Representation of Visual Motion in the Extrastriate Visual Cortex

Keiji Tanaka

In this chapter I describe how field motion caused by self-movements of the observer and object movements in the external space are extracted and represented in the neuronal circuitry in the extrastriate visual cortical areas. The finding that the extrastriate visual cortex is composed of multiple (>20) areas was one of the most important triggers of recent explosive development in neuroscientific research on vision. In studies of anatomical connections between these areas, several chains of areas starting at the primary visual cortex and leading to the temporal and parietal association cortical areas were identified. These temporal and parietal association areas had long been associated with unique cognitive functions, based on clinical case studies of human patients and studies on animal models, and therefore the elucidation of their afferent pathways has led to opportunities to study the neural mechanisms of the cognitive functions using bottom-up approaches.

Monkeys with lesions on area 7A in the inferior parietal lobule show deficits in visuo-motor control in space and spatial visual recognition but no deficits in the ability to recognize visual shapes (Mishkin, Ungerleider, and Kathleen, 1983; Goodale and Milner, 1992). Visual information related to the spatial functions is likely to be processed along the pathways from the primary visual cortex to area 7A. Among the several pathways from the primary visual cortex to area 7A, the one extending through the medial temporal area (MT) and the middle superior temporal area (MST) is characterized by an abundance of cells selective for direction of motion. Layer 4B of the primary visual cortex projects to MT directly (Zeki, 1971; Van Essen, Maunsell, and Bixby, 1981; Ungerleider and Mishkin, 1979) and indirectly via the CO thick stripes in V2 (DeYoe and Van Essen, 1985; Shipp and Zeki, 1985; Livingstone and Hubel, 1987). MT projects to MST (Maunsell and Van Essen, 1983a) which in turn projects to area 7A (Mesulam et al., 1977; Andersen et al., 1990). The proportion of cells which respond to visual motion with selectivity for the direction of movements is large in all of these layers, stripes, and areas. Therefore, it is expected that processing of visual motion information progresses along this pathway. Information of visual motion is useful for the visuo-motor control in space and the recognition of space.
Figure 10.1 Responses of a typical MT cell. They are selective for the direction of motion but not the orientation of the slit or the color of the stimulus.

**REPRESENTATION OF LOCAL MOTION**

Figure 10.1 shows response properties of a typical cell in MT. The area bounded by a broken line represents the receptive field. When a light spot moves across the receptive field in the preferred direction of the cell, the cell responds with high-frequency spike activity, whereas when the same spot moves across the receptive field in a nonpreferred direction, the cell does not respond. On average, a deviation from the optimal direction by 30 degrees halves the magnitude of the response (Maunsell and Van Essen, 1983b). This selectivity of response is called *selectivity for direction of motion or direction selectivity*. Responses of cells in MT are also selective for the speed of motion, although the tuning of responses for speed is generally not very sharp. On the average, the magnitude of the response is halved by changing the speed by 1.5 octaves from the optimal speed (Maunsell and Van Essen, 1983b; Lagaee, Raiguel, and Orban, 1993). While the responses are selective for the direction and speed of motion, they are generally insensitive to the shape and color of moving stimuli (Zeki, 1974) except that many of them preferentially respond to stimuli smaller than the size of the receptive field (Cheng et al., 1994). Cells in MT respond to stimuli composed of contours of luminosity contrasts as well as those composed of contours of color or texture contrasts without luminosity difference (Saito et al., 1989; Albright, 1992; Dobkins and Albright, 1994).

However, the direction selectivity does not appear first in MT along the pathway. Cells in layer 4B of V1 (Dow, 1974; Livingstone and Hubel, 1984) and those in the thick CO stripes of V2 (DeYoe and Van Essen, 1985; Hubel and Livingstone, 1987), which project to MT, show directionally selective responses to moving stimuli. It is not known whether the direction selectivity of cells in MT reflects only that of their inputs or whether there are additional mechanisms in MT that contribute to the selectivity (see Mikami, Newsome, and Wurtz, 1986; Rodman et al., 1989). In any case, the direction selectivity is not unique to cells in MT.

