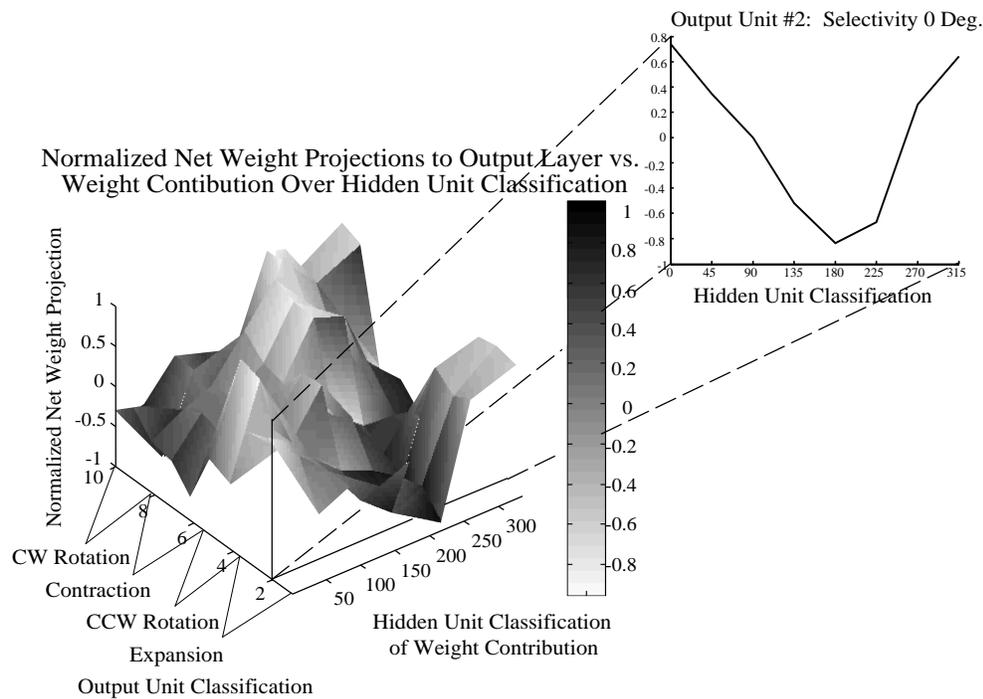


Figure 13. Surface plot showing functional structure in the weight connections between hidden and output units in the 16-hidden-unit network trained under the cardinal condition. Hidden units were grouped into eight classifications represented by their central angle (θ_c) following the cardinal/spiral segmentation of the optic flow space. Output units were classified into one of four cardinal regions and the net weight projection as a function of hidden-unit classification was calculated for each. The diagonal ridge and adjacent troughs indicate the presence of opponency between layers which is further illustrated by the two-dimensional slice corresponding to an output unit selective for expansion. The expansion unit received positive weight projections from expansion-selective hidden units and negative weight projections from contraction-selective hidden units.

the centred flow angle ($\theta_c = 0, 45^\circ, 90^\circ, \dots, 315^\circ$) of each region. Output units were classified into one of the four cardinal regions and the net weight projection to each output unit was calculated as a function of the eight hidden unit classifications. Figure 13 shows the three-dimensional surface formed across output classification. The ridge and adjacent troughs running diagonal to the classification axes illustrate the presence of opponency between layers. This is seen more clearly in the two-dimensional slice of an output unit selective for expansion. For this unit, the hidden units selective for expansion projected a large positive net weight while units selective for contraction projected a large negative net weight.

In the weights of the input–hidden layer, topographic structure was examined under the cardinal condition as a function of the weighted input activation mapped onto the input space. For each hidden unit the sum of weighted inputs was calculated for the 67 MT receptive fields using the ‘most preferred’ cardinal stimulus determined from the position-invariance tests. Figure 14 shows a typical surface map of the input activation to a hidden unit with the net activation from each MT receptive field projected onto its receptive field centre. Throughout all simulations, the input activation projected to the hidden units exhibited a

Net Input Activation to Hidden Unit #1 Mapped to the Input Coordinates of the MT Receptive Field Centers

