Communicated by Steven J. Nowlan

# A Simple Competitive Account of Some Response Properties of Visual Neurons in Area MSTd

**Ruye Wang**
*Department of Engineering, Harvey Mudd College, Claremont, CA 91711, USA*

**A simple and biologically plausible model is proposed to simulate the optic flow computation taking place in the dorsal part of medial superior temporal (MSTd) area of the visual cortex in the primates' brain. The model is a neural network composed of competitive learning layers. The input layer of the network simulates the neurons in the middle temporal (MT) area that selectively respond to the visual stimuli of the input motion patterns with different local velocities. The output layer of the network simulates the MSTd neurons that selectively respond to different types of optic flow motion patterns including planar, circular, radial, and spiral motions. Simulation results obtained from this model show that the behaviors of the output nodes of the network resemble very closely the known responsive properties of the MSTd neurons found neurophysiologically, such as the existence of three types of MSTd neurons that respond, respectively, to one, two, or three types of the input motion patterns with different position dependences, and the continuum of response selectivity formed by the three types of neurons.**

## 1 Introduction

Optic flow plays an important role in the perception of three-dimensional (3D) motion. When there exists a relative motion between an observer and the surrounding world, either some objects moving relative to the static observer, or, in the case of ego-motion, the observer moving in a stationary world, information about this overall motion can be extracted from the optic flow by the brain allowing for proper reaction.

Neurophysiological studies have found (Saito *et al.* 1986; Tanaka and Saito 1989a,b; Graziano *et al.* 1990; Duffy and Wurtz 1991a,b; Graziano *et al.* 1994) that most of the neurons in the dorsal part of the medial superior temporal (MSTd) area of the visual cortex in the primates' brain are responsive to several types of optic flow stimuli, called motion components, including radial, circular, and circuloradial (spiral) motions centered at different locations, and translational motions of different directions on the frontoparallel plane. Also, there exist three types of MSTd

 

cells that respond to one, two, or three of the motion components. They are referred to as single-, double-, and triple-component cells, respectively. The selective responses of some MSTd cells are position dependent, while those of others are position independent.

It has also been found that MST receives strong fiber projection from the middle temporal (MT) area (Maunsell and Van Essen 1983a,b; Ungerleider 1986) where the neurons are selectively responsive to the orientation and velocity of the visual stimuli (Albright 1984; Rodman and Albright 1987). Some of the MT neurons show a marked anisotropy in motion responses favoring directions oriented away from the center of gaze (Albright 1989). It is therefore natural to assume the MT area to be the preprocessing stage of the optic flow processing taking place in MSTd area.

To understand the mechanism of the MSTd neurons and to explain how the motion information is extracted from the optic flow along the visual processing pathway in the primates' brain, various hypotheses have been proposed, all based on the availability of the velocity selectivity from the MT cells as the input to the MST cells. A simple hypothesis was proposed by Tanaka and Saito (1989b, Fig. 12) where an MST cell responsive to one of the motion components receives the synaptic inputs from a set of directionally selective MT cells arranged in accordance with the pattern of that motion component. The input MT cells would be arranged radially in the case of an expansion/contraction MST cell, or arranged circularly in the case of a rotation MST cell. A similar hypothesis proposed by Saito *et al.* (1986, Fig. 14) further assumes that the receptive field of an MST cell responsive to circular or radial motions is composed of a set of overlapping compartments, each of which is in turn composed of a set of MT cells arranged in accordance with the preferred motion component of the MST cell. This assumption can explain the positional independence demonstrated by some of the MST cells. A neural network model based on Hebbian learning proposed by Zhang *et al.* (1993) can account for the position-independent responses of some of the MST cells.

Two different hypotheses called the *direction mosaic hypothesis* and the *vector field hypothesis* are proposed by Duffy and Wurtz (1991b, Fig. 1A and B). The direction mosaic hypothesis is essentially the same as the hypotheses discussed above, i.e., a set of directionally selective MT cells is arranged in a certain pattern to fit the particular motion component to which the MST cell responds; whereas the vector field hypothesis assumes that an MST cell is composed of many units that are distributed throughout the receptive field and are responsive to the same type of motion (but on a smaller scale) as the MST cell. Contradictory predictions may be made based on the two different hypotheses. The direction mosaic hypothesis is more consistent with position-dependent responses, and, therefore, can better explain the triple-component MST cells, whereas the vector field hypothesis is more consistent with position-

independent responses and can better explain the single-component MSTd cells. However, neither hypothesis is adequate to explain the mechanism of all MSTd cells.

While each of the above hypotheses can explain some of the observed properties of the MSTd neurons, they all have the weakness that other important properties are not accounted for. Most obviously, it is very difficult to explain why a multiple-component MSTd neuron can respond to different types of motion components, and why some MSTd cells have position-dependent responses and some have position-independent responses.

In this paper we propose a new model for the MSTd neurons. This model is biologically plausible, as it is based on a simple competitive learning network with unsupervised learning capability, and it can explain all of the major properties of the MSTd neurons found neurophysiologically. In the next section, we give detailed discussion of this model. In Section 3, we show the simulation results and compare the performances of the model with the biological properties of the MSTd neurons.

