BEWARE: These are preliminary notes. In the future, they will become part of a textbook on Visual Object Recognition.

Chapter VII. First steps into inferior temporal cortex

Inferior temporal cortex (ITC) is the highest echelon within the visual stream concerned with processing visual shape information[1]. As such, one may expect that some of the key properties of visual perception may be encoded in the activity of ensembles of neurons in ITC. The story of how inferior temporal cortex became accepted and described as a visual area is a rather interesting one; we encourage readers to consult (Gross, 1994) for a lucid historical discussion.

7.1.   Preliminaries

Imagine that you are interested in finding out the functions and properties of a given brain area, say inferior temporal cortex (ITC) within the primate ventral visual stream. As we have discussed before (**Chapter 4**), part of the answer to this question may come from lesion studies. Bilateral lesions to ITC cause severe impairment in visual object recognition in macaque monkeys and several human object agnosias are correlated with damage in the inferior temporal cortex (**Chapter 4**). Another piece of evidence for function could come from non-invasive functional imaging studies. For example, upon presenting images of human faces and comparing the patterns of blood flow against those obtained when the same subject looks at pictures of houses, investigators typically report increased activity in the fusiform gyrus (e.g. (Kanwisher et al., 1997)). To some, this may be enough. To many others, this is only the beginning. Even if we have some indication (through lesion studies, functional imaging studies or other techniques) of the general function of a given brain area, much more work is needed to understand the mechanisms and computations involved in the function and properties of neurons in that area. We need to understand the receptive field structure and feature preferences of the different types of neurons in that area, how these preferences originate based on the input, recurrent connections and feedback signals and what type of output the area sends to its targets. For this purpose, it is necessary to examine function at neuronal resolution and millisecond temporal resolution. In this lecture, we will give an overview of the heroic efforts of many investigators to characterize the activity of neurons in ITC.

7.2.   Neuroanatomy of inferior temporal cortex

---

[1] The famous Felleman and Van Essen diagram from 1991 places the hippocampus at the top. While visual responses can be elicited in the hippocampus, it is not a purely visual area and it receives inputs from all other modalities as well.

Inferior temporal cortex (ITC) is the last purely visual stage of processing along the ventral visual stream. It consists of Brodmann's cytoarchitectonic areas 20 and 21. It is subdivided into areas TE and TEO or PIT/CIT/AIT (Felleman and Van Essen, 1991; Logothetis and Sheinberg, 1996; Tanaka, 1996). ITC receives feed-forward topographically organized inputs from areas V2, V3 and V4. It also receives (fewer) inputs from areas V3A and MT, highlighting the interconnections between the dorsal and ventral streams. ITC projects back to V2, V3 and V4. It also projects (outside the visual system) to the parahippocampal gyrus, pre-frontal cortex, amygdala and perirhinal cortex. There are interhemispheric connections between ITC in the right and left hemispheres through the corpus callosum (splenium and anterior commissure). ITC includes a large part of the macaque monkey temporal cortex. Anatomically it is often divided into multiple different subparts as defined above but the functional subdivision among these areas is still not clearly understood. Although there are multiple visually responsive areas beyond ITC (e.g in perirhinal cortex, entorhinal cortex, hippocampus, amygdala, prefrontal cortex), these other areas are not purely visual and also receive input from other sensory modalities.

Most, if not all, ITC neurons show visually evoked responses. ITC neurons often respond vigorously to color, orientation, texture, direction of movement and shape. PIT or TEO show a coarse retinotopic organization and an almost complete representation of the contralateral visual field. The receptive field sizes are approximately 1.5 – 4 degrees and are typically larger than the ones found in V4 neurons. There is no clear retinotopy to area TE, but there is a clear topography such that nearby neurons show similar object preferences (Tanaka, 1996). The receptive fields in area TE are often large but there is a wide range of estimations in the literature ranging from some units with ~2 degrees receptive fields (DiCarlo and Maunsell, 2004) to descriptions of units with receptive fields that span several tens of degrees (Rolls, 1991; Tanaka, 1993). Most TE receptive fields include the fovea.

7.3.   Feature preferences in inferior temporal cortex

Investigators have often found strong responses in ITC neurons elicited by all sorts of different stimuli. For example, several investigators have shown that ITC neurons can be driven by the presentation of faces, hands and body parts (Desimone, 1991; Gross et al., 1969; Perrett et al., 1982; Rolls, 1984; Young and Yamane, 1992). Other investigators have used parametric shape descriptors of abstract shapes (Miyashita and Chang, 1988; Richmond et al., 1990; Schwartz et al., 1983). Logothetis and colleagues trained monkeys to recognize paperclips forming different 3D shapes and subsequently found neurons that were selective for paperclip 3D configurations (Logothetis and Pauls, 1995).

