Chapter 3
Inferior Temporal Cortex
as a Pattern Recognition Device*

Charles G. Gross†  Hillary R. Rodman†  Paul M. Gochin†
Michael W. Colombo†

1 Introduction

The sense of sight discerns the differences of shapes, wherever they are ... without delay or interruption, employing careful calculations with almost incredible skill, yet acting unnoticed because of its speed.

Ptolemy, Optics

How does the pattern of energy falling upon the retina yield the complexity of our visual experience and the richness of our visual memory? How do we recognize flowers and faces and all the other infinite inhabitants of our visual world? How do we continue to recognize them over transformations of lighting, viewpoint and size that totally replace the original mosaic of retinal excitation? Forty years ago, the textbook answer was clear: striate cortex (area 17), and only striate cortex, was responsible for all the mysteries of our perception of form and color, for our consciousness of the visual world. As a prominent introductory physiological psychology textbook put it, “There is no evidence that any area of the cortex other than the visual area 17 is important in the primary capacity to see patterns... whenever the question has been tested in animals the story has been the same” [67]. This view persisted, at least in some quarters, well into the period of research we will describe here. As recently as 1975, Krieg’s monumental treatise [52] on primate brain anatomy asserted that “image formation and recognition is all in area 17 and is entirely intrinsic...the connections of area 17 are minimal.”

Today, we know that a much larger expanse of the primate cortex, indeed about half of the entire cortical mantle, is involved in visual function [19]. Perhaps it should not be surprising that so much of the primate cortex is

---

*This work was supported by NIMH grant MH-19420, NSF grant BNS-8919538, and McDonnell-Pew grants 90-16 and 90-42.
†Department of Psychology, Princeton University, Princeton, NJ 08544.
concerned with vision since so much of the behavior of primates in general is visual in nature, and so much of the conscious experience of humans is likewise visual in nature.

2 Multiplicity of Visual Areas

As the recognized limits of primate visual cortex have crept forward over the past several decades to include more and more of the regions earlier assumed to serve “association” functions, we have come to realize that visual cortex is made up of a large number of distinguishable domains or “areas.” Some of these subdivisions are shown in Figure 1, which illustrates the cortical surface of a macaque monkey brain. All of the labelled areas are exclusively or primarily visual in function. A variety of converging criteria are used for distinguishing different visual areas.

Architectonic and other histological features. Cortical subregions, nonvisual as well as visual, were first distinguished at the turn of the century on the basis of differences in cell packing and cell size (cytoarchitectonics) and in appearance of patterns of myelinated fiber bundles (myeloarchitectonics). More modern histological techniques have allowed us to parcel out visual areas on the basis of their different concentrations of intrinsic substances such as the metabolic enzymes cytochrome oxidase and acetylcholinesterase, cell surface proteins such as CAT-301, and neurotransmitter substances or their precursors (e.g., [94,18,84]).

Retinotopic organization. Most primate visual cortical areas contain a systematic representation of some or all of the contralateral visual field. The arrangement and precision of the representation, the size of the receptive fields, and the degree to which ipsilateral as well as contralateral visual space is represented varies from area to area. Some of the more anterior zones, such as inferior temporal cortex (IT) and the superior temporal polysensory area (STP), lack retinotopic organization and are accordingly distinguished on the basis of other criteria.

Other neuronal response properties. In addition to receptive field size and organization, individual visual cortical areas differ in the degree to which cells within them are sensitive to stimulus attributes such as direction or speed of motion, size, color, shape, spatial frequency content, and so forth. Visual areas are also differentiable, in awake monkeys, by the degree to which their neuronal activity is modulated by such variables as attention, novelty, eye position, and stimulus significance.

Selective effects of removal. Damage to different visual areas leads to different patterns of deficits on tasks which probe the animal’s ability to use information about various stimulus attributes, to control eye movement responses, and to construct memories about a stimulus.

Patterns of connectivity. Finally, individual visual areas differ in their patterns of connections with other visual cortical areas, with subcortical structures, and with nonvisual cortical regions. Patterns of
Fig. 1. Lateral and ventral views of a macaque monkey brain illustrating some of the known subdivisions of visual cortex. In each of the smaller brain drawings, the shaded areas correspond to sulci which have been "opened up" to show the areas located within them. FEF, frontal eye field; FST, fundus of the superior temporal sulcus visual area; IT, inferior temporal cortex; LIP, lateral intraparietal area; MST, medial superior temporal area; PO, parieto-occipital area; STP, superior temporal polysensory area; TF and TH, cytoarchitectonically defined areas of the parahippocampal cortex; 36, cytoarchitectonic area 36 of perirhinal cortex; 7a, cytoarchitectonic area 7a of parietal cortex; TEO, visual area TEO; VIP, ventral intraparietal area; V1, V2, V3, V4, first, second, third, fourth visual areas; V4t, transitional V4 area.
interconnections between cortical regions vary, moreover, in their degree of convergence and divergence, in their "modularity" or patchiness, and, as described below, in their laminar features.