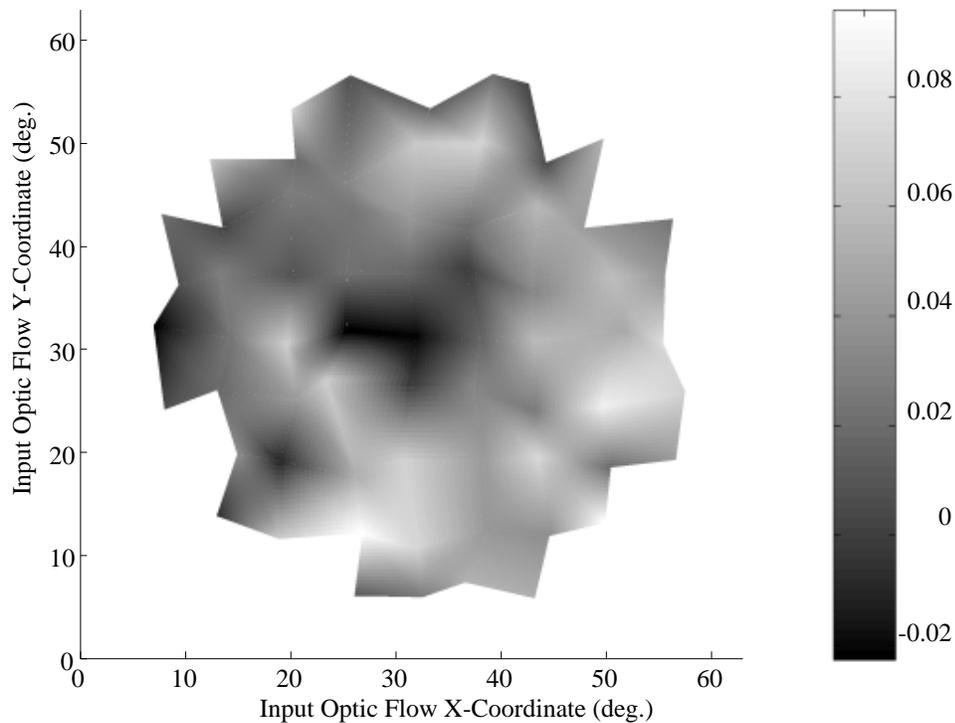


Figure 14. Surface map showing the topographic structure of the net weight connections to a hidden unit. For the hidden unit the sum of weighted inputs was calculated for each of the 67 MT receptive fields using the ‘most preferred’ cardinal stimulus determined from the position-invariance tests. The surface map was generated by projecting the summed weighted input from each MT receptive field onto its respective centre. Throughout network simulations, the hidden units exhibited a centre–surround topography with the net weighted input of the central 5–30° a factor of two to three lower than the periphery.

centre–surround topography with the central 5–30° exhibiting a significantly lower activation than the periphery. As figure 14 illustrates, input units with more central receptive field positions exhibited net activations near zero while units whose receptive fields were located peripherally in the input field exhibited significant positive net activations.

This reduced area of central activation helps to account for the position-invariant nature of the hidden layer to small shifts in the COM. As stimulus aperture increases, the peripheral activation progressively dominates the stimulus input. Since changes in the motion vectors of the stimulus periphery decrease with increasing aperture, the variability of hidden-layer responses to COM shifts also decreases. By definition, this decreased variability increases the position invariance of the hidden layer (figure 11).

4.4. Psychophysics

Figure 15 shows radial and rotational motion sensitivity as a function of stimulus area for the extended network and the experimental results of Morrone *et al* (1995). The dashed

lines represent the motion sensitivity of an ideal integrator model for complex motion discrimination. Under the no-mask condition the network agreed closely with experimental results and was generally consistent with an ideal integrator model. Under the mask condition, the motion sensitivity slope increased slightly from the no-mask condition but did not follow the increased slope predicted by an ideal integrator. The absence of a significant increase in the motion sensitivity slope over stimulus area under the mask condition indicates that the network was not affected by the presence of motion noise in the mask sectors.