A unique property of MT is that the preferred direction of individual cells is regularly arranged in localized regions. That is, cells preferring the same direction are clustered in a region elongated vertical to the cortical surface (column), and the preferred direction gradually changes in a clockwise or counterclockwise direction along the cortical surface. Based on data obtained in penetrations oblique to the cortical surface, the model shown in figure 10.2 has been proposed (Zeki, 1974; Albright, Desimone, and Gross, 1984). A region with dimensions of about 0.5 mm x 0.5 mm across the cortical surface contains columns tuned to directions covering 360 degrees. The preferred direction gradually changes along one axis, while opposing directions are paired along the orthogonal axis. The receptive fields of cells in this region containing a complete set of direction columns overlap with one another. Therefore, this cortical region contains a complete set of cells repre-
senting all directions of moving stimuli within a region of the visual field. This cortical region is a unit of the visuo-topical organization across the entire MT.

This simple model of MT composed of homogenous units has recently been modified. As will be explained, MT contains cells with receptive fields having inhibitory surround and those with receptive fields lacking inhibitory surround (Tanaka et al., 1986). These two kinds of cells are segregated in patches (Born and Tootell, 1992). How these patches are related to direction columns is not yet known.

FIELD MOTION VERSUS OBJECT MOTION

Movements of images on the retina can be caused by two kinds of events. One is the movement of objects in the external space, and the other is the movement of the eye, head, and/or body of the observer (self-motion). Movements of images are usually limited to a small part of the visual field when an object moves, while the whole visual field moves when the observer moves. The areal extent of coherent movements gives a useful cue for discrimination of these two sources of movements on the retina. Because object movement and observer movement often occur simultaneously, the movement of the image of an object relative to the movement of its background, but not the movement of the object image on the retina itself, should be extracted to perceive the movement of the object in the external space. I will describe how these two kinds of movements are extracted in MT and MST, to which MT projects.

REPRESENTATION OF FIELD MOTION

Figure 10.3 shows typical receptive fields of cells in MT and MST. The receptive fields of cells in MT are small in the central visual field. For example, the average diameter of the receptive field is 3 degrees at an eccentricity of 5 degrees from the fovea and 5 degrees at an eccentricity of 10 degrees from the fovea. The receptive fields become larger as the eccentricity of their centers increases, but their borders facing the fovea recede from the fovea. The receptive fields of cells in MST are larger than those of cells in MT, and many of them include the fovea. The large size of the receptive fields of MST cells raises the possibility that they extract the movements of wide-field movements.

The majority of cells in MST show directionally selective responses to moving stimuli as do the majority of cells in MT. In addition, most of the cells in the dorsal part of MST prefer coherent movements over a wide visual field to movements limited to a small visual field (Komatsu and Wurtz, 1988; Tanaka and Saito, 1989). Figure 10.4 shows responses of such a cell. It responded strongly to a translation of a dot pattern or that of a stripe pattern over an 80 × 80-degree field, whereas its responses to the movement
of a single slit were much weaker. Perception of coherent movements extending over a wide visual field is critical for the observer to be able to control the self-motion which caused the wide-field movements on the retina. The readers may confirm this by attempting to perform complex movements involving multiple body parts with the eyes closed. The selectivity of cells in the dorsal MST to wide-field movements is advantageous for extracting the self-movements from the retinal movements.

Wide-field movements caused by the movements of the observer are not limited to parallel translation in which image components move in the same direction with the same speed. When the observer moves forward with the visual axis of the eye fixed relative to the direction of motion, image components move in radial directions; for example, an expansion of the image occurs on the retina (see Hildreth and Royden, chapter 9; Warren, chapter 11). When the observer recedes, the image contracts. When the observer turns around an axis with the visual axis of the eye fixed relative to the axis of rotation, a clockwise or counterclockwise rotation occurs on the retina.

Corresponding to these ecological considerations, the dorsal MST contains cells which respond selectively to expansion, contraction, clockwise rotation, or counterclockwise rotation (Sakata et al., 1985; Sakata et al., 1986; Sakata et al., 1994; Saito et al., 1986; Duffy and Wurtz, 1991; Orban et al., 1992; Graziano, Andersen, and Snowden, 1994). Figure 10.5 shows responses of two such cells. Translation of a dot pattern in any direction in the frontoparallel plane did not evoke excitation, whereas expansion of the same dot pattern (figure 10.5, top) or a clockwise rotation (figure 10.5, bottom) evoked strong excitation. About one-third of all cells in the dorsal MST were exclusively activated by expansion, contraction, clockwise rotation, or counterclockwise rotation, whereas the remaining two-thirds of the cells could be maximally activated from topological translation in some direction. Responses of cells in the dorsal MST are selective not only for the mode and direction of field motion, but also for the speed of the motion. Cells are tuned to different speeds, and as a whole they cover the entire physiological range of speeds (Tanaka and Saito, 1989; Orban et al., 1995). Thus, the mode, direction, and speed of field motion are directly represented in the dorsal MST by the selective activation of cells.

