Neuron, Vol. 37, 525–535, February 6, 2003, Copyright ©2003 by Cell Press

Two-Dimensional Substructure of Stereo and Motion Interactions in Macaque Visual Cortex

Christopher C. Pack,* Richard T. Born, and Margaret S. Livingstone
Harvard Medical School
Department of Neurobiology
220 Longwood Avenue
Boston, Massachusetts 02115

Summary

The analysis of object motion and stereoscopic depth are important tasks that are begun at early stages of the primate visual system. Using sparse white noise, we mapped the receptive field substructure of motion and disparity interactions in neurons in V1 and MT of alert monkeys. Interactions in both regions revealed subunits similar in structure to V1 simple cells. For both motion and stereo, the scale and shape of the receptive field substructure could be predicted from conventional tuning for bars or dot-field stimuli, indicating that the small-scale interactions were repeated across the receptive fields. We also found neurons in V1 and in MT that were tuned to combinations of spatial and temporal binocular disparities, suggesting a possible neural substrate for the perceptual Pulfrich phenomenon. Our observations constrain computational and developmental models of motion-stereo integration.

Introduction

The faithful encoding of rapid, yet locally subtle, changes in the visual scene are critical for an animal's survival. This is especially true for motion and depth: objects may move rapidly in the three-dimensional world, and detecting their trajectories and the likelihood that they will collide with the observer is extremely important. In many cases, this information can only be obtained by drawing comparisons between different retinal images. Comparisons between the images from the two retinas yield stereoscopic vision, which can provide information about the relative distances of objects from the observer. Specifically, the relative position of an object on the two retinas, called *binocular disparity*, specifies the distance of that object from the plane of fixation. Similarly, comparing retinal images from successive moments in time yields information about the direction and speed of moving objects. The formal similarity between computations of motion and stereoscopic depth has been noted (Marr, 1982; Qian and Andersen, 1997), but it is not clear how these computations are combined within the receptive fields of individual cortical neurons.

One obstacle to this type of investigation is the technical difficulty of studying stereopsis. In alert animals, vergence movements of the two eyes may disrupt measurements of neural responses to stereoscopic stimuli during prolonged viewing. On the other hand, in anesthetized animals, the relative position of the two eyes cannot be known, so that the binocular disparity is basically unknown. Both of these difficulties can be overcome by using rapid stimulus presentations while animals are awake and fixating (Cumming and DeAngelis, 2001). The stimuli are presented at random positions within the receptive field, and a reverse-correlation analysis of second-order interactions (Szulborski and Palmer, 1990) can then be used to determine the neural selectivity for disparity. The result of this analysis is a high-resolution map of the receptive field's binocular interaction substructure, which in turn provides a picture of the receptive fields of afferent neurons.

The earliest level at which both disparity and motion sensitivity are found in primates is the primary visual cortex (V1). V1 projects to the middle temporal area (MT), where nearly all cells are selective for the direction of moving stimuli, and most cells encode binocular disparity (Maunsell and Van Essen, 1983). We have used the reverse-correlation technique to map the receptive fields of neurons in V1 and MT of alert macaques. Our results reveal a clear two-dimensional substructure for directional interactions, and, in the same cells, a two-dimensional substructure for disparity interactions. In both V1 and MT, the interactions for motion and disparity were much smaller than the receptive fields from which the maps were obtained. Both types of substructure were studied quantitatively in terms of their relationship to each other and to the responses of the same cells to more conventional stimuli.

A quantitative picture of V1 and MT receptive field substructure allows us to address a number of questions regarding the hierarchical organization of the primate visual pathways. First, we can relate the small-scale interactions within the receptive field to the behavior of the neuron when it is stimulated with larger stimuli (Gaska et al., 1994). For example, one might imagine that the simplest way to construct a large direction-selective receptive field would be to sum the outputs of many direction-selective neurons with small receptive fields. Indeed, this finding has been reported in V1 (Emerson et al., 1987) and MT (Movshon and Newsome, 1996; Britten and Heuer, 1999; Livingstone et al., 2001). In the present work we have extended this finding to disparity interactions. Second, it is of great theoretical interest to understand the spatial structure of the inputs to motion and disparity detectors in V1 and MT. A key question is whether or not they are oriented, and if so, how the orientation relates to the measured disparity or motion direction (Movshon et al., 1985; Morgan and Castet, 1997). We have found that, for the majority of cells in V1 and MT, the receptive field substructure is oriented. Furthermore, our results suggest that the direction and disparity inputs share a common orientation. Finally, we can measure the relationship between spatial and temporal interocular disparities. Because both cues are involved in depth and motion perception (Burr and Ross, 1979), it is useful to know how they are combined in the visual cortex (Qian and Andersen, 1997; Anzai et

*Correspondence: cpack@hms.harvard.edu

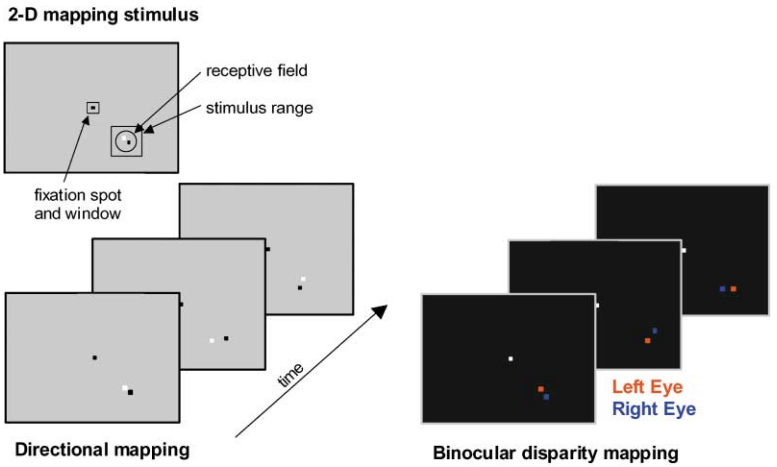

Figure 1. Stimulus Procedure for Two-Dimensional Mapping

(Upper left) Two small squares were presented during each frame at random positions within a predetermined range (large square) in the cell's receptive field (circle). For directional interaction maps, one square was white, and one was black. For binocular disparity maps, one square was visible to the left eye (red), and one square was visible to the right eye (blue). The monkeys maintained fixation on a small target while the stimuli were presented. For MT cells, the receptive field was typically much larger than the stimulus range.

al., 2001). Our findings suggest there are cells in both V1 and MT that integrate spatial and temporal disparity cues. This type of integration may be a substrate for a perceptual effect known as the Pulfrich phenomenon (Pulfrich, 1922).

Results

Two-Dimensional Spatial Interactions

Neurons were first screened with drifting bars or dots, and if they exhibited both direction and disparity tuning, they were further characterized using the one- and two-dimensional sparse noise stimuli (Szulborski and Palmer, 1990) depicted in Figure 1. Such cells, which were easily found in both V1 and MT, were then studied with white noise. All the neurons included in this study responded well to white noise, yielding clear and reliable interaction maps. The receptive field diameters ranged from 0.3° to 1.3° in V1 and from 3° to 10° in MT. In total, we obtained complete one- and two-dimensional disparity and directional maps from 25 single units in V1 and 25 single units in MT of four alert macaque monkeys. The majority (22/25) of the V1 cells were complex. For 2D direction-selectivity mappings (Figure 1, left), each frame of the stimulus consisted of two small squares (0.25°), one black, one white, on a gray background, flashed at 75 Hz at random positions within a 2.5° square stimulus range. For 2D binocular disparity mappings (Figure 1, right), the animals wore goggles containing a differently colored filter over each eye, and the stimuli were colored so that each square was only visible to one eye on each stimulus presentation. The luminances of the squares were adjusted so that they were equally bright when seen through their respective filters. 1D maps were obtained using a similar protocol, but the stimuli were bars instead of spots, and the positions were constrained to lie along a line perpendicular to the bars' orientation. The bar orientation was chosen to match the preferred orientation of the neuron under study. Disparity and directional interactions were measured separately, in sequential stimulus runs.