## 2 Competitive Learning Model for MSTd

The main structure of the model is a two-layer neural network composed of input layer simulating the MT neurons, and output layer simulating the MSTd neurons. The basic operation between the layers of the network is competitive learning (Rumelhart and Zipser 1985), through which the network can discover the salient features and use them to classify the input patterns.

**2.1 Balancing the Competitive Learning.** A competitive learning network can be trained by repeatedly presenting a set of patterns to the input units of network, and iteratively modifying the weights of the winning nodes. The trained network is able to recognize the structure of the input patterns in the sense that each cluster of similar patterns will always excite one particular output unit, and inhibit the others. There may also exist some output units that never win a competition and therefore never learn. These units are called dead units since they are never turned on to respond to any input. In some situations it may be the case that the input patterns do not form a set of nicely separable clusters. For example, the input patterns may form a continuum such that no obvious boundaries can be found to partition the continuum into clusters. This situation will be encountered later in the discussion of our MSTd model. In this case, the results of competitive learning may be very different, anywhere between two extremes: (1) the continuum of input patterns may be divided relatively evenly, but randomly, into several clusters, each represented by an output unit, or (2) the entire continuum is represented by one output unit and all other units become dead units and never turned on.

To achieve the preferred result of (1), the competitive learning can be modified to ensure some winning chance for all units. By including a bias term in computing the output of each node during competition, we can make winning harder for frequent winners and easier for frequent losers (Grossberg 1976, 1987). We can also adjust the learning rate so that the frequent winners learn more slowly than the frequent losers. As another way to ensure equal winning opportunity, DeSieno (1988) proposed a method that adds "conscience" to the competitive learning to reduce frequent winners' winning rate. Specifically, a bias term proportional to the difference between the equal winning probability and the actual winning frequency is used to enforce the equal winning opportunity among all output nodes. By adjusting the proportional factor, we can control the balance of the competitive learning. All of these methods are used in our model. By adjusting the relevant parameters the performance of the competitive learning can be controlled.

**2.2 Detecting the Optic Flow Components.** The input layer, as shown in Figure 1, is composed of MT nodes, each responding preferentially to a local translational motion of a particular direction. The visual area under consideration can be considered as being composed of $k \times k$ small patches of the same size as the receptive field of the MT cells. Each patch is represented by eight MT cells of eight different preferred motion directions (E, NE, N, NW, W, SW, S, SE). If a motion is detected, one of the eight nodes whose preferred direction is closest to the motion direction will be turned on. Otherwise, all nodes are off. The output layer is composed of a set of groups each containing $n$ MSTd nodes, which are fully connected to all of the MT nodes in the input layer. In other words, the receptive field of an MSTd node is as large as the entire visual area represented by the input layer of the network. The number $k$ can be chosen so that it properly relates the size of the receptive field of MT cells (input nodes) and that of the MSTd cells (output nodes). To simplify the model, only the directional flow field is used to represent the current scene. The speed tuning characteristic of MT cells is not modeled here.

Through competitive learning, each node in a group of the output layer learns to respond to one of the different motion patterns presented in the input layer of the network. Among all possible patterns, we are interested only in those patterns that represent local optic flow patterns such as those shown in Figure 2. After training, each of the motion patterns will be responded to by a unique node in each group of the output layer.

The discussion above is based on the ideal situation where the motion stimuli of the optic flow are presented accurately at every point in the visual field. This assumption is not realistic because there exists noise of various types in a real image. First, due to the aperture problem, the apparent velocity an MT cell sees may not represent the true motion. In addition, MT cells are not sharply tuned in motion direction (even though

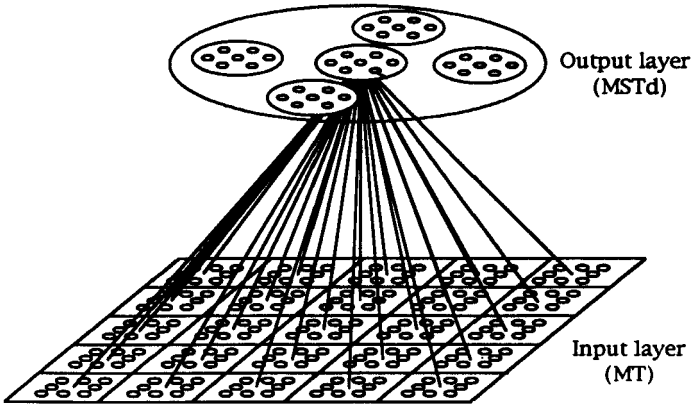


Figure 1: Configeration of the network. The input layer is composed of $k$ by $k$ patches ($k = 5$ in this figure), each represented by eight MT nodes. The output layer is composed of a set of groups (five groups in this figure) each containing $n$ MSTd nodes ($n = 7$ in this figure), which are fully connected to all of the MT nodes in the input layer. (Only the connections of one MSTd node are shown.)

most of them show some speed preference). According to Rodman and Albright (1987), the width of the direction tuning curve could be as wide as 90°. To account for this type of noise, we randomly choose some input nodes and change their direction by a random value in the range of −90° to 90°. Moreover, noise is also introduced when there exist homogeneous areas in the scene. The MT cells in these areas will not turn on because no gradient of brightness can be detected. This situation is simulated by setting some randomly chosen input nodes to zero.