Depth structure in the environment modifies the motion patterns on the retina caused by self-movements. Horizontal shifts of the body, if the eye is fixed on a point in the distance, will result in movement of environmental components in the same direction but at a speed inversely proportional to the distance of the environmental components from the observer. If the eye is fixed on a point at the middle of a depth gradient, such as a flat ground, a combination of opposing movements with symmetrical gradients (shearing) will be evoked. The speed of movement of the environmental components, caused by advance or retreat of the observer, is also inversely proportional to the depth of the environmental components. Of the cells which responded most strongly to an expansion or contraction among expansions, contractions, rotations, and frontoparallel translations, 4 percent responded more strongly to an axial expansion/contraction than to a symmetrical expansion/contraction, and of the cells which responded most strongly to a rotation among the three kinds of movements, 9 percent responded more strongly to a shearing than to a symmetrical rotation (Tanaka and Saito, 1989). These values are likely underestimates, because the cells were first tested using the symmetrical expansions/contractions, rotations, and frontoparallel translations in these experiments. It is thus possible that not only the mode, direction, and speed of the self-movement, but also the coarse depth structure of the environment, are represented in the activity of cells in the dorsal MST.

Movements of different modes can occur together. Frontoparallel translation is combined with expansion/contraction or frontoparallel rotation when the angular relation between the visual axis of the eye and the direction of advance or retreat or the axis of turns of the observer changes during the motion. A combination of expansion/contraction and frontoparallel rotation—for example, spiral motion—occurs on the retina when the observer turns its head or body while advancing or retreating. Advancing or retreating movement in a three-dimensional environment combined with eye movement results in retinal motion patterns containing spiral components (see Warren, chapter 11).

There are two possible ways to represent these combinatorial movements. One is to extract the magnitude of components in each mode and represent the component magnitude as activity of respective groups of cells. This way...
of representation is economical in terms of the number of required cell groups, but it has a difficulty in terms of the mechanisms of the selectivity. If the selectivity of the MST cells depends on the spatial arrangement of the preferred direction of excitatory inputs, as I will propose, the magnitude of responses will be determined by the degree of matching between the local vector of actual retinal image motion and the preferred vector of the input cells, and not by the magnitude of the mode component. The other way to represent the combinatorial movements is to have cells with responses tuned to particular combinations. Orban et al. (1992) and Graziano, Andersen, and Snowden (1994) showed that there are MST cells which respond more strongly to combinatorial movements than to pure expansion/contraction, rotation, or translation.

MECHANISMS OF THE SELECTIVITY FOR THE MODE OF FIELD MOTION

Expansion/contraction and clockwise/counterclockwise rotation of a wide textured field can be discriminated from each other and from frontoparallel translation according to three stimulus factors (figure 10.6). One is the spatial arrangements of direction of local motion. The directions of movements of the texture components are arranged radially in expansion and contraction, along concentric circles in frontoparallel rotation, and in parallel in fronto-parallel translation. The second factor is the gradient of speeds of local movements from the center to the periphery of the entire movement pattern. The speed of movements of texture components is proportional to the distance from the center of movement. The direction of the speed gradient and the direction of local movements are the same in expansion and contraction, whereas they are orthogonal to each other in rotation. The third factor is the size change of individual texture components in expansion/contraction, and the constant inward acceleration of movements of texture components in rotation.

Tanaka, Fukada, and Saito (1989) examined the relative contributions of the three factors to responses of the MST cells by comparing their responses to artificial motion patterns that contained only one or two factors, with those of the natural expansion, contraction, or rotation. The contribution of the spatial arrangement of directions was predominant in rotation cells (figure 10.7) and in expansion/contraction cells. The speed gradient seemed to contribute to some expansion/contraction cells, but even in these cells the contribution of the spatial arrangement of directions was larger. Graziano, Andersen, and Snowden (1994) and Lagae et al. (1994) obtained essentially the same results.