![Diagram of feedforward and feedback interconnections]

**Fig. 2.** Illustration of laminar patterns of interconnections used to define hierarchies of processing stages within visual cortex [55]. Feedforward projections originate mainly from cells (triangles) located in layers II/III of one visual area and terminate within layer IV of the target area. Feedback projections originate from cells located outside layer IV, particularly the deep layers, and terminate outside layer IV.

3 Hierarchical Organization of Visual Cortex

About a decade ago, neuroanatomists began to realize that corticocortical projection patterns could be used to specify the direction of information transfer between different functional areas and that the ensuing relationships could be used to construct hierarchies of cortical processing stations [77,55]. Although various intermediate patterns can be identified, the vast majority of interconnections between visual cortical regions are of one of the two general types illustrated in Figure 2. *Feedforward* projections originate mainly in supragranular layers II/III and terminate predominantly in layer IV, whereas *feedback* projections originate mainly in infragranular layers and terminate outside of layer IV. This principle has been used extensively to generate flow diagrams, such as that shown in Figure 3, of information processing within primate visual cortex. Even so, visual cortical regions are massively interconnected; a recent review tabulated 305 separate interconnecting pathways between a total of 32 differentiable visual or mostly visual cortical regions [19]. This figure represents 31% of the total number of pathways possible if each area were connected to all others.
FIG. 3. Diagram of some of the pertinent connections of IT cortex with other cortical areas and medial temporal lobe structures. The heavy line indicates the main afferent pathway to IT cortex, as discussed in the text. A, amygdala; ER, entorhinal cortex; H, hippocampus; PH, parahippocampal cortex; PR, perirhinal cortex. Other abbreviations are as for Figure 1. Faces indicate areas in which neurons selectively responsive to face stimuli have been found [37].

4 Processing Streams in Visual Cortex

In parallel with the discovery of this bewildering number of visual cortical areas has come the widespread realization that the primate visual cortex as a whole can be divided into a small number of processing subsystems. Most generally, primate visual cortex can be thought of as comprising two major processing streams, namely 1) a dorsally located set of areas subserving the analysis of movement and of space and 2) a ventrally located set of areas subserving the identification of visual objects and the storage of information about their form and color [97]. Although this dichotomy is undoubtedly an oversimplification, it remains a useful and valid expositional principle.

A summary of the major features of information flow within the ventral stream is shown in Figure 3. The ventral visual pathway begins in certain laminar subdivisions of V1 on the occipital pole, continues in subdivisions of area V2, and progresses forward and ventrally via synaptic stages in areas V4 and TEO before reaching the inferior temporal cortex, the central subject of this chapter.
5 Inferior Temporal Cortex

Inferior temporal (IT) cortex of the monkey is the final purely visual cortical processing stage within the ventral pathway [12,97]. After lesions of IT cortex, defined here as corresponding to architectonic area TE, monkeys have a severe and permanent deficit both in recognizing visual stimuli they have seen before and in forming new memories of visual stimuli [36,61]. Yet, their visuo-sensory capacities are intact. That is, they have normal visual acuity, visual fields of normal extent, and normal thresholds for a variety of psychophysical functions. Furthermore, their memory function in modalities other than visual is normal [36]. In 19th century terms, monkeys with IT lesions would be said to have a visual agnosia [25]. In more modern terms, IT damage in monkeys produces a deficit in processing and storing information about the physical attributes of visual stimuli, especially form and color. In humans, the temporal cortex, particularly its inferior and ventromedial portions, is similarly involved in high-level processing and storage of visual information. Temporal cortex of the right hemisphere in humans appears to play an especially important role in complex visual processing [60,10].

The focus of this chapter is on the properties of IT neurons and how these may account for the crucial role of IT cortex in visual pattern processing. We begin by briefly summarizing the characteristics of the successive stages of the ventral system that provide input to inferior temporal cortex. Then we consider inferior temporal cortex itself, particularly how its neurons may represent visual stimuli, their role in visual attention, how their activity is modulated by experience, how this area may be organized and how it may interact with medial temporal and frontal lobe structures to store information.