The performance of the competitive learning will get worse as a higher percentage of input MT nodes is contaminated by the two types of noise introduced. However, when the percentage of these nodes is lower than 50%, and with a longer training time, the learning can still reach a stable state where each input motion pattern is represented by basically the same output nodes.

By repeatedly presenting different input motion patterns to the input layer, the MSTd nodes in the output layer learn to respond to the optic flow motion patterns. It is important to note here that the motion patterns presented to the input layer do not form a set of separable clusters in the feature space. For example, two circular motion patterns may be very similar to each other, if their center locations are close to each other.

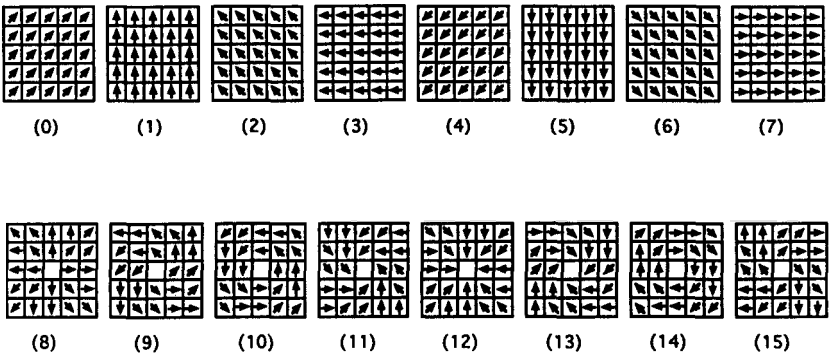


Figure 2: Optic flow patterns. The eight patterns in the first row are translational motions of eight different directions. The eight patterns in the second row are rotations (clockwise and counterclockwise), extractions, contractions, and spiral motions of different circular and radial directions. The orders of the patterns in both rows are arranged so that they form a periodic spectrum in the sense that each pattern can be obtained by counterclockwise rotating 45° the arrows in the pattern to its left, and the first patterns (0 and 8) can also be obtained the same way from the last patterns (7 and 15). Each of the patterns in the second row may have different center locations in the visual field. (This figure shows only those patterns whose center locations are in the center of the field.)

The central areas of these two patterns are represented by different nodes in the input layer, but the peripheral areas of the patterns may be represented by the same input nodes. Also we note that the neighboring patterns in the second row of Figure 2 may share some input nodes if their center locations do not coincide. Moreover, the boundaries between different types of motion patterns will be further blurred due to the existence of various types of noise as discussed above. In other words, the circular, radial, and spiral motion patterns with different center locations form a continuum, rather than separable clusters, in the feature space. Since there do not exist clearcut boundaries among these patterns, the clustering as the result of the competitive learning may not be predictable. The continuum of patterns may be divided randomly in different ways into different numbers of clusters, each of a different size. Moreover, it is also possible that two or even three different types of motion patterns are classified into one cluster and represented by the same output MSTd node because they share some nodes in the input layer. It is this learning mechanism that enables this model to simulate many important responsive features of the MSTd cells, such as the multicom-

ponent cells and their different position-dependent responses. These will be further discussed in the next section.

## 3 Simulation Results

We now present the simulation results of our network model and compare them with the responsive features of the MSTd neurons found neurophysiologically. The key features of MSTd responses are (Graziano *et al.* 1990; Duffy and Wurtz 1991a,b; Graziano 1994, etc.):

1. The receptive fields (ranging from 10 to 100°) of MSTd cells are much larger than that of the MT cells.
2. MSTd cells respond to different types of motion stimuli: unidirectional translations (planar motion), clockwise and counterclockwise circular motion (rotation), outward and inward radial motions (expansions and contractions), and various spiral motions (clockwise or counterclockwise, inward or outward).
3. There exist three types of MSTd neurons that respond to one, two, or three motion components, respectively. The double-component cells can be planocircular or planoradial (Duffy and Wurtz 1991a), or circuloradial (spiral cells) (Graziano *et al.* 1990; Graziano 1994).
4. The three types of neurons do not form three discrete classes but rather a continuum of response characteristics (Duffy and Wurtz 1991a; Graziano 1994, Fig. 6).
5. The responses of those MSTd cells that respond to circular, radial, and/or spiral motions can be plotted as a function, called a tuning curve, of the eight different types of motions (circular, radial, and spiral of different directions) that form a periodic horizontal axis. The tuning curve can be fitted with a gaussian curve very well (Graziano 1994, Fig. 7). For example, if a cell responds most strongly to a clockwise, contractive spiral motion (pattern 13 in Fig. 2), it will also respond (although less strongly) to the neighboring motion patterns, the contraction and clockwise rotation (patterns 12 and 14, respectively, in Fig. 2). (The tuning curves obtained from the simulation results will be shown in Fig. 6.)
6. MSTd cells have different position-dependent responses. Position-independent response selectivity is most prominent in single-component cells, while position-dependent response selectivity is most prominent in triple-component cells (Duffy and Wurtz 1991b). Many cells show a slope response profile, indicating that the response was stronger at some locations than at others, but a cell will never reverse its selectivity when the stimulus moves to a different position (Graziano 1994).

