Neural Mechanisms of Form and Motion Processing in the Primate Visual System

David C. Van Essen and Jack L. Gallant
Department of Anatomy and Neurobiology
Washington University School of Medicine
St. Louis, Missouri 63110

Our visual system uses the information contained in the two-dimensional images on the retina to generate a rich set of perceptions about the characteristics of our external three-dimensional world. These perceptions arise in a split second and require little conscious effort, yet they are the result of sophisticated information processing strategies that entail many stages of analysis. A major aim of visual neuroscience is to explain our subjective perceptions in terms of the properties of single neurons at these different processing stages. Much of our current understanding of higher visual processing comes from studies of the macaque monkey, whose visual system is in many respects similar to that of humans. The macaque visual system includes dozens of distinct areas within the cerebral cortex, plus numerous subcortical centers. The pattern of anatomical connections among these components suggests two basic organizational principles. First, the visual system is hierarchically organized but with extensive two-way information flow between levels. Second, each hierarchical stage contains multiple subdivisions, which are embodied as separate areas, separate anatomical compartments within an area, or (at subcortical levels) different morphological cell classes. Subdivisions at different levels are connected in specific patterns, giving rise to distinct processing streams which can be traced through many hierarchical stages.

This anatomical organization raises two broad questions about functional specialization within the visual system. First, what are the functions of the different processing streams? Second, how do the response characteristics of individual neurons differ at successive stages of the anatomical hierarchy? These questions have been tackled most successfully in relation to two key aspects of vision: the perception of form (the shapes of objects) and of motion (where things are moving). These explorations mainly involve physiological and behavioral techniques that are quite different from the cellular and molecular techniques most familiar to this journal's readership. Nonetheless, we hope that a review of recent progress in understanding visual cortex will interest a broad spectrum of neuroscientists who share the ultimate objective of attaining a continuum of explanations of brain function, from the most molecular to the most cognitive levels.

We will begin with a brief overview of the anatomical organization of the visual system. The remainder of the review will discuss the functional distinctions among the different processing streams, starting in the retina but emphasizing the early and intermediate stages of cortical processing.

The Visual Cortex Is a Distributed Hierarchical System

The general layout of the macaque visual system can best be visualized on a two-dimensional unfolded representation of its different subcortical and cortical centers (Figure 1). The major subcortical centers (lower left) include the two retinae, the lateral geniculate nucleus, the superior colliculus, and the pulvinar complex. In the cortex, 32 distinct areas associated with vision have been identified, which collectively occupy more than half of the total surface area of the neocortex in the macaque (Fellman and Van Essen, 1991; see also Desimone and Ungerleider, 1989). Areas specifically mentioned in this review are indicated in different colors in Figure 1; the remaining visual areas are unlabeled and shown in gray.

At the earliest stage of cortical processing are the two largest areas, V1 and V2, each of which occupies about 10% of the neocortex. At intermediate stages, there are more than a dozen visual areas, but only four of them will be discussed here: area V4, the middle temporal area (MT), and the dorsal and lateral subdivisions of the medial superior temporal complex (MSTd and MSTl, respectively). The highest stages of processing include two distinct clusters of areas—one located in the inferotemporal (IT) cortex and the other in the posterior parietal (PP) cortex.

Many visual areas contain internal compartments, or modules, that have been revealed by a combination of anatomical and physiological techniques. Figure 1 illustrates the spatial pattern and the relative sizes of these modules in five different areas. In V1, modularity is most clearly manifested by a fine grained pattern of "blobs" and "interblobs" that can be identified histochemically and by their pattern of connections (Livingstone and Hubel, 1984). The modules that have been identified in higher areas (V2, V4, MT, and posterior IT; see below) are larger in size and differ in their topology. For example, in V2 they form stripes rather than the spotted pattern found in V1.

