International Symposium on

Advanced Computing for Life-Science

As a part of
International Symposia on Information Sciences
(ISKIT'92)

July 12–15, 1992

Kyushu Institute of Technology
Iizuka, Fukuoka, Japan
Detection of Object Features in Primate Visual Cortices

Ichiro Fujita
Precursory Research for Embryonic Science and Technology Research Development Corporation of Japan,
and
Laboratory for Neural Information Processing Frontier Research Program, RIKEN Institute,
Wako, Saitama 351-01,
Japan

Abstract

The anterior part of the inferotemporal cortex (IT) of the monkey is situated at the final level in the hierarchical, distributed network of the visual cortical areas. Neurons in this area respond selectively to a particular complex object-feature such as a shape or a combination of shape with color or texture. The anterior IT as a whole show a wide variety of stimulus selectivities. Spatial distribution of these various stimulus selectivities within the anterior IT indicates that this cortex is organized in a columnar fashion: neurons responsive to similar object-features cluster together and are aligned vertically to the cortical surface. It is argued that this organization is suitable for representing visual images of objects with complex and diverse features.

1. Object Vision

We can distinguish and recognize an almost infinite number of objects in the natural world. Two ideas have been conceived to explain how this is done by the brain. One idea is that the brain contains neurons, each of which responds to each recognizable object. This idea is referred to as "recognition cell hypothesis" or "grandmother cell hypothesis" (because we might suppose that a single neuron or a small number of neurons in the brain may respond only to our own grandmother). The other idea is that any one of neuron in the brain does not encode a particular object, but distribution pattern of activity across a population of neurons constitutes or determines our visual perception. We refer to this idea as "population-coding hypothesis". Which of these ideas explain better what actually happens in the brain? This is not a settled problem in neuroscience.

When we recognize an object, our brain matches the visual image of that object with what we have in internal representation in the brain. For this matching process as well as for construction of the internal representation through learning processes, the brain must detect and analyze the visual features of objects. The question of how the brain does this job is directly related to the above two hypotheses. In this paper I shall review our own neurophysiological-anatomical studies on the inferotemporal cortex (IT) of the Japanese
monkey (Fujita et al., 1990, 1991a,b; Saleem et al., in press; Tanaka et al., 1991b). We have chosen the IT for the starting point of our study on the neural mechanism of object vision, because this cortex represents the final stage of the visual cortical stream crucial for object vision in the monkey, and, therefore, the functional architecture of the IT should reflect the final form of representation of object features in the brain.

2. The Visual Cortical Pathway in the Macaque

The anatomy of the brain visual pathway has been most extensively analyzed in the macaque monkeys (such as Japanese monkeys and rhesus monkeys). Cortical areas processing visual information occupy the posterior part of the brain. This cortical area is now divided into 32 subregions (Felleman and Van Essen, 1991), and more is going to be identified (e.g., Boussaoud et al., 1991). Figure 1 shows several of these areas labeled on the lateral surface of a monkey brain.

![Visual cortical areas](image)

**Figure 1** Visual cortical areas crucially involved in object vision. The areas are labeled on the lateral surface of the left hemisphere of a Japanese monkey brain. V1, V2, V4; primary and secondary visual cortices, visual area 4. PIT, AIT; posterior and anterior parts of the inferotemporal cortex.

Visual pathway in the primate cortex starts at the primary visual cortex, V1, which then projects to the second visual area, V2. From there, the visual pathway divides into two major routes, one directs dorsally towards the parietal cortex, and the other ventrally to the inferotemporal cortex. Several lines of evidence suggest that the dorsal stream deals with the visual location of objects and the ventral stream is involved in the identification of objects (Mishkin et al, 1983).

3. Neuronal Selectivity of Inferotemporal Cortex Neurons

How is shape information extracted along the ventral pathway from V1 to the IT? In early 1960's, Hubel and Wiesel (1962) found that a class of neurons in V1 respond to a slilt of a particular orientation ("orientation selective cells"). Later in 1980's an amazing class of neurons were found in the IT and its adjacent cortex; they respond selectively to faces of monkeys or experimenters, but not to other objects (see Desimone, 1991, for a recent review). Those cells are called "face-selective cells". As one might immediately notice, there is a big gap in the complexity of stimulus features to activate orientation-selective cells and face-selective cells, i.e., oriented bars and faces. What kinds of neurons exist beside the two classes of neurons?