By presenting stimulus sequences like those in Figure 1 for 5–30 min, we obtained a spike train from each cell, along with the corresponding record of several thousand stimulus presentations. Spikes were then reverse correlated with the difference in position of two different stimuli, one of which we refer to as the *reference* stimulus, and the other as the *probe* stimulus. The reference stimulus was defined for each spike as the stimulus that preceded the spike by a fixed correlation delay (the time to peak response— usually 40–60 ms). The other, *probe* stimulus was either the other stimulus in the same frame, for disparity interaction mapping, or one of the two stimuli in the immediately preceding frame, for directional interaction mapping. Activity was mapped as a function of probe stimulus position minus the reference stimulus position, in 2D retinal coordinates, with the horizontal axis corresponding to the horizontal separation, and the vertical axis corresponding to the vertical separation. Thus, position (0,0) represents occasions when the two stimuli fell in exactly the same location. To generate the direction interaction maps, we added the responses to same-contrast stimulus sequences (white-to-white and black-to-black), and subtracted opposite contrast responses (black-to-white and white-to-black), as described in Livingstone et al. (2001). This is equivalent to a second-order Wiener-like kernel (Emerson et al., 1987).

Two-dimensional direction and disparity interaction maps obtained from one V1 complex cell are shown in Figure 2. The first map (Figure 2A) shows the directional interactions. The red regions to the left and down from the center of the map indicate that the response to a reference stimulus was facilitated when it was preceded by a probe stimulus in which a spot of the same contrast appeared to its left and downward. Conversely, the opposite stimulus sequence suppressed the response, as indicated by the bluish regions upward and rightward of the map's origin. Both the red and blue regions also contain contributions from opposite contrast (white-to-black and black-to-white) sequences, as described in Livingstone et al. (2001). Figure 2B shows the directional tuning measured with a standard moving bar stimulus. Here the peak response occurred for motion to the right and slightly upward, in good agreement with the directional preference found in the interaction map.

A binocular interaction map for the same cell is shown in Figure 2C. Here we map the interactions between stimuli presented simultaneously at different retinal locations in the two eyes. The coordinates of the map

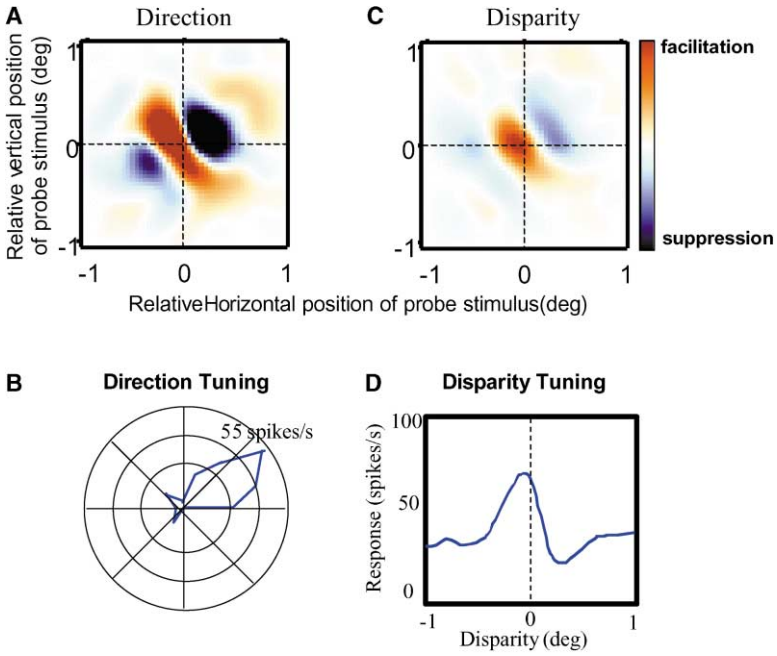

Figure 2. Interaction Maps for a V1 Complex Cell

Red indicates facilitation, and blue indicates suppression, as indicated by the color scale to the right of the map in (C).

(A) Directional interactions. The axes indicate the average effect of the 2D position of a probe stimulus relative to the position of subsequently presented reference stimulus. The reference position is always (0,0). The response of this cell was facilitated by a sequence of stimuli moving upward and rightward, and suppressed by the opposite sequence.

(B) The direction tuning curve as measured by drifting bars. The tuning curve is in polar coordinates, with angle representing the direction of stimulus motion, and radius representing the strength of the response. This cell preferred upward and rightward motion, which is consistent with the interaction map.

(C) Disparity interactions for the same cell. The axes indicate the position of a probe stimulus in the left eye relative to the position of a simultaneously presented reference stimulus in the right eye. The position of the reference stimulus is always (0,0). The response of this cell was facilitated when the right-eye stimulus was to the left of the left-eye stimulus (crossed disparity), and suppressed for uncrossed disparities.

(D) The disparity tuning curve as measured with flashed bars. This cell preferred near (crossed) disparities, which is consistent with the interaction map.

indicate the position of the right-eye stimulus relative to the position of the left-eye stimulus, which is assigned to (0,0). This cell showed facilitation (red regions in the interaction map) when the right-eye stimulus was to the left of the left-eye stimulus (a near stimulus), and was suppressed when the opposite configuration occurred. This preference for crossed disparity means that the cell should respond best to stimuli that appear to be in front of the plane of fixation, and indeed such a preference is observed. Figure 2D shows a more conventional disparity-tuning curve, for the same cell, obtained with flashed bar stimuli. The best response occurred for near disparities, and suppression was evident for far disparities.

Figures 3A and 3D show a pair of maps obtained from an MT cell, again using the stimulus configuration depicted in Figure 1. For this cell, and for many others, it was possible to obtain reliable disparity and directional interactions between stimuli that were displaced by only one pixel on the monitor (0.06°). This is remarkable considering that the MT receptive fields were typically on the order of 100 times this size. Nonetheless, the conventional directional and disparity tuning for bar stimuli shown in Figures 3B and 3E agree well with the interactions between pairs of spot stimuli shown in Figures 3A and 3D, respectively.

Subunits

The maps in Figures 2 and 3 describe the dependence of the neuronal response on the interactions between stimuli in the receptive field. In keeping with previous reports (Movshon et al., 1978; Emerson et al., 1987; Szulborski and Palmer, 1990; Anzai et al., 1999), we will refer to these interaction maps as *subunits*.

The subunits generally consisted of two or more subregions of facilitatory and suppressive interactions. It has been observed previously that subunits can be repeated fairly precisely across the receptive fields of individual neurons (Movshon et al., 1978; Baker and Cynader, 1986; Emerson et al., 1987; Szulborski and Palmer, 1990; Anzai et al., 1999; Livingstone et al., 2001), so it seems likely that the structure of the subunits embodies a principle by which these neurons select and organize their feedforward inputs. We therefore made quantitative measurements of the subunit structure for all of our interaction maps.