The 32 visual areas of the macaque are interconnected by more than 300 distinct cortico-cortical pathways. The great majority of these pathways are reciprocal; for example, V1 projects to V2, and V2 projects back to V1. In most cases, there are pronounced asymmetries in the specific cortical layers in which reciprocal pathways originate and terminate. On this basis, the entire collection of visual areas can be arranged in a hierarchy that contains ten distinct levels of cortical processing (Fellman and Van Essen, 1991). Figure 2 illustrates a simplified hierarchical scheme that includes only the subset of visual areas shown in color in Figure 1, but that also includes the various compartments...
The flattened cortical map encompasses the entire right hemisphere. Areas shown in color are discussed in the text; remaining visual areas are shown in gray. The cortical map also shows schematic representations of the various compartmental subdivisions that have been identified within individual visual areas, including the blobs, interblobs, and layer 4b of V1 (blue, green, and red, respectively); the thick stripes, thin stripes, and interstripes of V2 (red, blue, and green, respectively); and the recently described subdivisions of V4, MT, and posterior IT. The spatial configuration of compartments is illustrated schematically but approximately to scale. Partitioning within and between areas is based on information and references provided in Felleman and Van Essen, 1991, Van Essen et al., 1992, Van Essen and DeYoe, 1994, and DeYoe et al., submitted. Modified, with permission, from Felleman and Van Essen, 1991, and Van Essen et al., 1992.

within these areas which contribute to the different processing streams.

Three major processing streams have been identified at the subcortical level. In the lateral geniculate nucleus, these streams are represented by distinct populations of cells residing in different layers. The most prominent of these are the parvocellular (P) and magnocellular (M) layers (shown in Figure 2 as gray and pink boxes, respectively). In addition, there is a small population of koniocellular (K) neurons (yellow box) whose function remains poorly understood (Casagrande and Norton, 1991; Hendry and Yoshioka, 1994). As they converge in V1, the three subcortical streams are reorganized into three new streams, which can be traced through several subsequent stages of the extrastriate cortex (see Van Essen and DeYoe, 1994). The blob-dominated (BD) and interblob-dominated (ID) streams (green and blue boxes, respectively) are named after the compartments in V1 from which they originate. The magnocellular (MD) stream (red boxes) is named after its dominant source of subcortical inputs. At higher stages of the visual hierarchy, there is another reorganization into separate dorsal and ventral streams (Ungerleider and Mishkin, 1982; Desimone and Ungerleider, 1989). The dorsal stream is largely an extension of the MD stream, and it projects to the PP cortex (orange box), which is involved in the analysis of spatial relations (where things are). The ventral stream includes the BD and ID streams, and it projects to the IT complex (turquoise boxes), which is involved in pattern recognition (what things are). The outputs of the visual cortex are distributed to various other cortical areas in the temporal and frontal lobes and to numerous subcortical structures, which are all outside the scope of this review.

Overall, these anatomical findings constitute a progress report that represents a major revision of our understanding of the organization of the primate vi-
Figure 2. Hierarchical Organization of Concurrent Processing Streams in the Macaque Monkey

Boxes represent visual areas, compartments within an area, and subcortical centers; solid lines represent major connections between structures (usually reciprocal pathways); and icons represent characteristic neurophysiological properties. Subcortical streams in the retina and lateral geniculate nucleus (LGN) include the M, K, and P streams (grey, yellow, and pink, respectively). Cortical streams at early and intermediate stages include the MD, BD, and ID streams (red, green, and blue, respectively). The PP complex is shown in orange. The IT complex includes posterior inferotemporal areas (PIT), which are components of the BD and ID streams, and central and anterior areas (CIT and AIT). Pathways are based on information cited in Felleman and Van Essen, 1991, and Van Essen and DeYoe, 1994. Physiological specializations are based on information cited in Felleman and Van Essen, 1987, DeYoe and Van Essen, 1988, and in the present text.
visual cortex. Until the 1970s, it was widely presumed that the visual cortex contained only three subdivisions (areas 17, 18, and 19), which were linked by a strictly serial (unidirectional) flow of information, from area 17 to 18 to 19. Starting with the work of a few early pioneers (see Allman and Kaas, 1974; Zeki, 1975), this classic picture has evolved into our current scheme, which includes dozens of visual areas, many of which have an internal modular structure and a principle of distributed hierarchical organization that includes bidirectional information flow and multiple processing streams.

Neural Function Can Be Analyzed by Lesions and Single-Neuron Recordings

The hierarchical scheme illustrated in Figure 2 provides a basic structural framework for addressing questions about the specific functions associated with different streams and different stages of processing. The various icons within the figure summarize much of what is known about the distribution of different physiological and functional characteristics in the macaque visual system. This information derives mainly from two complementary approaches: lesion studies and single-unit recording.