Tanaka et al. (1991a) recently analyzed selectivities of IT neurons to object-features with the following method. While activity of a single IT neuron was monitored, they first presented to the monkey several tens of various 3D objects and 2D paper cutouts for the
initial survey for stimuli to activate the neuron. Those stimuli were shown with various viewing angles and orientations. When they found an effective object or paper cutout, then the stimulus feature essential for the activation was analyzed by simplifying the stimulus step by step (see Figure 2 of Tanaka et al., 1991b for an example). This process was originally performed by cutting and pasting paper models, and recently a computer-graphics system was introduced to this procedure (Fujita et al., 1991a).

From analysis of more than 700 neurons over a wide area of the temporal visual cortex, Tanaka et al. (1991a) demonstrated that the IT consists of two physiologically different subareas. Most neurons in the posterior 1/3 of the IT have relatively small receptive fields (3-5°) and respond well to simple stimuli such as slits and spots. On the other hand, the majority of neurons in the anterior 2/3 of the IT have large receptive fields (10-30°) and respond only to a particular, complex stimulus. Figure 2 shows such critical features determined for 11 anterior IT neurons. Those stimulus features are far more complex than simple slits and spots, although they are not complex enough to specify a particular object. Tanaka et al. (1991a) refer to this degree of complexity as "moderately complex". The exceptions to this are neurons selective to biologically significant stimuli such as faces and hands. The stimulus selectivity of the face-selective and hand-selective cells appear to be complex and sharp enough to signal the presence of faces or hands.

![Figure 2](image)

**Figure 2** Eleven representative examples of stimulus configurations critical for activating IT neurons. The features are more complex than simple dots, spots or edges, and not complex enough to correspond to a particular object.

4. **Modularity in the Inferotemporal Cortex**

4.1 **Relations between Adjacent neurons**

We examined the spatial arrangement of various stimulus selectivities within the anterior IT. As a first step we asked whether adjacent IT neurons have similar or identical stimulus selectivity or they respond to totally different visual features (Fujita et al., 1990). To answer this question, we recorded extracellular action potentials from more than 2 anterior IT cells with a single electrode (0.8-3.5 MOhm at 1 kHz). The signal pick-up radius of electrodes with similar impedance has been estimated be 50µm in the monkey cerebral cortex. The stimulus feature critical for activation was then determined for one of the neurons, while the second neuron simultaneously recorded was being separated with another window discriminator.

Figure 3 shows such an example where action potentials from one single neuron and other multiple neurons were simultaneously recorded, separated by window discriminators,
and counted to produce peristimulus time histograms. The threshold of the window for the single neuron was set higher than the amplitude of action potentials of the multiple units, so that the peristimulus histograms for the multiple neurons (lower row) did not count action potentials from the single neuron shown in the upper row. Both the single neuron and the multiple neurons responded better to horizontally striped squares and triangle than to horizontally striped circle or a square with vertical stripes. A white solid square or triangle or a dotted square did not evoke responses or even suppressed the cells. Thus these two units were similar to each other in that both of them required for their activation horizontal gratings as well as the proper shape of the contour of the stimulus. However, the best stimulus was not identical between them; the single neuron responded best to horizontally striped squares, whereas the multiple unit was activated most strongly by a triangle with horizontal gratings.

![Figure 3](image)

**Figure 3** An example of simultaneous recording from neighboring IT cells with a single electrode. Peristimulus histograms shown upper are from a single neuron in the anterior IT, and those shown lower are from the multiple unit simultaneously recorded. Both the single cell and the multiple unit responded well to squares or triangle filled with horizontal gratings.