6 From V1 to TEO

The different areas in the ventral visual cortical pathway share a number of characteristics. Consistent with its role in form vision and object identification, all areas in the ventral pathway contain some cells sensitive to stimulus form, color or both. In addition, all areas in the pathway beyond V1 appear to depend entirely on the information they receive from V1 for their ability to respond to visual stimuli, whereas many areas in the dorsal pathway do not [79,80,31,38]. However, as one moves along the ventral pathway from V1 through V2, V4 and TEO to IT, the properties of the neurons encountered, and their connectivity with other areas, change systematically in several ways. These changes are crucial to understanding how properties in IT cortex are generated.
7 RF Size and Topography

Receptive field (RF) size as a function of eccentricity within the visual field is shown in Figure 4 for the various areas within the ventral visual pathway. Moving from V1 to IT, the characteristics of receptive fields change in several ways. First, RFs become larger at any given eccentricity, and the increase in RF size with eccentricity becomes steeper. Second, as one moves along the ventral pathway, the visual topography in successive areas becomes progressively less fine, so that in V2 it is less precise than in V1, less pronounced still in V4, and coarse in TEO. Third, the amount of ipsilateral visual field representation increases from a degree or two in V1 and V2, slightly more in V4, and first becomes marked (10 deg. or more) in TEO [28,29,30,22,7].

![Figure 4](image)

**Fig. 4.** Regression lines relating multiunit receptive field size (square root of receptive field area) to receptive field eccentricity in various areas along the ventral visual pathway. The two lines for V2 show data separately for two different animals.

8 Stimulus Selectivity and Other Neuronal Properties

Moving along the ventral pathway from V1 to TEO, we find two major trends in the stimulus selectivity of single neurons in these areas. The first is a trend toward selectivity for increasingly complex features or conjunctions of features, and the second is an increase in the extent to which these features are signalled regardless of spatial location, i.e., in the degree of tolerance of retinal position of the stimulus.

Cells in V2, like those in V1, respond in a tuned fashion to the length, width, orientation, and color of a bar of light, or to the spatial frequency of a grating pattern [3,24,8]. However, some cells in V2 (but
not V1) respond to appropriately oriented "illusory" contours generated by stimulation of surrounding regions [99]. Moreover, some V2 cells are conjointly selective for the color and dimensions of small stimuli within RFs that are considerably larger than those at comparable eccentricities in V1 [3]. These properties suggest that some V2 cells provide a more generalized or abstract representation of stimulus features than cells at the previous stage. This trend is continued in V4 in several ways. First, many V4 cells are tuned for both length and width of a stimulus in an invariant manner over a considerably wider RF than seen at the earlier stages. Second, some V4 cells seem particularly selective for sharp edges, a property not seen at earlier stages, over their larger RFs [13]. Finally, many cells in V4 have large suppressive surrounds bordering their RFs, so that they respond maximally to stimuli that stand out from the background on the basis of differences of form; in the color domain, the complementary color selectivities of the central RF and suppressive surround have been proposed as a mechanism for color constancy [86]. Finally, extraretinal influences such as directed attention and the use of visual imagery are first apparent in V4 and, as described later, become more pronounced in IT [17,46].

Little is known at present about the receptive field properties of neurons in area TEO except that cells in this region have stimulus requirements intermediate in complexity between those of V4 cells and those of IT neurons [22,93].

9 Trends in Anatomical Patterns

Trends in connectivity. In parallel with the increasing coarseness of retinotopic organization at successive stages in the ventral pathway, the patterns of corticocortical connectivity patterns become less and less topographic or "point-to-point." Thus, at the far end of the ventral stream, the inputs to IT from V4 and TEO, as well as the feedback projections from IT to these areas, have no discernible retinotopic organization [7,21,23]. Likewise, connections with subcortical structures, both visual and nonvisual, become more and more diffuse at progressive levels of the ventral hierarchy [23]. Finally, callosal connections linking areas in the ventral stream with their counterparts in the opposite hemisphere become denser and more widespread at higher levels, consistent with the increasing extent of ipsilateral visual field representation. In fact, callosal terminations in V1 and V2 are limited to the vertical meridian representation at the border between the two areas, whereas in V4, TEO and IT they are much more extensive [98]. In V4, callosal connections appear important in generating the large suppressive surrounds which may play a role in figure-ground segregation and color constancy [16].

Other physical gradients. Opiate receptor concentration increases in a graded fashion at successive levels of the ventral visual pathway, possibly
related to an increasing influence of affect-based selective attentional processes [54]. Progressively higher levels of the pathway also show increasing levels of phosphorylation of several proteins which have been implicated in synaptic plasticity and hypothesized to play a role in long-term information storage [68].

10 Characteristics of Inferior Temporal Cortex

Overall, the biggest discontinuity in neuronal properties and function along the ventral pathway occurs as we progress forward into IT.

Modal specificity. As is the case for areas in the earlier stages in the ventral pathway, neurons in IT cortex respond only to visual stimuli; likewise, ablation of this tissue alters only visually guided behavior [45,36]. Cortical areas hierarchically beyond IT such as perirhinal and parahippocampal cortex and STP receive direct input from other sensory modalities as well as the visual, and thus may be termed “polysensory” or multimodal areas [12,50,88]. Under certain conditions, however, cells in IT cortex itself appear to be activated by nonvisual signals, namely when an auditory stimulus signals the immediate arrival of a visual stimulus important for the animal [49,40]. It is possible that the auditory stimulus elicits a visual image and that it is the visual image which Activates the IT neurons. A similar phenomenon has been reported for tactile stimulation for cells in V4 [46], but not earlier in the ventral stream.