7. The selectivity of the multiple-component MSTd neurons is mostly position-dependent, and responses to different motion components change differently while the locations of the motion stimuli change. (Duffy and Wurtz 1991a, Fig. 8). These different position-dependent responses indicate that responses of different motion components of a multiple-component MSTd cell have different preferred regions of response in the receptive field, which usually do not coincide.

8. There are neurons that do not respond selectively to any of the motion components (Duffy and Wurtz 1991a; Graziano 1994);

To compare the performance of our model with the biological features listed above, we trained the network by repeatedly presenting to its input layer a set of $8 + 8 \times k^2$ different flow motion patterns containing eight planar motion patterns in eight different directions (first row of Fig. 2), and eight types of circular, radial, and spiral motion patterns (second row of Fig. 2) each with $k \times k = k^2$ different locations in the visual field. Here we chose $k = 7$ to cover a visual area of $7 \times 7$ patches, each represented by eight MT nodes. The total number of optic flow patterns is therefore $8 + 8 \times 7^2 = 400$. We also chose to have $n = 30$ MSTd nodes in each group of the output layer. Since the learning taking place in different groups is independent of each other (units in different groups do not compete with each other), there can be as many groups in the output layer as desired without affecting the results. Due to the random nature of the learning process (random initial values for the weights, etc.), the responses of these groups to the input patterns are statistically similar but not identical. This means that we can simulate a large number of MSTd cells with more statistically meaningful results. We may also choose to use slightly different learning parameters in different groups to model a variety of MSTd cells. Here we used 10 clusters in the output layer.

After the learning phase of about 3000 iterations, the network became stable and had learned all the input motion patterns. In the testing phase, the 400 input patterns were presented again to the network sequentially, each time with the winning node recorded. The results showed that each of the 400 motion patterns was responded to uniquely by one of the 30 nodes in each group of the output layer. However, as explained before, a node may respond to more than one input pattern. The responses of two of the groups in the output layer are shown in Figure 3. Here the 30 numbers (from 0 to 29) represent the 30 MST nodes in each group. The eight winners that responded most strongly to the eight planar motions of eight different directions are listed first, followed by eight arrays containing winning nodes that responded most strongly to circular, radial, and spiral motions of five by five different center locations in the central area of the visual field.

Moreover, to see the analog responses of the output nodes to motion stimuli with different center locations, we also computed the ana-

```
Output Group 1

Translation (0--7):  25  8 19  7  5 16 13  1

(8) Expansion       (9) Spiral-exp-cc  (10) Rotation-cc    (11) Spiral-con-cc
24 24 22 22 22      25 25 25 17 17      12  2  2  2 25      14 14  3  3  1
24 24 22 22 22      25 25 10 10 10      12  2  2  2 25      14 14  3  3  1
24 24 22 21 21      12 12 10 10 10       6  6  2 17 17      14 14  3  3  1
 8  8 21 21 21      12 12 10 10 10       6  6 27 17 17      20 20 13 13 13
 8  8 21 21 21      12 12  6  6  6       6  6 27 17 17      20 20 13 13 13

(12) Contraction    (13) Spiral-con-c  (14) Rotation-c     (15) Spiral-exp-c
20 20 20 14 14       0  0 23 23 23      29 29  0  0  0       4  4  5  5  5
20 20 20 11 11       0  0 23 23 23      29 29 18 18 18       4  4  5  5  5
15 15 11 11 11       0  0 26 26 23      29 29 18 18 18       4 28 28  5  5
15 16 16 11 11      29 26 26 26 26       4  4 18 18 18      24 28 28 28  8
16 16 16  1  1      29 29 26 26 26       4  4  4 18 18      24 28 28  8  8


Output Group 2

Translation (0--7):  11 17 13 19  8 29  5 15

(8) Expansion       (9) Spiral-exp-cc  (10) Rotation-cc    (11) Spiral-con-cc
 6 29 29 29 29      14 14 14 25 25       0  0  0 22 22      20  9 28 28 28
 6  6 29  1  1      14 14 14 25 25       0  0  0 22 22      20  9 28 28 15
 6 17  1  1  1      14 14 14  8 25       0  0 21 21 22      20  9  9 28 15
17 17  1  1  1      13 13  8  8  8      20 20 21 21 25      23 23 23 22 22
17 17 17 13 13      13 13  8  8  8      20 20 21 25 25      23 23 23 22 22

(12) Contraction    (13) Spiral-con-c  (14) Rotation-c     (15) Spiral-exp-c
10 10  2  2  2      18 18 10 10 10      16 26 18 18 18      24 24 26 26 26
10 10  2  2  2      18 18 12 12 12      16 16 18 18 18      24 24 26 26 26
10 10  2  2  2      18 18 12 12 12      16 16  7  7  7      24  4  4 26 26
27 27 27 15 15      16 16 12 12 11      24 24  7  7  7       6  4  4  4 17
27 27 27 15 15      16 16 16 11 11      24 24  7  7  7       6  6  4 17 17
```