At 84% of the 43 recording sites so far examined, the second unit responded to at least some of the optimal and suboptimal stimuli for the first cell, i.e., two nearby cells have an overlapping range of effective stimuli. In most cases, however, they were slightly different in their selectivity to fine parameters. The exact optimal stimulus was different in some cases as in Figure 3 or they have the same most effective stimuli but differently respond to suboptimal stimuli in other cases (data not shown). In 10% of the cases, the two neurons behaved very similarly and we could not detect a clear difference between them. In the remaining cases (7%), the second cell or multiple units did not respond to any of effective stimuli for the first cell. The results suggest that neurons with similar selectivity tend to cluster together within the anterior IT.

### 4.2 Columnar Organization of the IT

We next analyzed neurons obtained successively along penetrations directed either vertically or tangentially (obliquely) to the surface of the cortex in order to assess the spatial dimension of the clustering of IT cells with similar selectivities (Fujita et al., 1991a). Because time limitation during an experiment did not allow us to determine the critical feature for many neurons along a penetration, we took the following procedure. We first advanced an electrode to a depth in the IT, recorded from a neuron, and determined the stimulus feature critical for activation. Then we prepared a set of stimuli which included the optimal, suboptimal and ineffective stimuli for that cell. With this same set of stimuli, we tested other neurons sampled at 100 or 200 μm steps along that penetration.
In vertical penetrations, the distance over which we recorded neurons with similar selectivities occupied 64-86% (1.2 mm on average) of the estimated cortical thickness, indicating that neurons with similar stimulus preferences span most of the cortical layers. Along tangential or oblique penetrations, neurons with similar response properties were localized within 0.2-0.7 mm (0.4 mm on average). This distance is shorter than that obtained in vertical penetrations (p<0.001, t-test). The results suggest that the clustering is vertically elongated in the direction of the cortical depth and patchy across the surface.

4.3 Anatomical Correlates for the IT Columns

Injection of a neuronal tracer such as biocytin or wheat germ agglutinin-conjugated horseradish peroxidase (WGA-HRP) in the anterior IT results in a patchy distribution of labeled terminals and cell bodies within the anterior IT itself. Labeled terminals are found all through the grey matter from layer 1 to 6, while labeled cells are found in layer 2/3 and 5 and 6, and the cells in layer 2/3 are aligned with those in layers 5 and 6 (Fujita et al., 1991b). Similarly, injection of a lectin (PHA-L) into the posterior IT lead to columnar distribution of labeled fibers and terminals in the anterior IT (Saleem et al., in press). These results indicate that intrinsic and extrinsic connections of the anterior IT occur in a column-to-column fashion. This may partly be the anatomical basis for the functional columnar organization of the anterior IT.

5. Functional Implication of the IT Columns

We suggest from the evidence described above that the anterior IT consists of functional columns: neurons responding to similar object features are aligned vertically to the cortical surface. Columnar organization has so far been demonstrated for one-dimensional stimulus parameters such as orientation of stimulus gratings, ocular dominance of the visual inputs and the direction of visual stimulus motion. The present results demonstrate that similar organization is employed for representing shape which has much higher dimensions, and lend a support to the notion that columns are a basic organizational feature of the cerebral neocortex.

![Figure 4](image) Columnar organization of the anterior IT. Neurons responsive to similar or related stimulus features are grouped together to form columnar structure. The width of each column is around 0.4 mm. Most anterior IT neurons respond to partial features of objects, although there is a population of neurons selectively responsive to biologically important objects, faces.

When an object is shown to a monkey, a particular subset of columns, not the entire anterior IT, will be selectively activated by the image. Different partial object-features in the image activate different columns. Within each activated column, by contrast, the activity must be distributed across many cells because of their graded and overlapping
stimulus selectivities. The feature is therefore not encoded in the activity of a particular cell, but in the distribution pattern of activities over a population of cells in the column. It is suggested that the composition of the entire image is represented in the activation of a particular subset of columns (composition coding; Tanaka et al., 1991), whereas the individual partial features are encoded in the activities distributed over a population of cells in corresponding columns (population coding).

If we divide the surface area of the IT (from which we made recordings) by a square of columnar width, we get the number of columns, approximately 2,000. The number of distinct object features represented in this area would be smaller than this number, because we observed multiple clusters of neurons with similar selectivities along tangential penetrations. This limited number of distinct object-features may well constitute 'visual alphabets' or a 'basis set' of object-features to which the brain decompose visual image of objects.

References