Based on these results, a rather simple model has been proposed for the mechanisms of the selectivity of MST cells. It is assumed that cells in MST receive their main excitatory inputs from cells in MT. Because the receptive fields of MST cells are much larger than those of MT cells, inputs from multiple MT cells with receptive fields at different visual field positions likely converge onto a single MST cell. If the preferred directions of the input cells are the same, the postsynaptic cell will preferentially respond to fronto-parallel translation in that direction (figure 10.8, left); if the input cells are radially arranged, it will respond selectively to expansion (figure 10.8, middle); if the input cells are arranged tangentially in concentric circles, it will respond selectively to rotation (figure 10.8, right); and if the input cells are arranged spirally, it will respond selectively to spiral motion. The selectivity for wide-field movements can be achieved by setting the threshold for spiking activity.
at a relatively high level so that the postsynaptic cell fires only when most of the input cells are activated simultaneously.

The expansion/contraction and rotation cells in the dorsal MST show an extensive invariance for changes in the position of the center of movements (Saito et al., 1986; Lagae et al., 1994; Graziano, Andersen, and Snowden, 1994). The models shown in figure 10.8 require the stimulus center to be located at the center of the input arrangement to evoke the maximum responses. To achieve the position invariance, multiple sets of inputs with the center of arrangement at different positions should converge onto a single postsynaptic cell (figure 10.9). Summation of inputs in a set should be isolated from that in the other sets, because otherwise the selectivity for the mode of motion will be lost. As a possibility, it is assumed in figure 10.9 that different sets of inputs converge on different branches of dendrites.

Figure 10.9 A possible mechanism behind the position invariance of MST cells. (Reprinted with permission from Saito et al., 1986.)

Figure 10.10 Surround inhibition selective for the direction of motion, found in half of MT cells. (Reprinted with permission from Tanaka et al., 1986.)

REPRESENTATION OF RELATIVE MOVEMENT OF OBJECTS ON THEIR BACKGROUND

When an object moves in the external space while the observer moves the eyes, head, and/or body, the object movement in the external space can be perceived only by extracting the relative movement of the object image on the background. The extraction of the relative movement is partially seen in responses of cells in MT (Allman, Miez, and McGuinness, 1985; Tanaka et al., 1986).

A slit was presented at the center as an object and a wide dot pattern around it as a background, as shown in figure 10.10, in the experiments of Tanaka et al. (1986). The excitatory receptive field was masked from the dot pattern stimulation for the simplicity of data analysis. When the background moved together with the slit in the same direction and at the same speed, the response was considerably reduced in half of the cells in MT (figure 10.10, middle). This means that the stimulation of the surrounding region stimulated by the dot pattern exerted inhibition on the cell. The excitatory receptive field is surrounded by an inhibitory region. Because the response to the slit was not suppressed when the slit and background moved in opposite
directions (figure 10.10, bottom), the surround inhibition is direction selective. Typically, the inhibition was halved when the direction of the background movement deviated by 60 degrees from that of the slit movement, and the inhibition disappeared when the deviation was 90 degrees (figure 10.11, top).

The surround inhibition was also selective for the speed of background movement. The speed of the background movement was changed while the speed of the slit movement was fixed (see figure 10.10 of Tanaka et al., 1986). In some cells, the amount of inhibition decreased when the speed of the background movement deviated from the speed of the slit movement in either faster or slower directions (figure 10.11 lower). In other cells, the amount of inhibition decreased upon decrease of the background speed but remained constant when the background speed was increased. There were also cells in which the amount of inhibition decreased only when the background speed was increased. The proportions of the three kinds of cells were roughly equal. These variations in speed tuning of surround inhibition may be useful for the system when the object and background are located at different distances from the observer. It is interesting to examine the surround effect under conditions of various combinations of disparity between the object and background (see Roy and Wurtz, 1990).

The inhibitory surround extends more than 20 degrees from the center of the excitatory receptive field, because movement of a dot pattern presented only outside a 40 x 40-degree area centered at the excitatory receptive field evoked weaker but definite inhibition. The shape and size of the inhibitory surround were quantitatively examined in Raiguel et al. (1995) and Xiao et al. (1995).

The directionally selective surround inhibition of MT cells can be explained using two simple models. One is convergence of inhibitory inputs from multiple MT cells having receptive fields in surrounding regions (figure 10.12). The other is feedback inhibition from an MST cell which has a large receptive field centered at the receptive field of the MT cell. However, the latter model faces a difficulty in that cortico-cortical connections are generally excitatory. The inhibition must be relayed by an inhibitory cell in MT, and thus there must exist cells with receptive fields larger than 40 degrees in diameter in MT. Such cells are very rare in MT.