Figure 3: Responses of the output nodes. The 16 different types of input patterns [represented by (0) to (15), in the same order as in Fig. 2] are responded by the 30 output nodes (represented by 0 to 29) in each group. The first eight are translational motions of different directions, and the second eight are circular, radial and spiral motions of different directions and center locations in the visual field. The winner of each input pattern is listed.

log output (instead of the binary output generated by the winner-takes-all computation in competitive learning) of the output nodes. In the above example, five output nodes (14, 3, 1, 20, and 13) in group 1 all responded favorably to counterclockwise, contractive spiral motions with different center locations in the visual field. For each node, we obtained its responses to the spiral motion patterns of different center locations over the receptive field. From the results shown in Figure 4, we see that a node has the strongest response and becomes the winner when the center of the stimulus is located within a small area favored by that node, while it still responds (with smaller output and therefore no longer the winner) to the stimuli whose centers are located elsewhere.

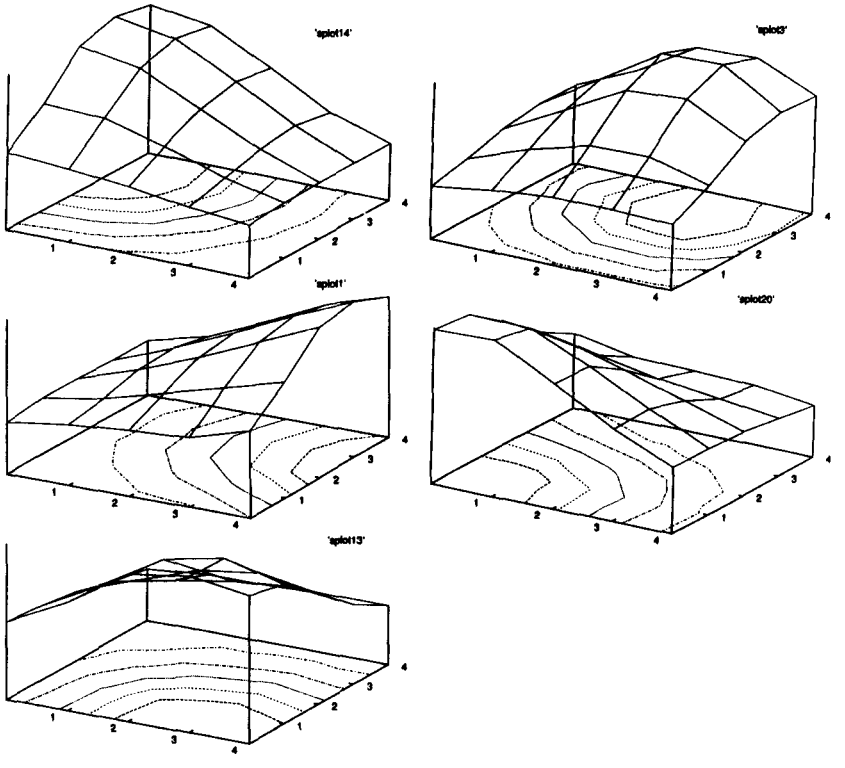


Figure 4: Analog responses of some output nodes to spiral motions. The analog responses of five output nodes to the counterclockwise contractive motions of different center locations are plotted. The vertical axis is the intensity of the responses, and the horizontal plane represents the visual area in the receptive field. The contour lines of the responses are also shown on this plane.

From the responses of the MSTd nodes in the above two groups, the following features of the network can be observed:

1. The nodes in the output layer of the network (simulating MSTd neurons) have much larger receptive fields than the nodes in the input layer (simulating MT neurons) because each of the output nodes is connected to all input nodes and therefore receives complete information present in the visual area.

2. The output nodes of the network respond to different types of motion stimuli, including planar translational motion, circular motion of clockwise and counterclockwise rotation, radial motion of expansion and contraction, and various spiral motions.

3. There exist three types of output nodes: single-component nodes (nodes 7, 19, 10, 18, 21, 22, etc. in group 1, nodes 5, 19, 2, 7, 14, etc. in group 2), double-component nodes (nodes 5, 13, 16, 20, 24, etc. in group 1, nodes 8, 11, 15, 20, 29, etc. in group 2), and triple-component nodes (nodes 1, 8, 25 in group 1, nodes 13, 15, 17 in group 2). It may appear that there are too few triple-component nodes compared to the biological findings (29% triple-component cells according to Duffy and Wurtz 1991a). However, note that here we count a node as a triple-component node only when it is the winner for all three motion patterns, while actually other nodes may also respond to three motion patterns (although not necessarily always the winner), as indicated by their responses shown in Figure 4 (also see discussion below).