Receptive field size and location. At the level of IT cortex, the visual receptive fields are no longer organized in a topographic manner [44,12]. Rather, virtually all IT RFs include the center of gaze, so that information about stimulus location has been thrown away to concentrate on processing information from the center of the visual field, usually the center of the animal’s attention. At least in anesthetized monkeys, the size of RFs in IT varies from several square degrees to virtually the entire visual field with a median size about 25 x 25 degrees, much larger than those of cells in areas earlier in the ventral visual pathway. The majority of IT RFs are bilateral, extending well into both visual half-fields. Thus, unlike in earlier stages of the visual system, the representations of the two halves of visual space are “united” by individual cells. This unity of space is provided by the forebrain commissures, namely the anterior commissure and the posterior portion of the corpus callosum; these fiber pathways provide converging input from IT, TEO and V4 in one hemisphere to IT cells in the other [12,78]. Section of both of these commissures eliminates the ipsilateral portion of IT RFs [76,42]. The larger size and bilateral extent of single IT RFs provide greater opportunity for signalling a stimulus independent of its retinal position than in earlier areas in the ventral stream. Indeed, stimulus selectivity remains similar throughout at least the large central portion of the RF, providing a basis for the perceptual phenomenon of stimulus equivalence across retinal
Fig. 5. Receptive fields of single neurons recorded on a series of penetrations through IT cortex. Receptive fields are represented by best-fitting rectangles; field borders extending beyond the 60 x 60 tangent screen are represented by dashed lines; and the meridians are represented by vertical and horizontal lines. Receptive fields smaller than 8 are shown as dots. The inset shows the location of the section on a side view of the brain and the portion of the section enlarged in the main figure. STS, superior temporal sulcus [12].

Some examples of IT receptive fields are shown in Figure 5.

11 The Representation of Stimuli in IT Cortex

Elephants and ensembles: general features of stimulus selectivity in IT cortex. Both the frequency and temporal pattern of the spike discharges of IT cells can be altered by a number of stimulus variables including shape (most commonly), color, orientation, texture, and even direction of motion [44,14]. Thus, as for other areas in the ventral pathway, IT can be said to "code" aspects of shape, color, and so forth. However, whereas for a small proportion of IT cells the optimal stimulus appears to be
FIG. 6. Average responses of each of 41 IT neurons to five stimuli presented 20 times in pseudo-random order. Neuronal firing rate is shown on the y-axis and stimulus number on the x-axis. The actual stimuli are illustrated at the bottom right. The face stimulus was colored, the other stimuli, black and white. The horizontal lines are the mean spontaneous activity for 400 ms of the ITI, and +/- 1 SD confidence limits. The curves are the mean activity and +/- 1 SD confidence limits for a 400 ms period during presentation of the sample stimuli [39].
a line segment of a particular width, length, color and orientation, for most it appears to be a more complicated stimulus, often a three-dimensional one with a fair amount of internal detail. It should be pointed out that the term \textit{optimal stimulus} is not used to imply that we have conclusively determined the stimulus configuration that will maximally activate the cell, but merely to denote the stimulus among those tested which elicited the greatest increase or decrease in firing. As we noted some time ago, in searching for the optimal stimulus we often feel like Kipling’s proverbial blind men examining an elephant.

As a whole, IT cells are very heterogenous in their stimulus selectivity. Some will fire only or very much more to a particular stimulus or class of stimuli whereas others will fire about equally well to a great variety of stimuli. Still others show what we sometimes refer to as \textit{paradoxical selectivity}, namely that they fire best or only to a small number of often highly complex stimuli which do not share obvious common physical characteristics [41]. Figure 6 illustrates response profiles for a random sample of 41 IT neurons presented with each of five standard stimuli. Note that while a few cells (e.g., No. 30) show a strong response to only one stimulus, most cells respond well to two or more stimuli and many to all five (Nos. 13, 41). However, they tend to respond to different stimuli with significantly different firing rates. The findings that many IT cells respond to a large variety of stimuli with differing response magnitudes and that a given stimulus will activate a number of IT cells have led to the view that visual stimuli are represented in IT cortex by the pattern of firing across a population or ensemble of cells [37,14]. To study population coding in IT, Gochin et al. [33] have developed a method for estimating the accuracy with which a sample of IT cells can discriminate among members of a particular set of visual stimuli. Some results obtained with this method, which bears some similarities to multi-variate analysis, are illustrated in Figure 7. Figure 7a shows the ability of an ensemble to discriminate among stimuli as a function of the number of stimuli; Figure 7b shows ensemble performance as a function of the number of cells in the ensemble; and Figure 7c shows that different stimuli are differentially discriminable by a set of IT cells.