To determine how the network 'ignored' the noise sectors under the mask condition, the weight distribution in the input-hidden layer connections was analysed by plotting a histogram (30 bins) of the weight values to each hidden unit. The weighted projections to the hidden layer were well approximated by Gaussian distributions with means centred near zero (± 0.0011) and an average standard deviation of 0.0048. During presentations of stimuli with varying degrees of noise, the multiplication of positive noise inputs with the weighted

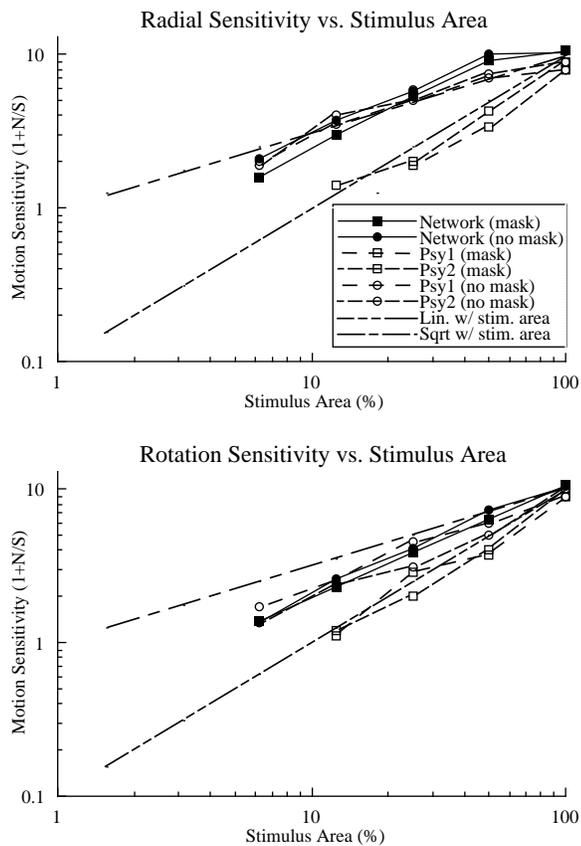


Figure 15. Network motion sensitivity for discriminating radial and rotational motion compared with experimental results from Morrone *et al* (1995). Dashed lines represent the motion sensitivity of an ideal integrator model of complex motion detection. Under the no-mask condition, the network results agreed closely with the experimental results and were generally consistent with an ideal integrator model. Under the mask condition, the network results decreased slightly from the no-mask condition and did not follow the increased slope of motion sensitivity predicted by an ideal integrator. The results under the mask condition indicate that the network was not affected by the presence of motion noise in masked sectors.

connections produced weighted noise inputs which were also well approximated by zero-mean Gaussian distributions. Under these conditions the input contribution of the summed weighted noise to the hidden layer was approximately zero prior to the activation function and its effect on the network motion sensitivity was minimized. For rotation sensitivity the weight distributions were closely centred near zero. This increased the noise cancellation and produced virtually no change between the mask and no-mask results with increasing noise. In contrast, for radial sensitivity, the contributing weight distributions showed more variance in the mean. This reduced the noise cancellation allowing the mask and no-mask results to diverge with increasing noise.

Throughout training and testing, the network measurements of MS as a function of signal area remained robust to changes in a variety of parameters including the distribution of noise used to simulate the internal noise of the observer, the range of noise levels presented in the training stimuli, the presence of internal noise during training, and changes in the threshold used to calculate the MS. Since several of these parameters characterized the decision layer, their inability to affect the motion sensitivity of the network significantly suggests that the decision layer did not play a significant role in the psychophysical responses. Instead, the decision layer acted only to reduce the amount of post-processing necessary to obtain psychophysical responses from the primary network.

5. Discussion

In sections 4.1–4.3 we presented results and analysis for the properties of hidden units in the primary network. These results include the types of complex motion tuning observed in the hidden layer (section 4.1), which we discuss in section 5.1, and an analysis of the weights and position-invariance properties (sections 4.2 and 4.3), which we discuss in section 5.2. We also presented psychophysical results of complex motion sensitivity for the extended network in section 4.4 which we discuss in section 5.3 in the context of the input–hidden layer weight structure.