4. There is no underlying mechanism to separate the three types of nodes and therefore they form a continuum of response selectivity. This is best illustrated by the triangular diagram in Figure 5, which closely resembles the result found biologically by Graziano (1994, Fig. 6). The diagram shows the responses of each MSTd cell, represented by a dot in the diagram, to three types of motion stimuli (translational, circular and radial) represented by the three vertices of the diagram. The closer a dot is to a vertex, the stronger the cell responds to the corresponding motion than to the others. In other words, the nodes located close to the vertices are single-component nodes, those located close to the edges are double-component nodes, and those located in the central area are triple-component nodes. As these dots have random but relatively even distribution, they do form a continuum in the space.

5. The responses to various motion stimuli are obtained for some output nodes in group 1, as listed in Table 1. The numbers in the first column represent eight output nodes, and the numbers in the first row represent eight different motion patterns in the same order as in the second row of Figure 2. Since a node's response to a motion pattern varies, depending on where the center of the motion is in the field, listed in Table 1 are the maximum responses for each type of motion pattern. From these data, the tuning curves of four of the nodes (20, 0, 24, 17) are plotted, as shown in Figure 6. For convenience, the data points are properly shifted horizontally (based on the fact that the eight motion patterns are periodic, i.e., pattern 15 is also a neighbor of pattern 8) so that the peak is always in the middle of the curve, although actually these nodes favor different types

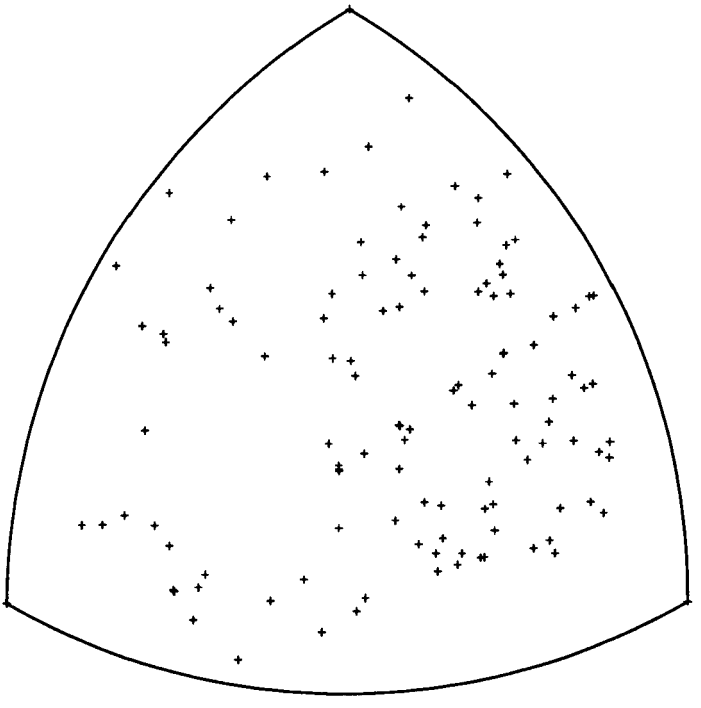

Figure 5: Relative response strength to different types of motions. The distance $D_X$ from a dot to a vertex $X$ is computed as $1/(1 + kR_X)$, where $R_X$ is the maximum response to the motion stimuli represented by $X$, and $k$ can be found numerically for each node so that the three distances so computed indeed define a unique point in the diagram.

Table 1: Maximum Responses to Various Motion Stimuli.

| Nodes | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) |
|---|---|---|---|---|---|---|---|---|
| 8 | 7.4 | 4.2 | 1.7 | 1.0 | 1.1 | 1.5 | 3.5 | 7.2 |
| 24 | 7.9 | 4.6 | 2.0 | 1.0 | 1.0 | 1.5 | 3.5 | 6.8 |
| 6 | 3.2 | 7.4 | 7.1 | 4.0 | 1.3 | 0.3 | 0.4 | 1.1 |
| 17 | 3.3 | 6.2 | 7.9 | 5.0 | 2.0 | 1.1 | 0.9 | 1.3 |
| 14 | 0.9 | 2.2 | 4.9 | 7.9 | 6.4 | 3.0 | 1.3 | 0.8 |
| 20 | 0.7 | 1.2 | 3.5 | 6.6 | 7.2 | 4.2 | 1.5 | 0.6 |
| 0 | 0.7 | 0.1 | 0.1 | 1.3 | 3.9 | 6.9 | 5.9 | 2.5 |
| 4 | 3.4 | 0.9 | 0.2 | 0.3 | 1.2 | 3.8 | 6.9 | 6.7 |