\textbf{Selectivity for faces.} The most striking class of highly specific IT cells are those that fire preferentially to faces, and indeed these are the IT cells that have been most extensively studied [11,14,71,82,100]. An example of selectivity for face stimuli is shown in Figure 8. Are face cells an exception to the view that IT cortex represents the visual world through the pattern of firing across a neuronal ensemble? Are face selective cells truly face “detectors”? There is no convincing evidence for any such “grandmother cells,” that is, for IT cells narrowly selective for a specific face across transformations of size, orientation, and color. Even the most selective face cells discharge to a variety of individual faces and usually also discharge, though to a lesser degree, to other stimuli as well. Thus, the
Fig. 7. Factors which affect the capacity of an ensemble of IT cells to discriminate a set of stimuli during performance of a visual delayed matching from sample task. A. Ensemble performance (percent correct stimulus identifications) for five different stimuli. Based on all neurons recorded in each of two monkeys. B. Ensemble performance in distinguishing five stimuli as a function of ensemble size. The first point on each curve is chance and the last point is the total population sampled in the study. C. Ensemble performance as a function of the number of stimuli. Based on 10 neurons in each of two monkeys [33].

representation of faces by IT cortex probably involves the pattern of firing across a population of cells with varying selectivity, just as for other visual objects. However, there may be important quantitative differences in the way IT deals with faces. First, the large proportion of IT cells that give statistically significant responses to face stimuli suggest that more IT cells participate in encoding faces than other types of stimuli. Second, the high specificity shown by at least some “face-selective” cells suggests that they are relatively dedicated to the task of face processing, although not, we would argue, in the sense of a detector. Finally, we speculate that many of the
Fig. 8. Responses obtained from a face-selective IT cell in an alert infant monkey to the set of stimuli shown beneath the graph. The cell gave a significant response (asterisk) only to one of the profiles shown. Very few spikes were obtained with other profiles, scrambled profiles, front views, and control stimuli such as lines or FDs. (Adapted from [81].)

ensembles participating in face coding are themselves relatively dedicated, in that these ensembles may not participate in coding many other patterns. The more explicit representation of faces in IT cortex is probably not the result of a different type of neural circuitry devoted to them, but rather the result of the general properties of a network architecture capable of encoding both general categories and subtle differences within categories [33].

There are several reasons why the visual systems of primates might be expected to possess relatively dedicated circuitry for representing facial configurations. For one, it is more crucial for a monkey to be able to differentiate among individual faces than among individual exemplars of other classes of stimuli. For another, faces are more similar to each other in their overall organization and fine detail than most other stimuli a monkey must discriminate among, such as different types of fruit. That faces appear to be represented somewhat more explicitly than other stimuli reflects their importance for primates. Faces are crucial in primate social behavior, they are perceived extra-ordinarily early in life, they have special perceptual properties, and our memory for them is prodigious [101].

Shape constancy. One of the central mysteries of perception is that we recognize a particular shape across changes in its retinal location, size, contrast, color or even spatial frequency spectrum. The basis of these perceptual phenomena might be the neurons in IT cortex whose selectivity for specific shapes remains invariant over such transformations [87,83]. That is, they show shape constancy. Data from some such cells are shown in
Figure 9.

**Fig. 9.** Responses of three IT neurons to "FD stimuli." The shape selectivity remains similar over changes in stimulus size (A), contrast (B), and retinal location (C). Responses are plotted as percent change of each neuron's mean spontaneous rate. Each point is a mean of ten responses. f, fovea; c, contralateral field; i, ipsilateral; u, upper; l, lower [37].

An alphabet for shape? How do IT cells extract information about overall shape from the information they receive about local contours from areas V4 and TEO? One possibility is that they do so by acting as detectors or filters for Fourier Descriptors. The method of Fourier Descriptors is a method for describing boundary curvature that is sometimes used in computer pattern recognition systems [103]. This method depends first on determining the boundary orientation function for the shape and then expanding it in a Fourier series. Each term in the Fourier expansion is associated with a particular frequency, amplitude and phase and is known as a Fourier Descriptor. Any closed boundary or shape is fully described by its set of Fourier Descriptors. The inverse transform of a single Fourier Descriptor uniquely determines a plane closed boundary having a specific number of lobes (frequency), lobe indentation (amplitude) and phase (orientation). (We call these shapes, some of which are shown in Figure 9, "FD stimuli"). This method of describing shapes is independent of both the position and size of the stimulus. Thus, Fourier Descriptors are an efficient method for representing shapes. If IT cells were selective for a particular FD stimulus, i.e., for a closed boundary described by a specific Fourier Descriptor, then a population of such neurons could specify or code any shape. This hypothesis yields several specific and testable predictions. First, at least some IT cells should be selective for specific FD stimuli. Second, this selectivity should cover an adequate range of parameters for
coding an arbitrary shape. Third, this selectivity should remain invariant over variations in stimulus size, retinal location and contrast. All three predictions were confirmed in a study from which Figure 9 is drawn [87].