The directionally selective surround inhibition observed in MT cells is not perfect for detecting the relative movements of an object on a background. There are cases in which responses of MT cells are dissociated from the relative movement of objects on the background. When the image of an object remains stationary on the retina while the background moves (figure 10.13, top right), the ideal cells that extract relative movements of objects on the background respond in the direction opposite to the direction in which they respond to movement of the object on a stationary background (figure 10.3, top left). MT cells do not respond in this stimulus configuration (figure 10.13, middle).

Cells that respond to background movement behind a stationary object and movement in the opposite direction of an object on a stationary background were found in the ventral part of MST (figure 10.13, bottom, Sugita and Tanaka, 1991; Tanaka et al., 1993). Cells in the ventral MST have receptive fields as large as those of cells in the dorsal MST. Most of them do not respond to movements of a wide textured field but respond to movements of a small stimulus. However, when a stationary object was placed in front of a moving wide textured field, one-third of the cells responded to movement in the direction opposite to the direction in which they responded to
movements of a small object. The stationary object must be small but can be placed anywhere in the large receptive field. Thus, compared with the responses of the MT cells, those of the ventral MST cells carry less information about the position of the object but more precise information about the direction of the relative movement of the object on the background. Responses of individual ventral MST cells are insufficient to carry the information of the speed of the relative movement. The speed tuning of the responses to the background movement behind an object did not necessarily match that of the responses to the object movement on the background.

Such responses of ventral MST cells cannot be explained by integration of pure motion information, but require integration of a kind of occlusion information with motion information (Sugita and Tanaka, 1991). The simplest model we should first consider is the double opponent structure between center and surround (Frost and Nakayama, 1983). The center induces excitation in one direction and inhibition in the opposite direction, whereas the surround induces inhibition in the first direction and excitation in the opposite direction. This model does not explain two aspects of responses of the ventral MST cells. First, based on the model, it is expected that responses are evoked only with an object located at the center, but actually the response magnitudes are largely invariant to changes of the position of the object. Second, the responses disappeared when the border of the object was blurred (figure 10.14). This elimination of responses was not caused by a change in the exact position of the border, because the responses were evoked with objects of various sizes. The disappearance and appearance of

Figure 10.13 Unlike cells in MT, those in the ventral MST respond to the background movement behind a stationary object. (Reprinted with permission from Tanaka et al., 1993.)

Figure 10.14 The responses of ventral MST cells to the background movement disappear when the border of the foreground object is blurred (bottom right). The same blurring does not influence the response to the movement of the object (bottom left). (Reprinted with permission from Sugita and Tanaka, 1991.)

Figure 10.15 Occlusion cues used to extract the relative movements between objects and their background. (Reprinted with permission from Sugita and Tanaka, 1991.)

texture components of the background at the border of the foreground object, which are kinds of cues of occlusion, must contribute to the responses (figure 10.15). The occlusion cues alone do not evoke the responses, because the responses were eliminated by decreasing the outer diameter of the background stimulus (see figure 10.4 of Sugita and Tanaka, 1991). Integration of the occlusion cues with the background motion over a wide field is necessary for the responses to be evoked.
CONCLUSION

MT represents a single visuo-topical map as a whole, and patches with cells subject to surround inhibition and ones with cells not subject to surround inhibition are intermingled in regions of MT. MST has a crude visuo-topical organization with two maps, one in the dorsal part and the other in the ventral part. Cells near the border between the dorsal and ventral parts have receptive fields covering only the peripheral visual field, and those near the dorsal and ventral ends have receptive fields including the central visual field. Cells in the dorsal MST preferentially respond to movements of a wide textured field, whereas those in the ventral MST preferentially respond to movements of a small object. Therefore, signals of field movements and object movements, which are intermingled in MT, are divided into the dorsal and ventral parts in MST (figure 10.16). In addition to the separation, signals of each kind of movement are further processed to represent behaviorally significant information more explicitly by activity of single cells. Different modes of field movements are represented by activity of different groups of cells in the dorsal MST, and the relative movements of an object on a background are represented in the ventral MST more faithfully and over a wide range of conditions than in MT.

REFERENCES