Inspection of Figures 3A and 3D suggests several types of measurements that could be of interest in understanding the functional properties of this MT cell's receptive field. First, as we mentioned previously, the angle at which the center of the facilitatory region (the bright red region) occurs suggests a motion preference (in this case up and to the left), which should be reflected in the cell's tuning for more conventional stimuli. We can therefore compare the angle between the facilitatory center and position (0,0) in such maps to the cell's actual preferred direction. Second, the distance of this peak from the (0,0) point gives us an indication of the spatial scale of the interactions between stimuli inside the receptive field. Third, the shape of the subunit itself in Figure 3 is elongated, which suggests a selectivity for orientation in the afferents. We can therefore ask how this elongation relates to the cell's preferences for motion and disparity. A fourth quantity, the curvature of the subunits, has been studied extensively in other work (Livingstone et al., 2001; Livingstone and Conway, 2003) and will not be examined further here.

To obtain objective, quantitative descriptions of the binocular and directional subunit structure in V1 and in

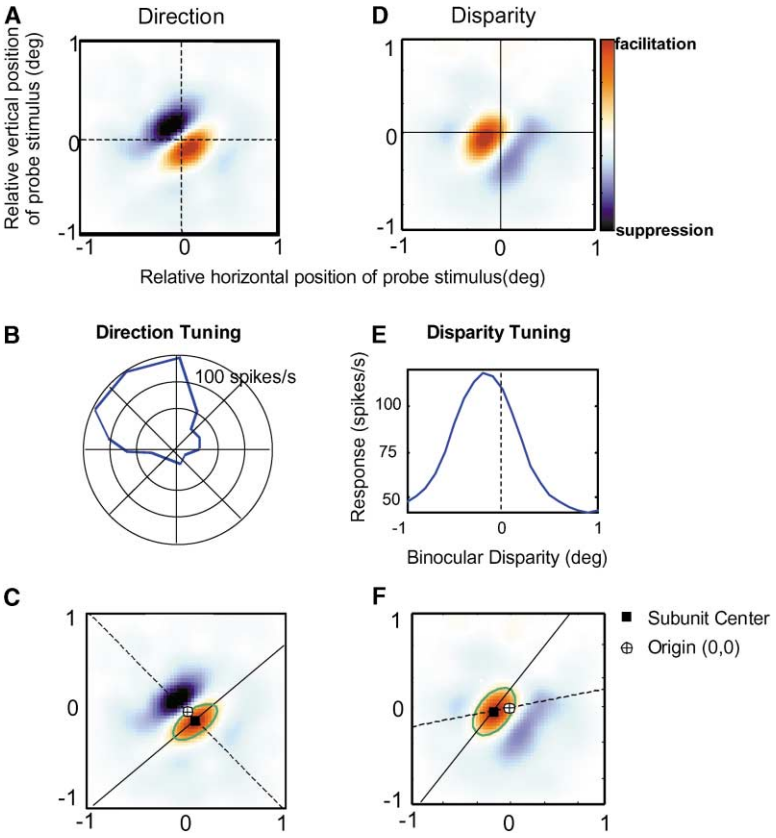

Figure 3. Interaction Maps for an MT Cell

(A) and (B) show direction interactions and direction tuning. (D) and (E) show the disparity interactions and disparity tuning for the same cell. (C) and (F) show the result of fitting the interaction maps in (A) and (D) to elliptical Gaussians. The green ellipses indicate the boundaries of the Gaussian. The solid oblique lines indicate the orientations of the Gaussians, and the dashed oblique lines connect the origin of the map (position (0,0)) with the centers of the Gaussians, which are indicated by the black squares.

MT, we fit maps for each of the 50 cells with a two-dimensional elliptical Gaussian. Each map was truncated at a point 20% below its peak, and the remaining structure was fit with a 7-parameter model that described the center, width, height, and orientation of the subunit. (We also examined other analytical methods, such as Fourier analysis, the Radon transform, and fits to a Gabor function. Not surprisingly, the different methods yielded similar results, but the Gaussian had the advantage of providing a simultaneous measure of the orientation and center of the subunits, with a modest number of parameters.) Figures 3C and 3F show the results of the Gaussian fits to the interaction maps shown in Figures 3A and 3D. The green ellipses denote the boundaries of the Gaussians, and the black squares indicate the centers. The orientations of the Gaussians are indicated by the solid oblique lines, and the angle between the position of the reference stimulus (0,0) and the center of Gaussian in each map is indicated by the dashed oblique lines.

The positions of the centers of the Gaussians are plotted in Figure 4 for both the V1 and the MT population. In panels A and B, the *location* of each arrow indicates the center of the directional interactions for one neuron, and the *direction* of each arrow is the preferred direction of the cell, as measured with moving dots or bars. In panels C and D, the circles indicate the centers of the disparity interactions—open circles are the cells that preferred far disparities, and closed circles indicate near disparities, as measured with flashed bars.

The positions of the subunit centers correlate well with the preferences for motion direction and disparity. For motion direction, the correlation between the subunit position and the preferred direction is seen in the tendency for the arrows in Figures 4A and 4B to point toward the center of the plots. This correlation is highly significant (angular-angular correlation, $p < 0.0001$) (Mardia, 1975). Similarly, for disparity, the segregation of closed and open circles on either side of the vertical line in Figures 4C and 4D indicates that the disparity preference of most cells was well predicted by the position of the subunit centers. For both motion and disparity, the optimal interaction distances were extremely small in both V1 and MT. In V1 the mean subunit center was 0.13° from (0,0) for direction and 0.09° for disparity. In MT the mean center positions were 0.19° and 0.10°. Although the V1 and MT populations were not precisely matched for receptive field eccentricity, there was substantial overlap in the distributions. The V1 cells had receptive field eccentricities between 2° and 15°, and the MT cells between 2° and 11°. It is clear that the two areas respond to a similar range of directional and disparity interactions. However, this conclusion is probably limited to two-stimulus interactions, since it has been shown previously that MT responds to larger interstimulus intervals than V1 if more than two successive stimuli are analyzed (Mikami et al., 1986).

As described above, we also analyzed the orientation of the subunits. By inspection of Figures 2 and 3, it appears that the subunits are oriented, and it has been suggested that this orientation reflects the orientation preference of afferent neurons (Szulborski and Palmer,

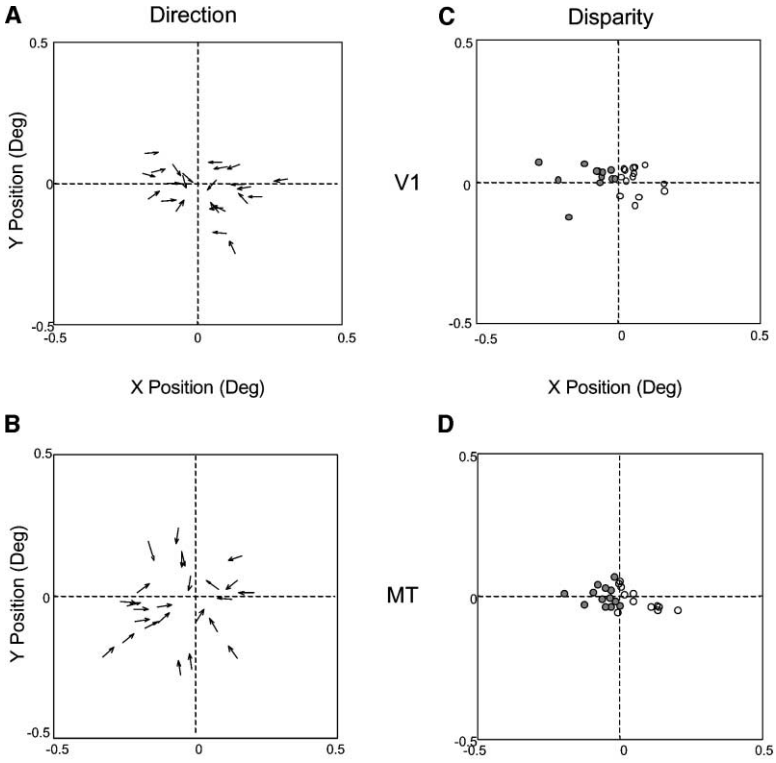

Figure 4. Subunit Centers

The centers were determined by fitting interaction maps like those in Figures 2 and 3 to elliptical Gaussian functions.

