

INFEROTEMPORAL CORTEX AND OBJECT VISION

Keiji Tanaka

The Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa,
Wako-shi, Saitama, 351-01, Japan

KEY WORDS: macaque monkey, extrastriate visual cortex, object vision, optical imaging,
population coding

ABSTRACT

Cells in area TE of the inferotemporal cortex of the monkey brain selectively respond to various moderately complex object features, and those that cluster in a columnar region that runs perpendicular to the cortical surface respond to similar features. Although cells within a column respond to similar features, their selectivity is not necessarily identical. The data of optical imaging in TE have suggested that the borders between neighboring columns are not discrete; a continuous mapping of complex feature space within a larger region contains several partially overlapped columns. This continuous mapping may be used for various computations, such as production of the image of the object at different viewing angles, illumination conditions, and articulation poses.

Introduction

Recognizing objects by their visual images is a key function of the primate brain. This recognition is not a template matching between the input image and stored images but a flexible process in which considerable change in images—due to different illumination, viewing angle, and articulation of the object—can be tolerated. In addition, our visual system can deal with images of novel objects, based on previous visual experience of similar objects. Generalization may be an intrinsic property of the primate visual system. In this article, I discuss the neural organization essential for these flexible aspects of visual object recognition in the anterior part of the inferotemporal cortex.

The inferotemporal cortex (IT) of the monkey brain has been divided into subregions in several different manners. Our own division into posterior IT and anterior IT, based on the size of the receptive fields and the properties of responses (Tanaka et al 1991, Kobatake & Tanaka 1994), roughly corresponds to the previous cytoarchitectural division into TEO and TE (Iwai & Mishkin 1967; von Bonin & Bailey 1947, 1950): Posterior IT corresponds to TEO, and

anterior IT to TE. I use TEO and TE in this article because they are more popular.

TE receives visual information from the primary visual cortex (V1) through a serial pathway, which is called the ventral visual pathway (V1-V2-V4-TEO-TE). Although there are also jumping projections, such as that from V2 to TEO (Nakamura et al 1993) and that from V4 to the posterior part of TE (Saleem et al 1992), the step-by-step projections are more numerous. The IT projects to various brain sites outside the visual cortex, including the perihinal cortex (areas 35 and 36), the prefrontal cortex, the amygdala, and the striatum of the basal ganglia. The projections to these targets are more numerous from TE, especially from the anterior part of TE, than from the areas at earlier stages (Iwai & Yukie 1987, Ungerleider et al 1989, Saleem et al 1993a, Cheng et al 1993, Suzuki & Amaral 1994). Therefore, there is a sequential cortical pathway from V1 to TE, and outputs from the pathway originate mainly in TE.

Monkeys that have had their TE bilaterally ablated showed severe but selective deficits in learning tasks that required the visual recognition of objects (Gross 1973, Dean 1976). These behavioral results, together with the above-described important anatomical position of TE, suggest that TE is the site of neural organization essential for the flexible properties of visual object recognition.

In this review, our own data are emphasized, and the citation of other references is selective. This selection is not based only on the value of the studies but also on their relevance to the subject. The readers should read other reviews to get an overview of studies in the IT, e.g. Rolls (1991), Miyashita (1993), Gross (1994), and Desimone et al (1994). In particular, mechanisms of short-term memory of object images are not discussed in this article. I first summarize the data from unit-recording experiments to show that cells in TE respond to moderately complex object features and that those that cluster in a columnar region respond to similar features. I then consider the process by which the selectivity is formed in the afferent pathways to TE. I introduce the data of optical imaging of TE in order to discuss the function of the TE columns. Finally, I consider how the concept of the object emerges in the brain. The selections of our recordings that are introduced in this article were all conducted in anesthetized preparation, and they were from the lateral part of TE, lateral to the anterior middle temporal sulcus (AMTS). This part is often referred to as TED (dorsal part of TE).

Stimulus Selectivity of Cells in TE

One obstacle in the study of neuronal mechanisms of object vision has been the difficulty in determining the stimulus selectivity of individual cells. There

is a great variety of object features in the natural world, and we do not know how the brain scales down the dimension of this variety.

Single-unit recordings from TE were initiated by Gross and his colleagues (Gross et al 1969, 1972). They found that cells in TE had large receptive fields, most of which included the fovea, and that some cells responded specifically to a brush-like shape with many protrusions or to the silhouette of a hand. They extended the study of the stimulus selectivity by using two different methods: a constructive method and a reductive one. In the constructive method, they used Fourier descriptors that were defined by the number (frequency) and amplitude of periodic protrusions from a circle. Any contour shape can be reconstructed by linearly combining elementary Fourier descriptors of single frequency and amplitude. Some cells responded specifically to Fourier descriptors of a particular range of frequencies, with a considerable invariance for the overall size of the stimulus (Schwartz et al 1983). This method was not very promising, however, because the same group of authors found that the response of a TE cell to a composite contour was far from the linear combination of its responses to the elementary component contours (Albright & Gross 1990). Fourier descriptors are not the basis functions that the IT uses for the representation of object images.

The other direction that Gross's group pursued was reductive. They first presented many object stimuli for individual cells in order to find effective stimuli. Next, the images of the effective stimuli were simulated by paper cutouts to determine which features were critical for the activation (Desimone et al 1984).

We expanded this latter method and have developed a systematic reduction method with the aid of a specially designed image-processing computer system (Tanaka et al 1991, Fujita et al 1992, Kobatake & Tanaka 1994, Ito et al 1994, 1995). After spike activities from a single cell were isolated, many three-dimensional (3D) animal and plant imitations were presented to find the effective stimuli. Different aspects of the objects were presented with different orientations. Next, the images of the effective stimuli were recorded with a video camera and presented on a TV monitor by the computer to determine the most effective stimulus. Finally, to determine which feature or combination of features contained in the image was essential for the maximal activation, the image of the most effective object stimulus was simplified step-by-step while the activity of the cell was monitored. The minimal combination of features that evoked the maximal activation was determined as the critical feature for the cell. Figure 1 exemplifies the process for a cell for which the effective stimulus was reduced from the view of a water bottle to a combination of a vertical ellipse and a downward projection from the ellipse.

After the reduction was completed, the image was modified so that the selectivity could be further examined. Figure 2 exemplifies this latter process