While this wide range of responses may appear puzzling at first, it is perhaps not too surprising given a simple model where ITC neurons are tuned to
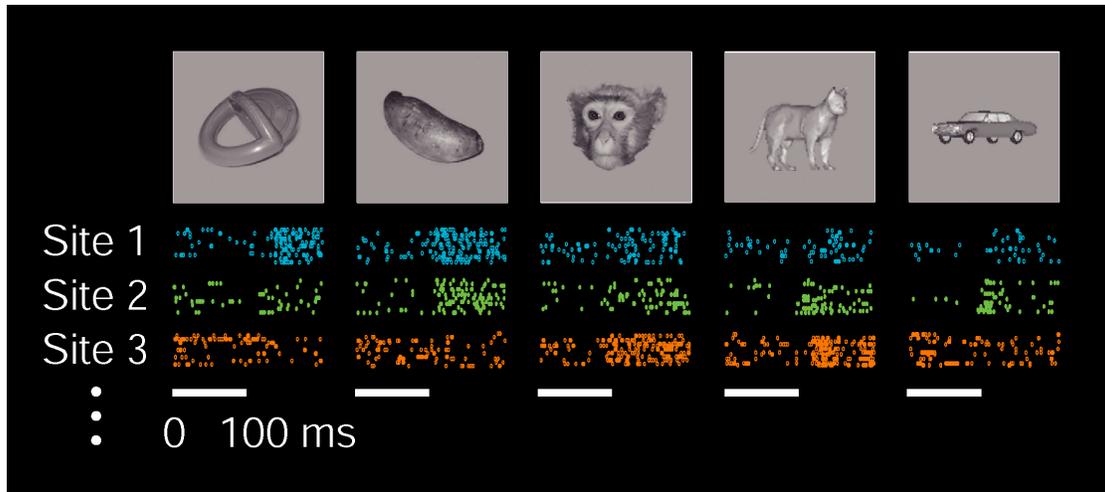
Figure 7.1. Example responses from 3 neurons in inferior temporal cortex (labeld "Site 1", "Site 2", "Site 3" to 5 different gray scale objects. Each dot represents a spike, each row represents a separate repetition (10 repetitions per object) and the horizontal white line denotes the onset and offset of the image (100 ms presentation time). Data from Hung et al, 2005.

"complex shapes". My interpretation of the wide number of stimuli that can drive ITC neurons is that these units are sensitive to complex shapes which can be found in all sorts of 2D patterns including fractal patterns, faces and paperclips. This wide range of responses also emphasizes that we still do not understand the key principles and tuning properties of ITC neurons.

As emphasized earlier, the key difficulty to elucidate the response preferences of neurons involves the *curse of dimensionality*: given limited recording time, we cannot present all possible stimuli. A promising line of research to elucidate the feature preferences in inferior temporal cortex involves changing the stimuli in real-time dictated by the neuron's preferences (Kobatake and Tanaka, 1994; Yamane et al., 2008).

Tanaka and others have shown that there is clear topography in the ITC response map. By advancing the electrode in an (approximately) tangential trajectory to cortex, he and others described that neurons within a tangential penetration show similar visual preferences (Fujita et al., 1992; Gawne and Richmond, 1993; Kobatake and Tanaka, 1994; Tanaka, 1993). They argue for the presence of "columns" and higher-order structures like "hypercolumns" in the organization of shape preferences in ITC.

While each neuron shows a preference for some shapes over others, the amount of information conveyed by individual neurons about overall shape is limited (Rolls, 1991). Additionally, there seems to be a significant amount of

A   **B**   **C**
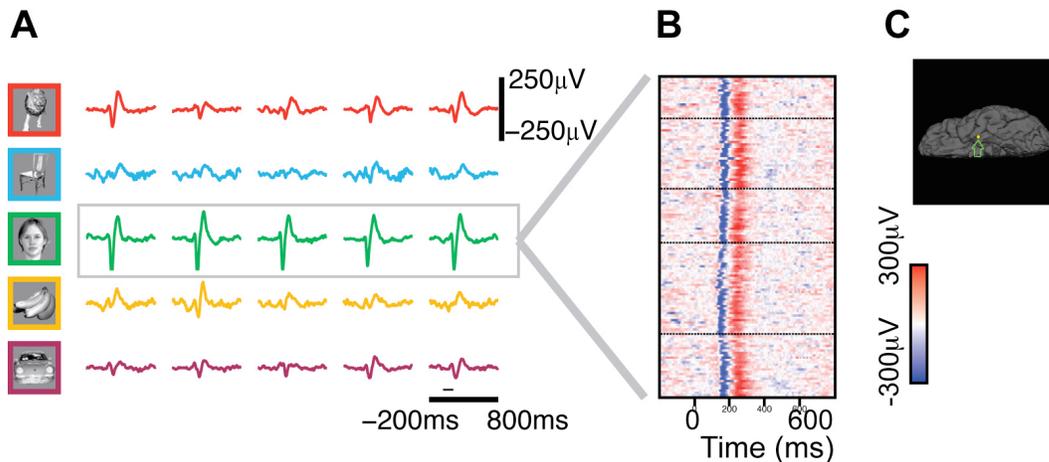
Figure 7.2. Example electrode describing the physiological responses to 25 different exemplar objects belonging to 5 different categories. **A**. Responses to each of 25 different exemplars (each color denotes a different category of images; each trace represents the response to a different exemplar). **B**. Raster plot showing every single trial in the responses to the 5 face exemplars. Each row is a repetition, the dashed lines separate the exemplars, the color shows voltage (see scale bar on right). **C**. Electrode location.