5.1. The development of optic flow selectivity in the hidden layer

During network training and testing several primary network parameters including network size, input response profiles, and output response biases were varied across implementations of the network. In the primary network, the hidden layers developed units with preferred responses to optic flow consistent with the multi-component (Duffy and Wurtz 1991a,b) and continuum of single-component (Graziano *et al* 1994) methods of classification used to characterize MSTd neurons. In addition to the single-component units identified in neurophysiological results, a large proportion of hidden units developed stimulus selectivities characterized by multi-component radio–circular responses. Consistent with findings in MSTd neurons (Graziano *et al* 1994), the hidden units classified as radio–circular exhibited ‘most preferred’ responses to the intermediate spiral stimuli formed by the scalar product of the respective radio-circular motions. Since translational motions were not included during network training, multi-component classifications including translation were not specifically tested. However, the position-invariance tests for large COM shifts did show units with significant responses to multiple directions of translational motion.

Analysis of the hidden-layer responses to stimuli spanning the optic flow space identified a wide range of complex response profiles. Over 70% of the hidden units developed near-Gaussian responses to optic flow stimuli producing good Gaussian fits ($r \geq 0.9$) to the eight-pattern stimulus set used by Graziano *et al* (1994) with average standard deviations

and correlations which were in good agreement with neurophysiological results. For these units, a more extensive sampling of the optic flow space revealed a subset of units which contained secondary peaks not detected with the eight-pattern stimulus set. The structure of the secondary peaks varied considerably in amplitude but typically exhibited standard deviations less than 15° .

The development of secondary peaks in the hidden layer is interesting but not unexpected. While we have constrained the network through minimization of the hidden layer to optimally encode the input to the output we have not imposed any explicit constraints to force Gaussian responses in the hidden layer. In its current implementation we have allowed the network a significant amount of freedom in developing suitable weight structures. Because the network uses a small training set to constrain a large number of weights, we suggest the secondary peaks may result from an overfitting of the computational representation in the hidden layer.

In this context, the fact that the hidden units developed responses well approximated by Gaussian fits is actually more significant than the presence of secondary peaks. Although the variety of tuning profiles is reasonable based on the network structure, the ability to fit them accurately to Gaussian profiles by rough sampling of the optic flow space is interesting and suggests that the tuning profiles reported in MST need not be fully Gaussian throughout the stimulus space.

Finally, an analysis of optic flow selectivity in the hidden layer across networks revealed a continuum of preferred responses. The hidden units in all network simulations developed a continuum of preferred responses to optic flow stimuli, regardless of the biases associated with the specification of preferred responses in the output layer. Specifically, when the output layer was biased to contain units with cardinal selectivities, the hidden layer continued to develop a continuum of preferred responses including preferences for spiral motion. These observations support results reported by Graziano *et al* (1994) and suggest that a continuum of optic flow selectivity may be an efficient method of encoding the visual motion components of optic flow.

5.2. Testing for position invariance in the hidden layer

As with the development of hidden-unit selectivity in the network, position-invariance testing in the hidden layer identified units with characteristics consistent with the neurophysiological results reported by Duffy and Wurtz (1991a,b) and Graziano *et al* (1994).

For small shifts in the stimulus COM ($< 10^\circ$), the hidden units produced position-invariant responses which increased with stimulus aperture. Network analysis suggests that the increased position invariance resulted from a combination of the computational increase in input and the emergent weight structure in the network.

Because the hidden units contained logistic activation functions, small changes in the weighted input tended to produce large changes in the output near the maximum slope of the activation function. As stimulus aperture increased, the input to the network increased proportionally with area. This caused the output to shift to regions of the logistic with decreased slope, reducing sensitivity to small changes in the input. As a result, for small changes in the input associated with shifting the COM, the variability in the output decreased causing an increase in position invariance with aperture.

In addition to the position invariance associated with the increased input, the emergent weight structures in the input–hidden layer connections also contributed significantly to the position invariance of the hidden layer. During training, weight structures developed which placed more emphasis on input from the periphery of the network receptive field. As the

stimulus aperture increased, the weighted input from the periphery increasingly dominated the input to the hidden layer. Since the periphery of the optic flow stimulus did not change significantly for small shifts in the COM, the bias toward peripheral weighting caused a significant increase in the position invariance of the hidden layer.

For larger shifts in the COM ($> 22.5^\circ$), the degree of position invariance in the hidden layer spanned a continuum of responses with a predominantly position-dependent response profile. In a manner similar to that for small shifts, the resulting position dependence was caused primarily by changes in the preferred optic flow of the periphery. As the COM shifted toward the periphery and beyond, the peripheral optic flow dominating the input changed significantly from the preferred centred stimulus. This decreased both the output from the hidden layer and, as a result, the position invariance.