Lesion studies assess the behavioral deficits resulting from ablation or inactivation of a specific cortical or subcortical structure. This approach can suggest how the information processed within the lesioned structure actually contributes to different perceptual and visuomotor tasks. However, it has inherent limitations, one of which is that simply identifying a deficit says little about the detailed neural processing carried out within the lesioned region.

Single-unit neurophysiological recordings are used to determine the specific stimulus characteristics that are effective in eliciting responses from neurons in any given region. Typically, the first step is to determine the region of the visual field from which a cell can be directly activated (the classical receptive field). Various stimuli are then presented within this receptive field, and the firing rate of the cell is determined in response to each stimulus. There are an infinite number of possible visual stimuli, but usually no more than a few hundred can be presented to any given cell during the limited period it is available for study. Choosing stimuli appropriate for investigation of any particular issue is a major challenge, and the choice often depends on the investigator's assumptions about the way information is encoded by individual neurons. Some tend to regard visual neurons as feature detectors that convey information about the likelihood that a particular trigger feature is present in the image (Barlow, 1972; Tanaka, 1993). From this perspective, a key objective is to determine what particular stimulus is most effective in driving any given cell. Others tend to regard visual neurons as filters, which are typically tuned along more than one stimulus dimension (orientation, wavelength, etc.). This perspective puts greater emphasis on determining which stimulus dimensions are most relevant to each cell and on characterizing the tuning curves along each of these dimensions. The underlying assumption is that differences in firing rate convey information useful for discriminating among stimuli that lie on the slopes of each cell's multidimensional tuning surface. Our emphasis will be on the filtering point of view, which has been very useful in computational vision (see Marr, 1982) and is becoming more widely accepted in neurophysiology.

Low Level Analysis Occurs in the Subcortical Streams

The different subcortical processing streams carry out a basic division of labor that has been revealed by neurophysiological recordings and by behavioral tests after selective lesions (see Merigan and Maunsell, 1993; Van Essen et al., 1992). This division is best understood for the P and M streams, which differ in the information they carry along several low level stimulus dimensions.

P cells are the most numerous cell type, and they provide a higher resolution (finer grained) spatial representation of the visual world than the less numerous M cells. In the time domain, P cells convey information mainly about relatively slow changes (low temporal frequencies, corresponding to patterns that are static or moving slowly), whereas M cells emphasize relatively rapid changes (high temporal frequencies, corresponding to rapid motion or sudden transients). In the color domain, P cells convey most of the information about the colors of different objects, and M cells play only a minor role.

Given that the P and M streams are selective for distinct yet partially overlapping aspects of the information contained in images, it should not be surprising that each stream contributes to many different aspects of perception (Schiller et al., 1990; Merigan and Maunsell, 1993). For example, if P cells are selectively lesioned, a monkey cannot discriminate colors, but it can discriminate different shapes, textures, depths, and motion patterns, as long as the stimuli contain the appropriate spatial and temporal frequencies for activating cells in the surviving M stream. Likewise, if M cells are selectively lesioned, a monkey can discriminate different shapes, textures, depths, and even motion patterns, as long as the stimuli contain the appropriate spatial and temporal frequencies for activating cells in the surviving P stream.

Motion Analysis Is Emphasized in the MD Stream

The MD stream (Figure 2, red boxes) includes portions of V1 and V2 plus several higher extrastriate areas (MT, MSTd, and MSTf). At all stages of the MD stream, there is a relatively high incidence of cells selective for the direction of stimulus motion (arrow icon). The highest incidence of direction selectivity occurs in MT, whose involvement in motion analysis has been...
established more directly by lesion and microstimulation experiments. In particular, lesions of MT lead to specific deficits in motion perception and also in the smooth eye movements used to track a moving target (Newsome et al., 1985; Newsome and Paré, 1988). Electrical stimulation of a small region of MT (by passing current through a microelectrode) can markedly perturb a monkey's perception of the direction of stimulus motion, causing the monkey to report, for example, leftward movement for a stimulus that is actually moving to the right (Salzman et al., 1992; Salzman and Newsome, 1994).