A fourth and crucial test of the notion that IT cells use the method of FDs as an “alphabet” for shape is that the response of an IT cell to a complex shape should be predictable from the cell’s responses to FD stimuli. To carry out this test we first determined the response of a sample of IT neurons to a set of FD stimuli, varying only frequency. Then for the cells tuned to the frequency of the FD stimuli, we tested their responses to stimuli constructed by linear combinations of two FD stimuli. We also tested their responses to an FD stimulus halved and quartered, manipulations that produce stimuli with borders resembling part of the border of the original FD but with very different Fourier Descriptor components. As shown in Figure 10, our results were clearly incompatible with the hypothesis that single IT units act as linear filters for boundary curvature. More generally, the response to the individual FD stimuli could not be related to the response to their compounds, that is, we failed to find evidence for any linear or systematic non-linear coding scheme involving Fourier Descriptors. In summary, IT cortex does not appear to code shape by a method analogous to the method of Fourier Descriptors [1]. Perhaps there is some other alphabet or basis set that IT does use. In any case, FD stimuli do provide a useful set of systematically varying shapes for studying the properties of IT cells.

12 Intrinsic Circuitry

Shared and direct connections. As an initial step in studying the intrinsic circuitry of IT cortex, we measured the functional interaction of IT cells by computing cross-correlograms of the spike discharges of simultaneously recorded IT cells [34]. If two cells receive synchronous or nearly synchronous excitatory input from a common source, the cross-correlogram of their activity would be expected to form a peak spanning zero delay, usually a relatively wide one [64,70]. By contrast, if one cell provides excitatory input to another, their cross-correlogram would be expected to be very narrow and offset to one side of zero. Figure 11b shows an example of a correlogram from IT cells indicating shared input and Figure 11a shows a correlogram indicating direct connections. Cross-correlation analysis of 288 pairs of IT cells (in which the two members of each pair were within about 0.5 mm of each other) indicated that 19% had shared inputs and 16% had direct interconnections. Interestingly, both figures indicate less common excitatory connectivity than has been reported for V1 cells using the same analysis in the cat [95,96,56]. In a simple model network, effective connectivity increases through successive layers. Since this does not appear to have occurred along the ventral stream on the way to IT cortex, it suggests that some inhibitory mechanism has been incorporated [34].

Inhibitory connections. Neurons with similar stimulus selectivities
Fig. 10. A. Selectivity of an IT neuron for FD frequency. Response in spikes per second above baseline rate on the y-axis and frequency in cycles per perimeter on the x-axis. B. Actual responses (bars) and predicted responses (asterisks) of the same neuron for compound FDs. C. Actual and predicted responses to halved FDs [1].

were not more likely to show either shared or direct connections than were neurons with different selectivities. This is consistent with our suggestion that IT cortex utilizes “competitive learning” circuits involving inhibitory as well as excitatory competitive interactions. Further support for the importance of inhibitory interactions in IT comes from several studies which show that the responses to two simultaneous stimuli were usually less than to one presented alone, indicating that the second stimulus suppressed the response to the first [85,59] (see Figure 12).

Synaptic efficacy. The complex stimulus selectivity of IT neurons would appear to be the result of integration of a large number of input connections from preceding areas where neurons have simpler selectivity and smaller receptive fields. Since IT appears to be a site of visual information storage, it seems likely that the efficacy of connections between IT cells would be quite variable and depend on experience. In an effort to address the issues of number and typical efficacy of connections, we have estimated these parameters based on spike train statistics [34]. As described in the previous
section, when two neurons are directly connected, their cross-correlogram shows a narrow peak such as that illustrated in Figure 11a. The area of the peak serves as a measure of the number of times the post-synaptic cell discharged as a consequence of the firing of the presynaptic neuron. The efficacy of the connection can be estimated by dividing the peak area by the total number of spikes from the presynaptic neuron. The number of connections responsible for activation of the post-synaptic neuron can be estimated by dividing the total number of spikes from the post-synaptic neuron by the peak area. The average number of estimated inputs to IT neurons is about 40, and the distribution is wide and skewed (Figure 13).
FIG. 12. Histogram of suppression indices for 34 IT neurons. The suppression index was the ratio of the difference between the responses to the pair and the strongest individual stimulus versus the best individual stimulus, times 100. Negative values indicate that the cell responded more weakly to a pair of simultaneously presented stimuli than to a single stimulus. Most IT neurons (30/34) had a negative suppression index [59].

The distribution of values for synaptic efficacy is similarly broad and skewed. These measures are of particular interest when compared with the results of similar analyses applied to V1. Reports on V1 [95,96,56] indicate that synaptic efficacy is typically high and distributed over a narrow range of values. Moreover, very few inputs (2-10) are estimated to be necessary to activate typical V1 neurons. These observations are consistent with a change from a tightly wired, static network in V1 to a more variably wired, dynamic network in IT.