(A) Centers of the directional subunits in 25 V1 cells. The position of each arrow indicates the position of the center, and each arrow's direction indicates the preferred direction of the cell, as measured with drifting bars or dots.

(B) As in (A), but for 25 MT cells.

(C) Centers of the disparity subunits in the same 25 V1 cells. The position of each dot indicates the position of the center. Closed dots represent cells preferring near disparities, and open dots represent cells preferring far disparities, as measured with flashed bars.

(D) As in (C), but for the same 25 MT cells.

1990). If this is true, then the orientation of the subunits can tell us something about the way in which V1 and MT neurons organize their inputs. Specifically, the relationship between subunit orientation and each neuron's preference for direction and disparity bears directly on the controversial issues of how global direction and disparity are computed from local inputs (Movshon et al., 1985; Morgan and Castet, 1997).

We first examined the hypothesis that the orientation of the V1 subunits was perpendicular to the angular position of the centers. That is, subunits centered on the horizontal axis of the interactions maps should have vertical orientations, those centered on the vertical axis should have horizontal orientations, and so forth. The difference between the subunit orientation and angular position is shown graphically by the intersections of the solid and dashed oblique lines in Figures 3C and 3F.

For this analysis, we excluded six V1 cells because the Gaussian fits did not indicate that the major axis of the fitted ellipse was at least 50% longer than the minor axis. For the remaining 19 V1 cells, the subunit orientations were well predicted by the positions of the centers. An angular-angular correlation revealed that the orientations of the directional subunits were generally perpendicular to the position of the centers of the directional interactions ($p < 0.01$). Similarly, the orientations of the binocular subunits were perpendicular to the positions of the centers of the binocular interactions ($p < 0.01$). However, because the V1 subunits for direction and disparity tended to cluster close together, we cannot say whether the subunits' orientations are determined by direction, disparity, or some combination of both. We can only say that their orientations are reasonably close to being perpendicular to their angular positions.

The orientation distribution in MT was broader than in V1, which enabled us to observe a clearer organization. Figure 5A illustrates the relationship between the center of the directional interaction and the orientation of the corresponding subunit for 21 MT cells. The other 4 MT cells did not have clear subunit orientations and so are not included in this plot. In the majority of cases, the subunits had orientations that were perpendicular to the cells' preferred motion direction, although a few cells deviated somewhat from this tendency. As in V1, the tendency for the subunit orientation to be perpendicular to its position was highly significant ($p < 0.001$; angular-angular correlation); however, the same was not true of the binocular subunits. The tendency for the orientations of the disparity subunits to be perpendicular to the angular positions of the disparity subunit centers was much weaker and did not reach significance (angular-angular correlation, $p > 0.2$). This is shown in Figure 5B. However, the orientation of the disparity subunits did tend to be perpendicular to the positions of the *directional* subunits. Figure 5C shows this relationship between the centers of the directional interaction and the orientation of the corresponding disparity subunit for the same 21 MT cells. This relationship between the disparity subunit orientations and the directional subunit positions was statistically significant (angular-angular correlation, $p < 0.02$). Moreover, the orientations of the directional and disparity subunits were well correlated with each other (angular-angular correlation, $p < 0.001$). The histogram in Figure 5D shows the difference between the subunit orientations for disparity and direction for the MT population. The distribution is centered near zero, with a mean absolute angular deviation of 22.6°. Thus the orientations of the MT subunits for disparity and direction

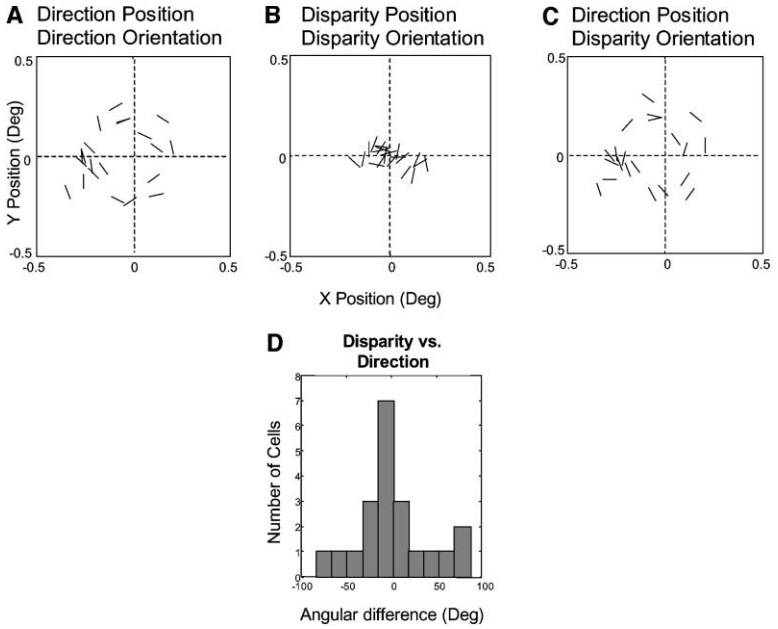

Figure 5. Relationship between the Location of MT Subunit Centers and Subunit Orientations

(A) Orientation of the directional subunits for 21 MT cells. The position of each line indicates the position of the subunit center, and the orientation of the line indicates the subunit's orientation. The orientation is generally perpendicular to the angle that generates the strongest facilitation.

(B) Orientation of disparity subunits for the same 21 MT cells. The position of each line indicates the position of the *disparity* subunit, and the orientation of the line indicates the orientation of the corresponding disparity subunit.

(C) Orientation of disparity subunits for the same 21 MT cells. The position of each line indicates the position of the *directional* subunit, and the orientation of the line indicates the orientation of the corresponding *disparity* subunit.

(D) Distribution of the difference between the orientations of the directional and disparity subunits for the population of 21 MT cells. The distribution is centered near zero.

are quite similar and tend to be linked to each cell's direction preference. Indeed, both of these tendencies can be observed in the example MT cell shown in Figure 3, where the subunit orientation is perpendicular to the position of the motion interaction (Figure 3A), but not to the position of the disparity interaction (Figure 3D). In addition, close inspection of Figures 3A and 3D reveals another typical feature of the maps: the binocular subunit was slightly more vertically oriented than the directional subunit. In this case, the binocular and directional subunit orientations were 37° and 53° from vertical, and a similar trend is apparent in the population in Figure 5C. Thus the orientation of the MT subunits appears to reflect a compromise between the directional and disparity centers, but one that is weighted toward the directionality of each cell.