When tested with simple stimuli (such as moving spots, bars, or gratings), most cells in MT are tuned relatively broadly for the direction of stimulus motion and also for stimulus speed. On the presumption that MT is likely to be involved in higher order motion analysis as well, a number of investigators have studied the responses of MT neurons to more complex motion patterns. One such example involves plaid patterns, which consist of two superimposed gratings that differ in orientation and direction of movement. Under some conditions, human observers perceive the two gratings as a pair of transparent gratings moving in different directions. Under other conditions, the percept is of a single coherent plaid pattern moving in a direction different from either of the component gratings. Some cells in MT have responses that correlate with the subjective reports of human observers, suggesting that they play a role in the representation of complex motion (plaid icon in Figure 2; Movshon et al., 1985; Stoner and Albright, 1993).

Our visual system uses motion information for a variety of purposes. One useful distinction is between the local motion signals used to analyze the trajectory of an individual object and the global motion flow fields that arise when we move our eyes or navigate through the environment. Neurophysiological studies suggest that these two types of motion analysis are segregated into distinct subregions within MT (red and orange in Figure 1), whose outputs are then distributed to separate visual areas, MSTd and MSTf, at the next stage of the hierarchy (Born and Tootell, 1992; Komatsu and Wurtz, 1988a, 1988b).

The motion flow fields that arise during navigation typically include components of image expansion (when moving forward), contraction (when moving backward), rotation (when tilting the head), and shear (when moving past objects at different distances). The characteristics of cells in MSTd suggest a specific involvement in the analysis of these types of flow fields (Saito et al., 1986; Duffy and Wurtz, 1991a; Orban et al., 1992). This is illustrated in Figure 3 for two different
neurons (Tanaka and Saito, 1989). The neuron on the top responded vigorously to an expanding motion field but not to contraction or ordinary linear motion. The neuron on the bottom responded best to clockwise rotation but not to counterclockwise rotation or linear motion. We refer to cells of this general type as non-Cartesian motion cells (rotation icon in Figure 2) because they are tuned along dimensions that are described most simply in polar or hyperbolic coordinates rather than conventional Cartesian (x,y) coordinates. However, motion analysis in MSTd involves more than just extracting the components of motion along the cardinal non-Cartesian axes (pure rotation, pure expansion, etc.). In particular, some MSTd neurons are selective for spiral motion patterns, which have components of both rotation and expansion (Graziano et al., 1994; see also Duffy and Wurtz, 1991a, 1991b). Thus, it appears that MSTd contains a population of motion filters whose peaks are scattered throughout a multidimensional stimulus space. This strategy for representing higher order motion information is presumably useful for mediating the variety of behavioral tasks (avoiding obstacles, guiding eye and limb movements, etc.) that are carried out during navigation through the environment.

The major outputs of the MD stream are directed dorsally to areas in the parietal lobe. Lesion studies and physiological recordings indicate that the PP cortex is involved in a variety of high level functions. These include analyzing spatial relations (where things are located), controlling eye movements, and determining where visual attention is to be directed (see Mountcastle et al., 1975; Ungerleider and Mishkin, 1982; Andersen, 1989). Although many details remain to be worked out, the various areas within the PP complex appear to be specialized for handling distinct (but partially overlapping) subsets of this constellation of visuomotor and attentional tasks (see Andersen, 1989; Colby et al., 1993).

Form and Color Analysis Are Emphasized in the Ventral Stream

The ventral stream encompasses the BD and ID streams (green and blue in Figure 2), which parallel one another over four stages of the cortical hierarchy, from V1 to the posterior IT complex (DeYoe et al., submitted). Originally, both streams were thought to be dominated by inputs from the subcortical P stream. More recent evidence indicates a much greater degree of convergence, with the ID stream receiving inputs from two subcortical streams and the BD stream from all three subcortical streams (Figure 2; see Nealey and Maunsell, 1994; Van Essen and DeYoe, 1994).