13 The Effects of Extra-Retinal Factors on IT Neuron Activity

The activity of IT cells can be affected by a number of factors besides the pattern of light falling on the retina. These include anesthesia, age, the attentional demands on the animal, and the animal's previous experience.

Anesthesia and age. The basic "sensory" properties of IT cells described in previous sections appear essentially the same in awake behaving animals and in immobilized animals anesthetized with nitrous oxide and oxygen [75,40,34]. For example, the proportion of cells that are visually responsive, the effects of different visual stimuli, and response invariance across site, size and color remain similar in the two states. There are two exceptions to this generalization. First, although there have been
no direct systematic comparisons on the same cells, it appears that the absolute magnitude of IT responses (measured by firing rate) is higher in awake monkeys than in monkeys under nitrous oxide anesthesia. The second exception concerns young animals [81]. In unanesthetized monkeys as young as five weeks old, the properties of IT cells are very similar to those in adult monkeys. However, under nitrous oxide anesthesia, the incidence of responsive cells in the infants dropped from about 80% (in adults and in unanesthetized infants) to 50% in animals between the ages of four and seven months and to about 10% in animals between five weeks and four months. In both the anesthetized infant groups, the response properties of the diminished number of responsive cells were similar to those in the adults and unanesthetized infants. By contrast, there is no effect of nitrous oxide anesthesia on the activity of cells in V1 or in area MT in either infants or adults. Thus, there seems to be some special sensitivity of IT neurons in infant monkeys to nitrous oxide. Finally, barbiturate anesthesia, even in adults, virtually eliminates all visual responsiveness in IT cortex, but not V1 [45].

Effects of attention. As soon as IT neuron activity was investigated in awake behaving animals, it became clear that the strength or even presence of responses to stimuli could be modulated by their behavioral context. For example, stimuli used as discriminanda in a learning task usually elicited stronger responses than the same stimuli when they had no significance for the animal [40,74,90]. Moreover, the mere presence or absence of a visual fixation point can alter the response of an IT cell to a patterned stimulus within its receptive field and even alter the size of that receptive field [74,75].
In a delayed match to sample task the magnitude and temporal pattern of the response of a cell to a particular stimulus can depend on whether it is presented as sample or match and if match whether the sample on the trial was the same stimulus or a different one [40,57].

![Graph A: 2 sec ISI](image1)

![Graph B: 20 sec ISI](image2)

**Fig. 14.** Responses of an IT neuron to a stimulus presented once every 2 sec (A) and once every 20 sec (B). On the left are responses on individual trials and post-stimulus time histograms. On the right, the neuron’s response on each trial is plotted, after smoothing with a 3 point moving average. The mean baseline firing rate is plotted on the far right (triangle) [58].

Desimone et al. [17,65] showed that when two different stimuli were presented within the receptive field of an IT cell, only the stimulus that the animal attended to drove the cell. The authors pointed out the double benefits of this attention mechanism. First, the gating of responses to unattended stimuli reduces the amount of irrelevant information or noise that must be processed. Second, the authors suggested that “the fact that the neurons responded to an attended stimulus as if their receptive fields had contracted around it may allow cells to communicate information with high spatial resolution in spite of their large receptive fields.” There is evidence that inputs from the pulvinar contribute to these attentional modulations of IT activity including alterations of effective receptive field size [39,15,53,72].

**Memory.** Although removal of IT cortex produces a severe and permanent deficit in both short and long term visual memory [36,48,91],
very little is known about how these functions are subserved by IT cells. We do have, however, a number of examples of how the responses of IT cells can be modified by experience, usually transiently, but in one case relatively permanently [62].

Perhaps the simplest example of the modification of IT neuronal responses with experience is the reduction of neuronal response with repeated presentation of the same stimulus, a process analogous to behavioral habituation [58]. As shown in Figure 14, when the stimulus interval is very short, the response usually declines to or near the level of spontaneous activity in as little as nine presentations. After a brief period of no stimulation the magnitude of the response returns to its original level (Figure 15). By contrast, with a substantially longer interstimulus interval there is usually little or no habituation. Although this phenomenon occurs in both unanesthetized and anesthetized animals, the time constant of the habituation is different under the two conditions. The reduction of response to a repeated stimulus is a neural correlate of the familiarity of a stimulus, but how the visual system uses this information is not clear.

In a delayed matching to sample behavioral task, some IT neurons continue to fire into the delay between sample and match and may in fact continue to do so throughout the entire delay interval. Fuster and Jervey [26], who discovered this phenomenon, argued that the delay activity could be a “neuronal memory trace,” a correlate of short term memory. Support for this view comes from the findings that the delay activity can be specific to one of the sample stimuli and remains invariant over changes in size, orientation, color and position [63], just as the response to the sample stimulus itself is tolerant to such manipulations for many IT neurons [87].