Although a comparison of the neurophysiological results (Duffy and Wurtz 1995) with the hidden layer responses for large shifts in the COM reveals differences in the degree of spatial specificity, these appear to be due to the form of stimulus presentation. In the neurophysiological experiments, most cells exhibited significant spatial specificity to the direction of shift when shown stimuli centred in the monkey's field of view, while in the network, hidden-unit responses remained largely independent of the direction of shift with stimuli centred in the unit's receptive field. Owing to the structure of the current network, stimuli could not be presented in a view-centred frame which did not coincide with the receptive field of units in the hidden layer. Had network stimuli been presented in a view-centred frame which did not coincide with the network receptive field, the hidden-unit responses would have been more consistent with neurophysiological results in MSTd showing increased spatial specificity for directions of COM shift.

5.3. Psychophysics

The motion sensitivity of the extended network was consistent with psychophysical results from human subjects in the no-mask condition and its performance was significantly better than psychophysical results in the masked condition. The linear increase in motion sensitivity (as a function of the square root of the signal area) observed in the no-mask conditions suggests that the network processed motion in a manner analogous to an ideal integrator under no-noise conditions (Morrone *et al* 1995). In the mask condition, the zero-mean Gaussian distribution of weights in the input-hidden layer connections caused the network to 'ignore' sectors containing noise. As a result, in the mask condition the overall motion sensitivity did not follow the linear increase with stimulus area observed in human subjects and predicted by a linear-integrator model.

Although the network results in the mask condition were not consistent with psychophysical results, the network did integrate local motions into a global motion percept as observed in the human results. Based on the weight analysis, the discrepancy in the mask condition arose primarily from a lack of constraints on the weights during learning. In the network, the distribution of positive (excitatory) and negative (inhibitory) weight values was freely determined within the constraint imposed by the output error. In the case of our network, this resulted in a roughly equal distribution of excitatory and inhibitory weighted connections which are not consistent with the predominantly excitatory connections in cortex. Because inputs were summed prior to the activation function, the resulting zero-mean weight distribution allowed the network to 'ignore' noise sectors in the mask condition even though the local motion information to the network was integrated into a psychophysical global motion percept in the decision layer.

6. Conclusion

Although the extensive recurrent information used while training the network precludes its literal application to learning in the brain, previous computational models of higher cortical areas utilizing back-propagation have been shown to discover algorithms and develop network interactions similar to those employed in cortex (Zipser and Anderson 1988, Zipser 1990, Sharon *et al* 1995). Similarly, our network has used back-propagation to discover network properties which suggest that the structure of the network in conjunction with the learned weight connections captured many of the salient features of neurons in MSTd. Analysis of the network has shown that units in the hidden layer developed complex motion properties consistent with a variety of neurophysiological results in MSTd. Specifically, the network developed units with multi-component classifications (Duffy and Wurtz 1991a,b) which were more accurately labelled using a continuum of classifications for cell specificity proposed by Graziano *et al* (1994). Testing with shifted COMs showed that the hidden layer developed position-invariant properties which were consistent with MSTd for both small ($< 10^\circ$, Graziano *et al* 1994) and large shifts ($> 22^\circ$, Duffy and Wurtz 1995) in the COM. During psychophysical tasks measuring motion sensitivity to complex motion stimuli, the network responses met or exceeded human performance and that predicted by a simple ideal integrator model. In the context of applying physically more valid optic flows, it would be interesting to examine these issues further using more natural stimuli similar to those of Zemel and Sejnowski (1998) which encompass larger fields of view and contain both observer and object motion.

In addition to being consistent with neurophysiological results in MSTd, the combination of properties observed in the hidden layer suggest a possible role in heading and navigation. Since it is postulated that neurons in MSTd contribute to heading and navigation, the specificity of network units for complex motion stimuli in conjunction with position-invariant responses to small shifts in the COM and increasing position dependence for larger shifts, suggest that individual units could be used to code roughly for heading. Taken in the context of a population coding mechanism, these units could be combined to obtain refined estimates of heading and relative motion for use in navigation.