Because the directional stimulus was binocular, it is possible that the correlation between the orientations of the directional and disparity subunits was due to disparity energy present in the direction mapping stimulus. In particular, visual persistence of the probe stimulus from one frame to the next might lead to spurious correlations between the directional and disparity subunit structure. We therefore performed a control analysis in which we calculated the subunits for *monocular* direction maps. These maps were obtained by reverse correlating the spike trains used to compute the disparity maps with the positions of the stimuli in one eye on subsequent frames. This analysis yielded elongated directional subunits in at least one eye for 16 MT cells. The same tendency was observed for the monocular directional subunits as for the binocular directional subunits: both the directional and disparity subunit orientations were perpendicular to the angular position of the direction subunits ($p < 0.05$, angular-angular correlation), and the disparity subunits were not perpendicular to the disparity subunit positions ($p > 0.2$, angular-angular correlation). We conclude that the binocular nature of the directional mapping stimulus was not the cause of the correlations between the disparity and directional subunits.

Space-Time Interactions

The measurement of spatial disparity becomes problematic for objects moving at high velocities because a rapidly moving object stimulates each retinal position very transiently. In this situation, the visual system seems to make use of the fact that an object moving in depth also generates a *temporal* disparity (Morgan and Castet, 1995). That is, a moving object will stimulate corresponding points on the two retinas with some interocular delay, and this cue is sufficient to generate a robust percept of depth, even in the absence of spatial disparity (Burr and Ross, 1979). A dramatic example of the interplay between spatial and temporal disparities is manifested in a class of perceptual phenomena known as Pulfrich effects (Pulfrich, 1922). The basic Pulfrich phenomenon can be observed by watching a pendulum swinging back and forth in a plane perpendicular to the line of sight. If one eye's view is slightly delayed, for instance by a light-attenuating filter, the pendulum appears to follow an elliptical path in depth (Figure 6).

A simple explanation for this phenomenon is that the temporal delay introduces a spatial disparity between the views of the pendulum in the two eyes. In Figure 6A, this is represented as the horizontal displacement of the black diagonal lines relative to the red diagonal lines. If the brain interpreted this horizontal displacement as a spatial disparity, a percept of stereoscopic depth would naturally result. However, this explanation is not sufficient to explain the percept in Figure 6B where the stimulus is a series of spots flashed stroboscopically with a slight delay between the two eyes. Because only one monocularly viewed spot is present at any instant, there is no spatial disparity, and so the resulting illusory percept of depth (Burr and Ross, 1979) must depend at least partially on a neural mechanism that is sensitive to temporal disparity.

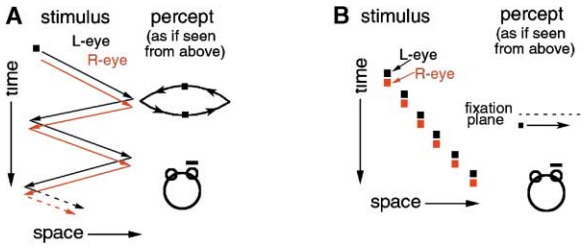

Figure 6. Pulfrich Phenomena

(A) The observer views a pendulum swinging along a straight path perpendicular to the line of sight. If one eye's view is delayed, the path of the pendulum differs in both space and time between the left eye (black zig-zag line) and the right eye (red line).

(B) If a moving spot is presented stroboscopically with a small delay between the two eyes, the spot appears to be outside the plane of fixation, even though there is no simultaneous interocular spatial disparity.

Such a mechanism could provide a robust means of encoding the depth of a moving object, because depth can be recovered from combinations of spatial and temporal disparity. Moreover, at a given depth, the spatial and temporal disparities are related linearly (by the object's velocity), suggesting a straightforward means of encoding depth (Qian and Andersen, 1997; Anzai et al., 2001). Physiologically, a receptive field tuned to combinations of spatial and temporal disparities should have a distinct signature: it should change its preferred spatial disparity as a function of temporal disparity (Qian and Andersen, 1997).

We investigated this possibility in our population of V1 and MT cells, again using the sparse noise reverse-correlation technique. However, in order to obtain greater temporal resolution, we limited the visual stimuli to one spatial dimension. The presentation of stimuli was as shown in Figure 1: each frame consisted of two differently colored stimuli so that each was viewed by only one eye, but the stimuli were now oriented bars instead of dots. The bars were oriented perpendicular to each cell's preferred direction and constrained to appear along a line parallel to the preferred null axis of motion. The analysis was conceptually similar to that used in the 2D maps—an optimal correlation delay was found, and the spike train was reverse correlated with the positions of the reference and probe stimuli at that delay. Spiking activity was mapped as a function of the relative positions of the reference and probe stimuli, irrespective of their actual positions in visual space. Space-time maps were generated by reverse correlating the spike train with pairs of stimuli separated by different temporal intervals. Specifically, we reverse correlated each spike with the positional difference between the reference stimulus and each of four preceding probe stimuli, spanning a total of 112 ms. The results were examined as a function of both time and one-dimensional space.

Figure 7 shows examples of this kind of analysis in three neurons from V1 (top) and three neurons from MT (bottom). In each panel, the *x* axis indicates the binocular disparity between the reference and probe stimulus, and the *y* axis indicates the interocular delay between stimulus presentations. Figure 7A shows space-time maps obtained from a V1 cell (top) and an MT cell (bottom). Both cells preferred near disparities and were relatively unresponsive to nonzero interocular delays. As a result, these cells can encode depth only through spatial disparity. Figures 7B and 7C show V1 (top) and MT cells (bottom) that were more sensitive to temporal disparities. The space-time slant evident in these maps means that these cells responded strongly to specific combinations of spatial and temporal disparities. As a result, they would not be able to distinguish a temporal delay between the presentation of the stimuli in the left and right eyes (as in Pulfrich's pendulum) from an actual retinal disparity. This explanation for the Pulfrich phenomenon is an explicit prediction of models that integrate depth and motion information (Qian and Andersen, 1997).

In order to quantify the degree of space-time slant in our maps, we measured the tilt direction index (TDI) described by Anzai et al. (2001). This index assigns to each map a value between 0 and 1, where 0 indicates no slant, and 1 indicates that the map is completely described by one direction of tilt. The V1 cells in the top row of Figures 7A–7C (from left to right) had TDIs of 0.02, 0.20, and 0.68, respectively. The corresponding MT cells in the bottom row of Figure 7 had TDIs of 0.10, 0.53, and 0.86. The mean TDI for the V1 population is 0.17, and in MT the mean TDI is 0.41. It is important to note that the mean TDI values in V1 and MT do not necessarily capture the prevalence of neural selectivity for spatiotemporal disparity. In fact, the white noise analysis probably underestimates the number of such cells in both areas, simply because of the low signal-to-noise regime in which it necessarily operates. Ideally, had we accumulated many more spikes for each neuron, we might have found a larger mean TDI across the populations in V1 and MT, but in practice, the spike counts are limited by the exigencies of recording single neurons from alert animals.

In a similar study, Anzai et al. (2001) used dense white noise (16 bars per frame) stimuli and found a larger mean TDI in their population of V1 complex cells recorded in anesthetized cats. In fact their distribution of V1 TDIs was quite similar to our distribution of TDIs in MT. We cannot say whether this discrepancy is due to species differences (monkey versus cat), differences in behavioral state (alert versus anesthetized), differences in the spatial distribution of stimuli (sparse versus dense noise), or temporal differences in the rate of presentation of stimuli. The latter possibility seems unlikely because we tried a variety of different presentation rates in our V1 population, ranging from 14 to 100 ms presentations, without observing any obvious difference in the TDI. However, at present none of the above possibilities can be ruled out.