Is the information in the delay period actually used, that is, does it have any functional significance? Recently, we have found evidence that it

![Graph showing mean response of 30 neurons to two runs of presentations of the same stimulus. For each run, the inter-stimulus interval was 2 sec and there were 14 min between the end of the first run and the start of the second. From [58].]
Fig. 16. Percent IT cells showing increased activity in the delay period of visual delayed matching from sample as a function of the animal's performance level on this task. Numbers above each bar indicate the number of cells studied during blocks in which the monkey's performance level fell within the specified range.

may: the incidence of activity in the delay period was higher on blocks of trials in which the monkey's performance was higher [9] (Figure 16). It will be valuable to further evaluate the possibility that this IT “delay activity” actually corresponds to a visual memory trace by studying the effect of manipulating performance level (varying the length of delay, introducing distractions, and increasing the number of intervening stimuli) on the incidence and magnitude of delay activity and by studying the effect of manipulating delay activity (by electrical stimulation) on performance level. At present, however, the role of the delay period activity of IT neurons in the visual memory process is unclear.

Functional architecture of inferior temporal cortex. Little is known about the functional architecture of IT cortex except that it can be loosely partitioned on the basis of various types of criteria. Anatomically, IT can be subdivided on the basis of cytoarchitectonics, and different regions vary in the strength and distribution of their subcortical and cortical connections [88,89,102,19,23,66]. Furthermore, neuronal response properties vary in different parts of IT [12,93]. Receptive field size is much larger in the anterior and anterior-dorsal part of IT, and the incidence of face selective cells is markedly higher in the anterior portion of the lower bank of the superior temporal sulcus than elsewhere in IT [6]. Functionally, the most anterior and ventral region, that is, the portion bordering on perirhinal cortex (areas 35 and 36) appears to be particularly important in short and long term memory [62,48]. Virtually nothing is known about the laminar or vertical organization of IT. As in other portions of cortex, nonvisual as well as visual, cells with similar response properties do show a marked
tendency to cluster [93,44,34] and it has been suggested that these clusters may be organized more vertically than horizontally [47] (K. Tanaka, personal communication).

14 Beyond IT Cortex

Relations with the hippocampal system. Evidence from ablation and cooling experiments indicates that IT cortex is necessary for both short term and long term storage of visual information [36,48,91]. Indeed, at least in the case of long term memory, IT cortex would appear to be the site of such storage. However, in order for long term storage to occur the hippocampus and adjacent cortical regions must be involved. As shown in Figure 3, IT cortex (as well as other high order sensory cortices) project to perirhinal and parahippocampal cortex which in turn project to the hippocampus both directly and by way of connections with entorhinal cortex. These connections are largely reciprocal and it is believed that through these connections the hippocampus and inferior temporal cortex interact to form the basis of the storage and perhaps the retrieval of visual declarative memories [92].

Interactions with the frontal lobe. The dorsal and ventral visual pathways described earlier appear to retain a large measure of segregation in their ultimate projections to the frontal lobe. Inferior temporal cortex projects to lateral orbital and ventral regions, whereas “high-order” areas in the dorsal pathway such as MST and areas within the intraparietal sulcus project to dorsolateral and medial zones [5]. Consistent with the proposed functions of the dorsal and ventral visual cortical pathways, damage to dorsolateral prefrontal cortex produces impairments on spatial memory tasks (e.g., [35,73,69]), whereas lesions of more ventral zones result in impairments on visual object recognition and pattern discrimination tasks [2,51]. Interestingly, analysis of the laminar pattern of interconnections between IT cortex and inferior frontal cortex suggest that the projection from IT to frontal cortex may not be easily described as either feedforward or feedback [89,4,78]. This observation is consistent with physiological studies indicating that the interaction between these two zones is relatively complex [27]. Interactions between the hippocampal system and IT may be more important for long term storage of visual information, whereas interactions between IT and ventral frontal cortex may be more important for short term or “working memory” of visual information.

15 Concluding Note

In primates, inferior temporal cortex is the final and crucial stage for the processing of visual information about form and color. Furthermore, it also appears necessary for the short and long term storage of visual information and is probably the actual storage site. Thus an understanding of the neu-
ral circuitry that underlies the properties of IT neurons and the behavioral effects of its removal is required for any biological understanding of pattern recognition. Since humans easily surpass any existing computer as pattern recognition devices, better pattern recognition devices are likely to come from a better understanding of IT cortex.

16 Acknowledgments

We would like to thank Nina Rebmann and Maida Rosengarten for their help in preparing the manuscript.

References


[38] C. G. Gross, Contribution of striate cortex and the superior colliculus to visual function in area MT, the superior temporal polysensory area and the inferior temporal cortex, Neuropsychologia, 29 (1991), pp. 497–515.