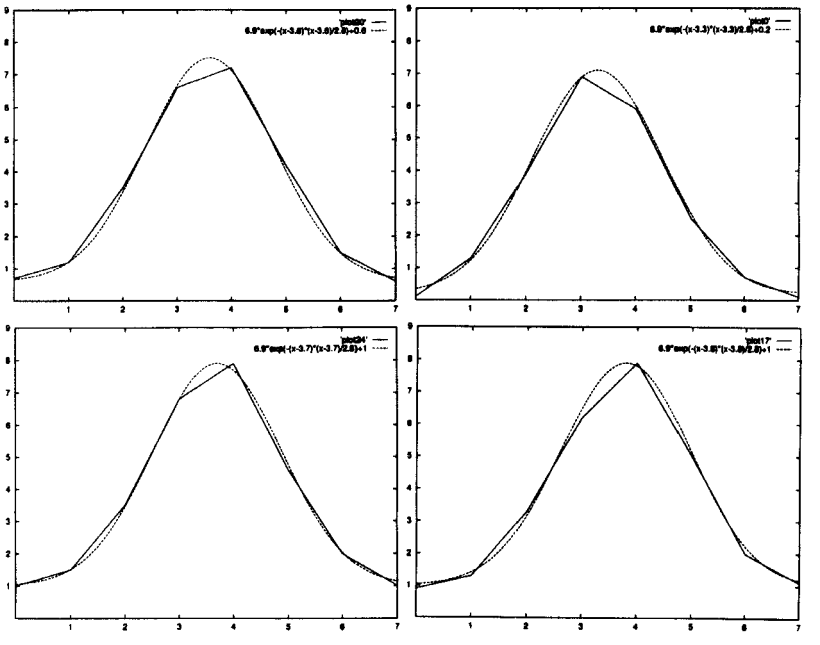

Figure 6: The tuning curves of some output nodes. The eight response intensities (the piecewise-linear line) of each node can be closely fitted with a gaussian function (the smooth curve). Each gaussian curve is slightly shifted in both vertical and horizontal directions (as indicated by its expression) to best fit the tuning curve of the node.

of motion patterns as shown in the table. It can be seen that each of these curves can be very closely fitted with a gaussian function, just like the result found biologically by Graziano (1994, Fig. 7B).

6. The single-component nodes tend to have larger responsive areas and therefore are more position independent (e.g., nodes 10, 18, 21, 22 in group 1, and nodes 2, 7, 14 in group 2 in Fig. 3), while the multicomponent nodes (double and triple-component) tend to have smaller responsive areas and are more position dependent (e.g., nodes 1, 8, 25 in group 1, and nodes 13, 15, 17 in group 2). This feature is consistently observed in more examples. Moreover, Figure 4 shows that the responses do have sloping profile as found biologically in (1994).

7. Multicomponent nodes have different responsive areas for different motion patterns (i.e., the responsive areas for different input patterns do not coincide). For example, as shown in Figure 3, the triple-component node 1 in group 1 responds strongly to counterclockwise, contractive spiral motions located close to the upper-right corner and a contraction located at the lower-right corner; and the triple-component node 13 in group 2 responds strongly to expansion motions located at the lower-right corner and counterclockwise, expansive spiral motion located at the lower-left corner of the field. This feature is consistently observed in more examples.

8. There exist dead nodes (node 9 in group 1 and node 3 in group 2) that never win and therefore never respond to any input motion stimuli.

All of these features match well with the properties of MSTd neurons found neurophysiologically, as listed at the beginning of this section.

## 4 Discussion

The main advantage of this model is that a set of important responsive properties of the MSTd neurons can all be accounted for by a network with simple competitive learning mechanism. As found in the recent study by Graziano (1994), there exist MSTd cells that respond to spiral motions of different directions much more strongly than to other expansion, rotation, or translational motion. This finding indicates that the three-channel decomposition hypothesis for visual perception of motion (optical flow is decomposed into three separate and discrete channels of translational, radial, and rotational motion components) does not appear to be correct. Instead, there is a continuum of patterns to which MSTd cells are selective. These responsive properties are well matched by the behaviors of the model presented here. As assumed in this model, all global motion patterns of different center locations are composed of a set of $k$ by $k$ small patches each represented by an MT node whose preferred direction best fits the local motion direction in the patch. Since different motion patterns may share some MT nodes, there do not exist clear boundaries in the feature space to separate different motion patterns, i.e., the motion patterns form a continuum rather than separate clusters. Consequently, the output nodes responding to these motion patterns also form a continuum composed of single-, double-, and triple-component nodes.

Further work in the future includes the following two aspects. First, the model has been tested only on simplified and idealized inputs. We will further test the model to assess its ability to respond to optic flow extracted from real images. The main difficulty that we have to overcome is the huge amount of data to process when real images are used. As

discussed above, the number of motion patterns (of different types and different center locations) is proportional to the size of the image squared. And the more motion patterns the network needs to recognize, the longer time (more iterations) it requires for training. The high computation demands required for simulating the responsive properties of the MSTd neurons using real images may be overwhelming.

Second, we can expand the model so that more responsive properties of the MSTd cells can be accounted for. According to Duffy and Wurtz's (1993) recent finding, there exist MSTd cells that are responsive to center of motion (COM) of both circular and radial motions (including focus of expansion, FOE). We want to show that this responsive property can be achieved from a hierarchical structure obtained by adding to the current network one or more layers. The nodes in the added layers should develop sharper tuning for the center positions of motions and therefore respond to the COMs. This is based on the feature of competitive learning that a learned node will represent the common features of the patterns represented by a set of nodes in the previous layer. In our case, a node may become responsive to a motion pattern whose COM is located in the intersection of the patterns represented by a set of nodes in the previous layer. In other words, the responses of the nodes in an added layer become more position-sensitive than those in the current output layer.