A previous study in alert macaque V1 (Perez et al., 1999) reported some cells that were tuned to nonzero temporal disparities for stimuli presented at zero spatial disparity. Our findings do not contradict this result, but when we examined the V1 and MT tuning to both spatial and temporal disparities, we found that the peak response always occurred at zero temporal disparity, often in conjunction with a nonzero spatial disparity. A similar result was reported by Anzai et al. (2001). Thus

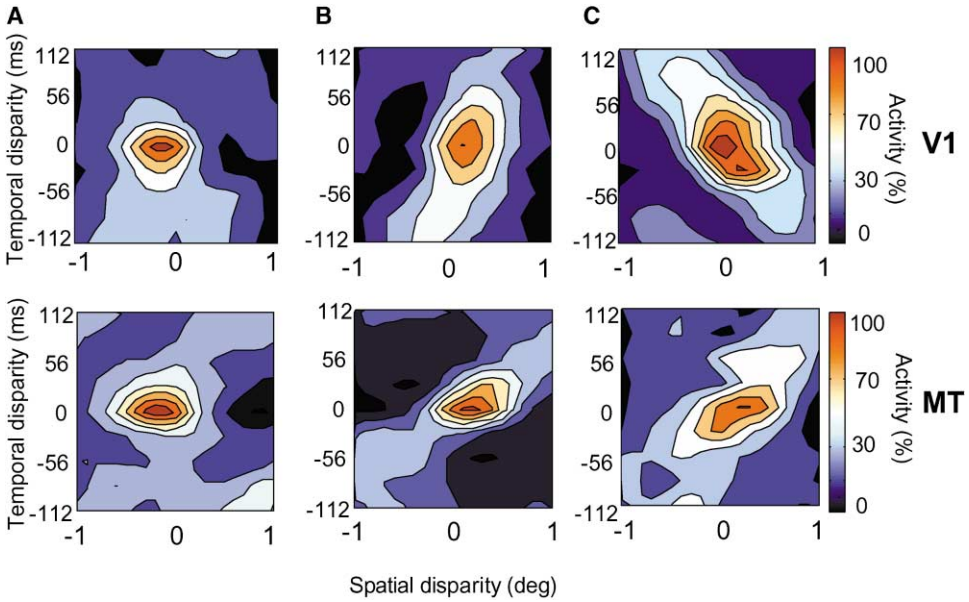

Figure 7. Neural Selectivity for Combinations of Spatial and Temporal Disparity

The *x* axis of the contour maps encodes spatial binocular disparity, and the *y* axis encodes temporal disparity.

(A) Space-time binocular interaction map for a V1 complex cell (top) and an MT cell (bottom). These cells are selective for instantaneous spatial disparity, but have no selectivity for temporal disparity.

(B) A V1 cell (top) and an MT cell (bottom) exhibiting modest slant in binocular space-time.

(C) A V1 cell (top) and an MT cell (bottom) that change their preferred spatial disparity when the temporal disparity is changed. These relationships are close to being linear, as indicated by the TDIs of 0.68 for the V1 cell and 0.86 for the MT cell.

it does not appear that the visual cortex explicitly measures interocular temporal disparity.

Discussion

Using sparse white noise, we have measured the two-dimensional substructure of directional and disparity interactions in V1 and MT of alert monkeys. The resulting subunit structure indicates interactions for both retinal disparity and motion direction that occur on a spatial scale that is much smaller than the receptive fields in either area. The interactions are extremely precise spatially, and correlate well with neuronal preferences for disparity and motion, as measured with conventional stimuli. The shape of the subunits of individual receptive fields is similar for motion and stereo, suggesting a common organizing principle for feedforward inputs. The temporal dynamics of the interocular interactions show an intriguing interplay between motion and stereo, which may be related to a perceptual phenomenon known as the Pulfrich effect.

Visual Hierarchy

Our two-dimensional mapping technique allows us to visualize the subunits of receptive fields in V1 and MT. The subunits are generally elongated along one dimension and contain alternating facilitatory and suppressive subregions along the orthogonal dimension. The subregions are sensitive to the sign of stimulus contrast for both direction (Livingstone et al., 2001; Livingstone and Conway, 2003) and disparity (Livingstone and Tsao, 1999) interactions, and are responsive only over a small range of stimulus separations. All of these observations are consistent with the idea that the subunits represent input from V1 simple cells. Of course, each subunit may represent an average of many inputs, but the clear structure that is evident in the interaction maps suggests that there is a remarkable degree of similarity among the receptive fields of the afferent neurons.

Simple cells provide input to complex cells (Martinez and Alonso, 2001), and complex cells provide input to MT cells (Movshon and Newsome, 1996). Our results on the orientation of subunits suggest that the basic physiological properties of the receptive fields are preserved throughout the hierarchy. Moreover, the properties of neurons at low levels of the hierarchy seem to be instrumental in determining the properties of neurons at higher levels. For example, the preferred motion direction of an MT cell can be predicted with reasonable precision from the orientation of the subunit (Figures 4 and 5). Because preferred orientation and preferred motion direction are linked in V1 cells, this suggests a natural basis for organizing a direction-selective receptive field. Furthermore, at least in MT, the disparity interactions seem to be subordinate to this organization, since their orientations are also perpendicular to the preferred motion direction. At the same time, it is clear that information is substantially altered as it moves along the visual hierarchy. Compared to simple cells, complex cells are relatively insensitive to the contrast polarity and spatial position of a stimulus, and MT cells are even less sensitive to these parameters. Compared to complex cells, the directional responses of MT cells are sensitive to motion over a broader range of spatial and temporal stimulus parameters (Mikami et al., 1986). Simi-

larly, our results suggest that MT cells may be sensitive to a broader range of spatiotemporal depth cues than are V1 complex cells. Furthermore, many MT cells are able to compute the direction of motion for stimuli that contain multiple orientations (Movshon and Newsome, 1996; Pack et al., 2001) or second-order motion (Albright, 1992), whereas neither behavior is found in complex cells (Movshon and Newsome, 1996; O'Keefe and Movshon, 1998). Lastly, even though the inputs to MT cells are orientation selective, the outputs of many MT cells are dominated by terminators and endpoints, which are not oriented (Duncan et al., 2001; Pack and Born, 2001). It remains to be seen whether these transformations are constructed by selective weighting of afferent input, or created de novo at each stage.

Functional Integration of Motion and Stereo

Our observations are consistent with previous findings that single neurons in primate V1 and MT are sensitive to both motion and disparity (Maunsell and Van Essen, 1983; Bradley et al., 1995; DeAngelis and Newsome, 1999; Prince et al., 2002). Functionally, this combination of selectivities is beneficial to both types of computations—information about depth guides the integration and segregation of ambiguous motion signals (von Grunau et al., 1997; Stoner and Albright, 1998; Duncan et al., 2001), and motion signals can facilitate binocular matching (Van Ee and Anderson, 2001). Joint selectivity for motion and disparity appears to be necessary to account for human stereoacuity, which can be maintained for stimulus velocities as high as 640°/s (Morgan and Castet, 1995). Cells tuned only to static disparities would need to resolve interocular temporal differences of less than 1 ms to maintain sharp disparity tuning for such rapidly moving stimuli (Morgan and Castet, 1995). Furthermore, the combination of motion and disparity information can be used for a broad range of perceptual and behavioral tasks, including figure-ground segmentation (Bradley et al., 1995; Bradley and Andersen, 1998), structure-from-motion (Bradley et al., 1998; Dodd et al., 2001; Fernandez et al., 2002), navigation (Roy et al., 1992; Lappe, 1996), and eye movement control (Howard and Simpson, 1989).