Physiological recordings and lesion studies imply that the ventral stream in the analysis of both form and color, but it has proven difficult to determine the specific role of the individual streams and individual areas in these different aspects of perception. There is some physiological evidence for a functional dichotomy in which the BD stream mediates color perception, and the ID stream mediates form perception (Livingstone and Hubel, 1984, 1988). However, there are difficulties with this hypothesis because there is significant intermixing of receptive field characteristics in different streams (see DeYoe and Van Essen, 1988; Peterhans and von der Heydt, 1991) and because blobs in V1 (and other structures in the BD stream) are present in the owl monkey, a primate species that has no color vision (Tootell et al., 1985). An alternative hypothesis is that the BD stream analyzes the surface characteristics of objects (including their texture and brightness as well as their color), whereas the ID stream analyzes the shapes of objects, based on their boundaries rather than their interiors (see Grossberg, 1987; Allman and Zucker, 1990; Van Essen and DeYoe, 1994).

At present, little is known about functional differences between the BD and ID streams at higher stages of the hierarchy (V4 and beyond) that would help to distinguish between these or other hypotheses. It has been suggested that V4 as a whole is primarily involved in color vision, both in the macaque (Zeki, 1973, 1983) and in a putative homologous area in humans (Zeki et al., 1991). However, lesions of V4 in the macaque have much greater effects on pattern discrimination than on color discrimination tasks (Schiller et al., 1990; Schiller and Lee, 1991; Heywood et al., 1992; Merigan, 1993). This suggests that V4 is more important for form vision than for color vision.

Physiological recordings from V4 have provided additional evidence in support of its involvement in form vision (Desimone and Schein, 1987). To obtain a more systematic analysis of how form information is represented in V4, we have used a novel class of visual stimuli that are more complex than conventional bars and gratings (Gallant et al., 1993; Gallant et al., 1993, Soc. Neurosci., abstract). These include concentric, radial, spiral, and hyperbolic gratings, which we collectively refer to as non-Cartesian gratings. Although the stimuli we used were stationary rather than moving, they are mathematically analogous to the non-Cartesian motion stimuli successfully used in the studies of MSTd mentioned above. Concentric gratings are analogous to rotation, radial gratings to expansion, etc.). As illustrated in Figure 4, we found many cells in V4 that responded better to certain non-Cartesian gratings than any conventional sine wave (Cartesian) grating. The stimuli in this figure are grouped into three subsets. Cartesian gratings (left) vary in orientation and spatial frequency; hyperbolic gratings (center) also vary in orientation and spatial frequency; and polar gratings (concentric, radial, and spiral; right) vary in concentric and radial frequency. Each stimulus is color coded according to the mean response that it elicited from the cell. Blue corresponds to stimuli that were relatively ineffective, and red to stimuli that were maximally effective (spiral patterns for the cell in Figure 4A and hyperbolic patterns for the cell in Figure 4B). Cells preferring polar gratings were more common than cells preferring hyper-
Figure 4. Responses of Two Cells in V4 to Different Cartesian and Non-Cartesian Spatial Patterns

Each icon represents a particular visual stimulus, and its color represents the mean response to that stimulus, relative to the spontaneous background rate, using the color scale shown above.

(A) A neuron that responded maximally to polar stimuli, particularly spirals, and much less to hyperbolic and Cartesian gratings. Modified, with permission, from Gallant et al., 1993.

(B) A neuron that responded maximally to hyperbolic stimuli of low to moderate spatial frequencies (J. G. and D. V. E., unpublished data).

Bolic gratings, and within this subpopulation, concentric preferences were more common than radial preferences. Many cells responded to a wide range of stimuli within both the Cartesian and non-Cartesian stimulus spaces. We suspect that these cells would also respond to a variety of other patterns containing appropriate combinations of curvature cues (e.g., curved edges, intersections, and other patterns that are not pure gratings). We consider the most significant aspect of non-Cartesian form cells to be their orderly tuning across a set of higher order stimulus dimensions rather than their best stimulus per se. An
interesting possibility is that V4 might include cells tuned along additional higher order dimensions distinct from the particular non-Cartesian dimensions illustrated here. Testing this hypothesis will require the development of new classes of visual stimuli that capture the complex aspects of shape information contained in natural images.