Finally we would like to compare the architecture of this model to that of the multilayer Hebbian learning network proposed by Linsker (1986a,b). Both models are based on a simple hierarchical architecture with increasing receptive field and use unsupervised learning algorithm. Linsker's model used white noise as the input data and successfully simulated enter-surround cells and orientation-selective cells found in V1 area, while the model presented here uses the data simulating the velocity selectivity found in the MT neurons as the input, and simulates the selective responses to various optic-flow motion components found in the MSTd area. These models show that important neuronal properties found in the biological visual system can be successfully accounted for by some simple network models.



## Acknowledgments

I thank Dr. Christof Koch for discussions in the course of this research and the two anonymous reviewers for their very helpful comments. The research was supported by the faculty research fund provided by the Harvey Mudd College.

## References


Albright, T. D. 1984. Direction and orientation selectivity of neurons in visual area MT of the Macaque. *J. Neurophysiol.* **6**, 1106–1130.

Albright, T. D. 1989. Centrifugal directional bias in the middle temporal area (MT) of the macaque. *Visual Neurosci.* **2**, 177–188.

Brenner, E., and Rauschecker, J. P. 1990. Centrifugal motion bias in the cat's lateral suprasylvian visual cortex is independent of early flow field exposure. *J. Physiol.* **423**, 641–660.

DeSieno, D. 1988. Adding a conscience to competitive learning. *IEEE Int. Conf. Neural Networks* (San Diego 1988) **I**, 117–124.

Duffy, C. J., and Wurtz, R. H. 1991a. Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *J. Neurophysiol.* **65**, 1329–1345.

Duffy, C. J., and Wurtz, R. H. 1991b. Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *J. Neurophysiol.* **65**, 1346–1359.

Duffy, C. J., and Wurtz, R. H. 1993. MSTd Neuronal responses to the center-of-motion in optic flow fields. *Soc. Neurosci. Abstr.* **19**, 5311.9.

Graziano, M. S. A., Andersen, R. A., and Snowden, R. 1994. Tuning of MST neurons to spiral motions. *J. Neurosci.* **14**(1): 54–67.

Graziano, M. S. A., Andersen, R. A., and Snowden, R. 1990. Stimulus selectivity of neurons in macaque MST. *Soc. Neurosci. Abstr.* **16**, 7.

Grossberg, S. 1976. Adaptive pattern classification and universal receding: II. Feedback, expectation, olfaction, illusions. *Biol. Cybernet.* **23**, 187–202.

Grossberg, S. 1987. Competitive learning: from interactive activation to adaptive resonance. *Cog. Sci.* **11**, 23–63.

Kohonen, T. 1988. *Self-Organization and Associative Memory*, 2nd ed. Springer-Verlag, Berlin.

Lappe, M., and Rauschecker, J. P. 1993. A neural network for the processing of optic flow from ego-motion in man and higher mammals. *Neural Comp.* **5**, 374–391.

Linsker, R. 1986a. From basic network principles to neural architecture: emergence of spatial-opponent cells. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7508–7512.

Linsker, R. 1986b. From basic network principles to neural architecture: emergence of orientation-selective cells. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8390–8394.

Maunsell, J. J. R., and Van Essen, D. C. 1983a. Functional properties of neurons in middle temporal visual area of the Macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J. Neurophysiol.* **49**, 1127–1147.

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

Orban, G. A., Lagae, L., Verri, A. *et al.* 1992. First-order analysis of optical flow in monkey brain. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2595–2599.

Perone, J. A. 1992. Model for the computation of self-motion in biological systems. *J. Opt. Soc. Am. A* **9**(2).

Rodman, H. R., and Albright, T. D. 1987. Coding of visual stimulus velocity in area MT of the Macaque. *Vision Res.* **27**, 2035–2048.

Rumelhart, D. E., and Zipser, D. 1985. Feature discovery by competitive learning. *Cog. Sci.* **9**, 95–112 (also, ch. 5, *Parallel Distributed Processing*, Vol. 1, The MIT Press.)

Saito, H., Yukie, M., Tanaka, K., *et al.* 1986. Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *J. Neurosci.* **6**, 145–157.

Snowden, R. J. *et al.* 1991. The response of area MT and V1 neurons to transparent motion. *J. Neurosci.* **11**, 2768–2785.

Tanaka, K., and Saito, H. 1989a. Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *J. Neurophysiol.* **62**, 626–641.

Tanaka, K., and Saito, H. 1989b. Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the Macaque monkey. *J. Neurophysiol.* **62**, 642–656.

Ungerleider, L. G., and Desimone, R. 1986. Cortical connections of visual area MT in the Macaque. *J. Comp. Neurol.* **248**, 190–222.

Zhang, K., Sereno, M., and Sereno, M. 1993. Emergence of position-independent detectors of sense of rotation and dilation with Hebbian learning: An analysis. *Neural Comp.* **5**, 597–612.

Received September 23, 1993; accepted June 22, 1994.