An open question concerns how individual neurons develop to combine selectivities for motion and disparity. There are indeed many ways in which simple cell receptive fields could be combined to make complex cells that are tuned to both motion and disparity. Recent theoretical work has shown that, under very modest assumptions about simple cell inputs, the Pulfrich phenomenon emerges at the complex cell level (Qian and Andersen, 1997). We have confirmed this theoretical prediction in macaque complex cells, as have previous experimental findings in the anesthetized cat (Anzai et al., 2001).

Our findings pose a challenge to existing MT models. MT neurons are primarily sensitive to motion direction, and the orientation preference for most neurons is within 30° of being perpendicular to the preferred direction (Albright, 1984). Similarly, we find that the orientations of the MT subunits are approximately perpendicular to each neuron's preferred motion direction. In contrast, orientation preference is unrelated to disparity tuning in V1 (Smith et al., 1997; Prince et al., 2002) and in MT (DeAngelis and Newsome, 1999). For our sample of V1 neurons, the relationship between subunit orientation and preferences for disparity and motion was not clear. It is tempting to conclude that the subunit orientations are perpendicular to their preferred motion direction, an idea that is borne out statistically. However, we cannot at present rule out the possibility that the V1 subunit orientations are tied to the preferred disparity as well. This will be the subject of further study.

In MT, we find that the orientations of MT subunits are unrelated to their disparity preferences (Figure 5B). However, the orientations of MT disparity subunits are perpendicular to their preferred motion directions, suggesting that the local MT receptive field structure is organized around the orientation preferences of afferent inputs. Whether this organization is inherited from V1 complex cells with similar receptive field structures or generated in the connectivity between V1 and MT neurons remains to be seen. In terms of functional implications, it seems reasonable that a correlation between preferred orientations for disparity and motion inputs would represent a sensible way to encode the velocity of moving edges at different depth planes. Developmentally, this may be achieved through correlations between depth and motion that occur naturally during self-motion (i.e., motion parallax). A strong prediction of this hypothesis is that individual neurons should exhibit a correlation between their preferred speed and the magnitude of their preferred disparity. Such a correlation has been observed in the cat (Anzai et al., 2001), but has not yet been found in macaque MT (DeAngelis and Newsome, 1999).

Experimental Procedures

Monkeys were prepared for chronic recording from V1 and MT (Livingstone, 1998; Born et al., 2000). All procedures were approved by the Harvard Medical Area Standing Committee on Animals.

We recorded from single units in V1 and MT of four alert rhesus macaque monkeys while they performed a simple fixation task. The monkeys were rewarded for maintaining fixation within 1° of a small fixation spot. For each single unit, isolated by spike height and waveform, we first determined the preferred direction of motion using fields of dots or moving bars. We then measured the preferred disparity with moving bars. If the cell appeared to be selective for both disparity and direction, we studied it further. All cells were studied with one- and two-dimensional white noise stimuli to measure both disparity and directional interactions. Each stimulus presentation lasted 13 ms for directional interactions, and 27 ms for disparity interactions.

To measure disparity interactions, color-separation filters were used to stimulate each eye separately and to generate stimuli at different interocular disparities. The energies of the phosphors were adjusted so that the luminance of the red phosphor through the red filter was the same as the luminance of the blue + green phosphors through the cyan filter, as measured with a Pritchard spot photometer. Through the red filter, the luminance of the red stimulus (at the maximum red phosphor setting; red = 255; green = 0; blue = 0) was 3.7 cd/m^2; the setting of the blue and green phosphors was adjusted to give a cyan stimulus that is also 3.7 cd/m^2 through the cyan filter (red = 0; green = 210; blue = 210). Through the cyan filter the luminance of the red stimulus was 0.170 cd/m^2; through the red filter the cyan stimulus was 0.24 cd/m^2. For light-on-dark stimuli, where the red and cyan bars overlap, the stimulus is white (red = 255; green = 210; blue = 210).

The 1D mapping stimulus consisted of two bars presented at random positions along a 1D stimulus range parallel to the preferred

motion axis. The orientation of the bars was perpendicular to the cell's preferred axis of motion.

The 2D mapping stimulus consisted of pairs of small squares presented at random positions at 75 Hz within a square stimulus range. For directional maps, we used white and black stimuli that were 19 cd/m^2 above and below the mean background gray luminance of 20 cd/m^2. When black and white stimuli overlapped, the resultant stimulus was the same gray as the background. For disparity maps, the stimuli were red and cyan, and the monkey wore filtered goggles, as described above. Directional and disparity experiments were conducted sequentially so that the spike trains that determined each map were completely distinct.

A computer recorded the evoked spike train (1 ms resolution), each stimulus position, and the monkey's eye position (4 ms resolution). Vergence was not monitored in this study, but it was monitored in a previous study and found to be stable (Livingstone and Tsao, 1999). For each map, between 5,000 and 30,000 spikes were collected over a 5–30 min period. The stimulus is "white" in the sense that the spatial and temporal autocorrelation functions are flat. It is "sparse" in that the stimulus at each position is the same as the background luminance most of the time. Because the probability distribution of luminances used in our stimulus is not Gaussian, it would not fit the criteria for Wiener kernel analysis. However, Emerson et al. (1987) developed a modified Wiener kernel analysis for use with ternary white noise, and our difference maps are equivalent to their second-order Wiener-like kernel calculation. Spikes were reverse correlated with the positions of sequential stimuli at a delay corresponding to the peak of the response to the second stimulus (typically between 40 and 60 ms); stimulus position was corrected for eye position at stimulus onset. For the 2D interaction maps, spikes were reverse correlated with the *difference* in position between sequential pairs of spots. For the disparity, interaction maps, a baseline first-order map was generated by reverse correlating the same spike train with pairs of stimuli presented 250 ms apart (a time delay when all interactions had disappeared). This baseline map was subtracted from each of the individual paired-contrast maps. The directional maps were calculated by summing the same-contrast individual maps (white-to-white and black-to-black) and subtracting the different-contrast maps (white-to-black and black-to-white), as described in Livingstone et al. (2001).

To study the positions and orientations of the subunits, we fit each interaction map with a 2D elliptical Gaussian function, which has seven free parameters. The fits were optimized via a least-squares criterion with the Levenberg-Marquardt algorithm in Matlab (Mathworks, Natick, MA). This procedure generally agreed with qualitative observations, even for subunits in which the subregions were curved (crescent shaped).

To measure the degree of tilt in the binocular space-time maps, we adopted the method used by Anzai et al. (2001). Briefly, this method involves computing the optimal spatial and temporal frequencies (F_s and F_t), which are found at the peak of the Fast Fourier Transform of each interaction map. The tilt direction index is then given by $(R_p - R_n)/(R_p + R_n)$, where R_p and R_n are the response amplitudes at (F_s, F_t) and $(F_s, -F_t)$.

Acknowledgments

This work was supported by NIH NS07484 (C.C.P.), NIH EY13135 (M.S.L.), and NIH EY11379 and EY12196 (R.T.B.).

Received: May 6, 2002
Revised: November 18, 2002

References

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

Albright, T.D. (1992). Form-cue invariant motion processing in primate visual cortex. Science *255*, 1141–1143.