Acknowledgments

We would like to thank Venkataraman Sundareswaran, Robert Pitts, and Colin Clifford for their helpful suggestions on this work. This work was supported by NIH grant EY-2R01-07861-08 to LMV

References

- Albright T D 1984 *J. Neurophysiol.* **52** 1106–30
- Albright T D and Desimone R 1987 *Exp. Brain Res.* **65** 582–92
- Andersen R A 1997 *Neuron* **18** 865–72
- Beardsley S A and Vaina L M 1997 *Invest. Ophthalmol. Visual Sci.* **38** 392
- Beardsley S A, Vaina L M and Poggio T 1996 *Soc. Neurosci. Abstr.* **22** 635.11
- Beverley K I and Regan D 1979 *Vision Res.* **19** 1093–104
- Burr D C, Morrone M C and Vaina L M 1998 *Vision Res.* **38** 1731–43
- Duffy C J and Wurtz R H 1991a *J. Neurophysiol.* **65** 1329–45
- 1991b *J. Neurophysiol.* **65** 1346–59
- 1995 *J. Neurosci.* **15** 5192–208
- 1997 *J. Neurosci.* **17** 2839–51

- Freeman T C A and Harris M G 1992 *Vision Res.* **32** 81–7
- Graziano M S A, Anderson R A and Snowden R J 1994 *J. Neurosci.* **14** 54–67
- Heeger D J and Jepsen A D, 1990 *Neural Comput.* **2** 129–37
- Lagae L, Maes H, Raiguel S, Xiao D K and Orban G A 1994 *J. Neurophysiol.* **71** 1597–626
- Lappe M and Rauschecker J P 1993 *Neural Comput.* **5** 374–91
- 1995 *Biol. Cybern.* **72** 261–77
- Maunsell J H R and Van Essen D C 1983 *J. Neurophysiol.* **49** 1127–47
- Morrone M C, Burr D C and Vaina L M 1995 *Nature* **376** 507–9
- Orban G A, Lagae L, Verri A, Raiguel S, Xiao D, Maes H and Torre V 1992 *Proc. Natl Acad. Sci., USA* **89** 2595–9
- Perrone J A and Stone L S 1994 *Vision Res.* **34** 2917–38
- Pitts R I, Sundareswaran V and Vaina L M 1997 *Computational Neuroscience: Trends in Research 1997* (New York: Plenum) pp 171–6
- Rodman H R and Albright T D 1987 *Vision Res.* **27** 2035–48
- Rumelhart D E, Hinton G E and Williams R J 1986 *Parallel Distributed Processing* ed D E Rumelhart *et al* (Cambridge, MA: MIT Press) pp 318–63
- Saito H A, Yukie M, Tanaka K, Hikosaka K, Fukada Y and Iwai E 1986 *J. Neurosci.* **6** 145–57
- Sharon M, Deligeorges S, Kacher D, Sundareswaran V and Vaina L M 1995 *Invest. Ophthalmol. Visual Sci.* **36** 1780
- Smith A T, Snowden R J and Milne A B 1994 *Vision Res.* **34** 2425–30
- Snowden R J and Milne A B 1996 *J. Cognitive Neurosci.* **8** 412–29
- Tanaka K, Fukada Y and Saito H 1989 *J. Neurophysiol.* **62** 642–56
- Tanaka K, Hikosaka K, Saito H, Yukie M, Fukada Y and Iwai E 1986 *J. Neurosci.* **6** 134–44
- Tanaka K and Saito H 1989 *J. Neurophysiol.* **62** 626–41
- Te Pas S F, Kappers A M L and Koenderink J J 1996 *Vision Res.* **36** 259–70
- Wang R 1995 *Neural Comput.* **7** 290–306
- 1996 *Neural Networks* **9** 411–26
- Watamaniuk S N J and Sekuler R 1992 *Vision Res.* **32** 2341–7
- Zhang K, Sereno M I and Sereno M E 1993 *Neural Comput.* **5** 597–612
- Zemel R S and Sejnowski T J 1998 *J. Neurosci.* **18** 531–47
- Zipser D 1990 *Neuroscience and Connectionist Theory* eds M A Gluck M A and D E Rumelhart D E (Hillsdale, NJ: Erlbaum) pp 355–83
- Zipser D and Andersen R 1988 *Nature* **331** 679–84