The IT complex (Figure 2) includes posterior, central, and anterior subdivisions that have been collectively implicated in pattern recognition. Lesions of the IT cortex impair the ability to recognize previously learned patterns (Ungerleider and Mishkin, 1982). Many cells in the IT cortex respond best to complex shapes, including various geometrical stimuli (Fujita et al., 1992; Tanaka, 1993) and natural shapes such as faces or hands (Perrett et al., 1982; Desimone et al., 1984). These cells generally respond to a range of stimuli (e.g., several different faces or hands when tested with a large number of stimuli), suggesting that they are tuned along dimensions associated with certain classes of natural stimuli but not for unique examples within that class.

It is unlikely that the complex receptive field characteristics encountered in higher visual areas are genetically hardwired. Instead, these properties are probably related to visual experience so that the specific pattern of higher order receptive fields present in any particular brain depends upon the individual's particular sensory history. Indeed, recent experiments have shown that the selectivity of cells in the anterior IT cortex can be markedly altered by repeated exposure to a limited set of behaviorally relevant stimuli (Miyashita, 1993; Sakai and Miyashita, 1993).

Concluding Remarks

Functional Specialization Reflects the Diversity of Visual Cues and Visual Tasks

To illustrate the nature of functional specialization in the visual cortex, we have emphasized the roles of the MD stream in motion analysis and the ID and BD streams in form and color analysis. However, each stream contains a diversity of response characteristics suggestive of involvement in a range of visual tasks (Peterhans and von der Heydt, 1993; Sáry et al., 1993). These characteristics make sense, given that high level visual functions generally require access to a variety of low level cues, and low level cues can contribute to a variety of perceptual and visuomotor tasks. Viewed from this perspective, the extensive convergence, divergence, and cross talk between processing streams provide an anatomical substrate that enables the visual system to perform a wide variety of tasks in a complex visual environment (DeYoe and Van Essen, 1988; Van Essen et al., 1992).

There Are Parallels between Form and Motion Analysis

We have noted a parallel between the non-Cartesian spatial dimensions linked to form analysis in V4 and the non-Cartesian spatiotemporal dimensions linked to motion analysis in MSTd. The specific perceptual functions of non-Cartesian selectivity in V4 and MSTd are surely quite different, and there are many differences in the specific receptive field properties of each area. Yet, the neural computations underlying higher level form and motion analysis may have important features in common, as has been suggested for the neural circuitry underlying conventional orientation and direction selectivity in V1 (Douglas and Martin, 1991). If such commonalities indeed exist, it may help to explain how the cerebral neocortex, which is characterized by a remarkable degree of structural uniformity, is able to carry out such a wide diversity of functionally distinct tasks.

Visual Neurons Are Multidimensional Filters

At every stage of the visual hierarchy, neurons act as filters that are tuned along several distinct dimensions. At early cortical levels, the dimensions, such as orientation and spatial frequency, tend to be relatively simple. At higher levels, inputs are combined in ways that lead to selectivity along more complex dimensions, such as the non-Cartesian dimensions relevant for V4 and MSTd. In addition, as one moves up the visual hierarchy, responses become progressively less dependent on exactly where stimuli are placed in the visual field and progressively more dependent on attention and other aspects of behavioral state (see Maunsell and Newsome, 1997; Miyashita, 1993; Connor et al., 1993; Soc. Neurosci., abstract; Motter, 1994). Treating neurons as filters rather than feature detectors has inherent analytic advantages even when dealing with these highly nonlinear aspects of neural selectivity.

Biological Systems Are Complex but Decipherable

The organization and function of the primate visual system is far more complex than most neuroscientists appreciated as recently as a decade ago. Similar statements can be made for many important topics in cellular, molecular, systems, and developmental neuroscience. Processes that were once discussed in relatively simple mechanistic terms have been shown to involve complex interacting components, whether they be transmitters, channels, and trophic factors (in studies focusing on the cellular/molecular level), or cell classes, receptive field types, and cortical areas (in studies focusing on the systems level). Whatever the level of analysis, it remains a great challenge to understand how these diverse components interact dynamically with one another to produce coordinated, reliable performance in a living organism.

Acknowledgments

We thank Drs. J. Senes, J. Lichtman, D. Gottlieb, and C. T. Connor for insightful comments on the manuscript. Work from the authors' laboratory was supported by grants from NIH (EY02091) and ONR (N00014-91-J-1192).
References