Anzai, A., Ohzawa, I., and Freeman, R.D. (1999). Neural mechanisms for processing binocular information II. Complex cells. J. Neurophysiol. *82*, 909–924.

Anzai, A., Ohzawa, I., and Freeman, R.D. (2001). Joint-encoding of motion and depth by visual cortical neurons: neural basis of the Pulfrich effect. Nat. Neurosci. *4*, 513–518.

Baker, C.L., and Cynader, M.S. (1986). Spatial receptive-field properties of direction-selective neurons in cat striate cortex. J. Neurophysiol. *55*, 1136–1151.

Born, R.T., Groh, J.M., Zhao, R., and Lukasewycz, S.J. (2000). Segregation of object and background motion in visual area MT: effects of microstimulation on eye movements. Neuron *26*, 725–734.

Bradley, D.C., and Andersen, R.A. (1998). Center-surround antagonism based on disparity in primate area MT. J. Neurosci. *18*, 7552–7565.

Bradley, D.C., Qian, N., and Andersen, R.A. (1995). Integration of motion and stereopsis in middle temporal cortical area of macaques. Nature *373*, 609–611.

Bradley, D.C., Chang, G.C., and Andersen, R.A. (1998). Encoding of three-dimensional structure-from-motion by primate area MT neurons. Nature *392*, 714–717.

Britten, K.H., and Heuer, H.W. (1999). Spatial summation in the receptive fields of MT neurons. J. Neurosci. *19*, 5074–5084.

Burr, D.C., and Ross, J. (1979). How does binocular delay give information about depth? Vision Res. *19*, 523–532.

Cumming, B.G., and DeAngelis, G.C. (2001). The physiology of stereopsis. Annu. Rev. Neurosci. *24*, 203–238.

DeAngelis, G.C., and Newsome, W.T. (1999). Organization of disparity-selective neurons in macaque area MT. J. Neurosci. *19*, 1398–1415.

Dodd, J.V., Krug, K., Cumming, B.G., and Parker, A.J. (2001). Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT. J. Neurosci. *21*, 4809–4821.

Duncan, R.O., Albright, T.D., and Stoner, G.R. (2001). Occlusion and the interpretation of visual motion: perceptual and neuronal effects of context. J. Neurosci. *20*, 5885–5897.

Emerson, R.C., Citron, M.C., Vaughn, W.J., and Klein, S.A. (1987). Nonlinear directionally selective subunits in complex cells of cat striate cortex. J. Neurophysiol. *58*, 33–65.

Fernandez, J.M., Watson, B., and Qian, N. (2002). Computing relief structure from motion with a distributed velocity and disparity representation. Vision Res. *42*, 883–898.

Gaska, J.P., Jacobson, L.D., Chen, H.W., and Pollen, D.A. (1994). Space-time spectra of complex cell filters in the macaque monkey: a comparison of results obtained with pseudowhite noise and grating stimuli. Vis. Neurosci. *11*, 805–821.

Howard, I.P., and Simpson, W.A. (1989). Human optokinetic nystagmus is linked to the stereoscopic system. Exp. Brain Res. *78*, 309–314.

Lappe, M. (1996). Functional consequences of an integration of motion and stereopsis in area MT of monkey extrastriate visual cortex. Neural Comput. *8*, 1449–1461.

Livingstone, M.S. (1998). Mechanisms of direction selectivity in macaque V1. Neuron *20*, 509–526.

Livingstone, M.S., and Conway, B.R. (2003). Substructure of direction-selective receptive fields in macaque V1. J. Neurophysiol., in press.

Livingstone, M.S., and Tsao, D.Y. (1999). Receptive fields of disparity-selective neurons in macaque striate cortex. Nat. Neurosci. *2*, 825–832.

Livingstone, M.S., Pack, C.C., and Born, R.T. (2001). Two-dimensional substructure of MT receptive fields. Neuron *30*, 781–793.

Mardia, K.V. (1975). Statistics of directional data. J. R. Stat. Soc. B *34*, 102–113.

Marr, D. (1982). Vision (New York: W.H. Freeman).

Martinez, L.M., and Alonso, J.M. (2001). Construction of complex receptive fields in cat primary visual cortex. Neuron *32*, 515–525.

Maunsell, J.H., and Van Essen, D.C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. J. Neurophysiol. *49*, 1148–1167.

Mikami, A., Newsome, W.T., and Wurtz, R.H. (1986). Motion selectiv-

ity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. J. Neurophysiol. *55*, 1328–1339.

Morgan, M.J., and Castet, E. (1995). Stereoscopic depth perception at high velocities. Nature *378*, 380–383.

Morgan, M.J., and Castet, E. (1997). The aperture problem in stereopsis. Vision Res. *37*, 2737–2744.

Movshon, J.A., and Newsome, W.T. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. J. Neurosci. *16*, 7733–7741.

Movshon, J.A., Thompson, I.D., and Tolhurst, D.J. (1978). Receptive field organization of complex cells in the cat's striate cortex. J. Physiol. *283*, 79–99.

Movshon, J.A., Adelson, E.H., Gizzi, M.S., and Newsome, W.T. (1985). The analysis of moving visual patterns. C. Chagas, R. Gattas, and C.G. Gross, eds. (New York: Springer), pp. 117–151.

O'Keefe, L.P., and Movshon, J.A. (1998). Processing of first- and second-order motion signals by neurons in area MT of the macaque monkey. Vis. Neurosci. *15*, 305–317.

Pack, C.C., and Born, R.T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. Nature *409*, 1040–1042.

Pack, C.C., Berezovskii, V.K., and Born, R.T. (2001). Dynamic properties of MT neurons in alert and anaesthetized macaque monkeys. Nature *414*, 905–908.

Perez, R., Gonzalez, F., Justo, M.S., and Ulibarrena, C. (1999). Interocular temporal delay sensitivity in the visual cortex of the awake monkey. Eur. J. Neurosci. *11*, 2593–2595.

Prince, S.J.D., Pointon, A.D., Cumming, B.G., and Parker, A.J. (2002). Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. J. Neurophysiol. *87*, 191–208.

Pulfrich, C. (1922). Die Stereoskopie im Dienste der isochromem und herterochromen Photmetrie. Naturwissenschaft *10*, 553–564.

Qian, N., and Andersen, R.A. (1997). A physiological model for motion-stereo integration and a unified explanation of Pulfrich-like phenomena. Vision Res. *37*, 1683–1698.

Roy, J.P., Komatsu, H., and Wurtz, R.H. (1992). Disparity sensitivity of neurons in monkey extrastriate area MST. J. Neurosci. *12*, 2478–2492.

Smith, E.L., Chino, Y.M., Ni, J., Ridder, W.H., and Crawford, M.L.J. (1997). Binocular spatial phase tuning characteristics of neurons in the macaque striate cortex. J. Neurophysiol. *78*, 351–365.

Stoner, G.R., and Albright, T.D. (1998). Luminance contrast affects motion coherency in plaid patterns by acting as a depth-from-occlusion cue. Vision Res. *38*, 387–401.

Szulborski, R.G., and Palmer, L.A. (1990). The two-dimensional spatial structure of nonlinear subunits in the receptive fields of complex cells. Vision Res. *30*, 249–254.

Van Ee, R., and Anderson, B.L. (2001). Motion direction, speed and orientation in binocular matching. Nature *410*, 690–694.

Von Grunau, M., Dube, S., and Kwas, M. (1997). The effect of disparity on motion coherence. Spat. Vis. *7*, 227–241.