"noise"[2] in the neuronal responses in any given trial. Can the animal use the neuronal representation of a population of ITC neurons to discriminate among objects in single trials? Hung *et al* addressed this question by recording (sequentially) from hundreds of neurons and using statistical classifiers to decode the activity of a pseudo-population[3] of neurons in individual trials (Hung et al., 2005a). They found that a relatively small group of ITC neurons (~200) could support object identification and categorization quite accurately (up to ~90% and ~70% for categorization and identification respectively) with a very short latency after stimulus onset (~100 ms after stimulus onset). Furthermore, the pseudo-population response could extrapolate across changes in object scale and position. Thus, even when each neuron conveys only noisy information about shape differences, populations of neurons can be quite powerful in discriminating among visual objects in individual trials.

7.4.   Tolerance to object transformations

---

[2] The term "noise" is used in a rather vague way here. There is extensive literature on the variability of neuronal responses, the origin of this variability and whether it represents noise or signal. For the purposes of the discussion here, "noise" could be defined as the variability in the neuronal response (e.g. spike counts) across different trials when the same stimulus was presented.

[3] Because the neurons were recorded sequentially instead of simultaneously, the authors use the word pseudo-population as opposed to population of neurons.

As emphasized in **Lecture 1**, a key property of visual recognition is the capacity to recognize objects in spite of the transformations of the images at the pixel level. Several studies have shown that ITC neurons show a significant degree of tolerance to object transformations.

ITC neurons can show similar responses in spite of large changes in the size of the stimuli (Hung et al., 2005b; Ito et al., 1995; Logothetis and Pauls, 1995). Even if the absolute firing rates are affected by the stimulus size, the rank order preferences among different objects can be mainained in spite of stimulus size changes (Ito et al., 1995). ITC neurons also show more tolerance to object position changes than units in earlier parts of ventral visual cortex (Hung et al., 2005b; Ito et al., 1995; Logothetis and Pauls, 1995). ITC neurons also show a certain degree of tolerance to depth rotation (Logothetis and Sheinberg, 1996). They even show tolerance to the particular cue used to define the shape (such as luminance, motion or texture) (Sary et al., 1993).

An extreme example of tolerance to object transformations was provided by recordings performed in human epileptic patients. These are subjects that show pharmacologically-resistant forms of epilepsy. They are implanted with electrodes in order to map the location of seizures and to examine cortical function for potential surgical treatment of epilepsy. This approach provides a rare opportunity to examine neurophysiological activity in the human brain at high spatial and temporal resolution. Recording from the hippocampus, entorhinal cortex, amygdala and parahippocampal gyrus, investigators have found neurons that show responses to multiple objects within a semantically-defined object category (Kreiman et al., 2000). They have also shown that some neurons show a remarkable degree of selectivity to individual persons or landmarks. For example, one neuron showed a selective response to images where the ex-president Bill Clinton was present. Remarkably, the images that elicited a response in this neuron were quite distinct in terms of their pixel content ranging from a black/white drawing to color photographs with different poses and views (Quian Quiroga et al., 2005). As discussed above for the ITC neurons, we still do not have any understanding of the circuits and mechanisms that give rise to this type of selectivity or tolerance to object transformations.

7.5.   The path forward

Terra incognita (extrastriate ventral visual cortex), has certainly been explored at the neurophysiological level. The studies discussed here constitute a non-exhaustive list of examples of the type of responses that one might see in areas such as V2, V4 and ITC. While the field has acquired a certain number of such examples, there is an urgent need to put together these empirical observations into a coherent theory of visual recognition. In our Lecture 6, we will

discuss some of the efforts in this direction and the current status in building computational models to test theories of visual recognition.

As a final note, I conclude here with a list of questions and important challenges in the field to try to better describe what we do *not* know and what needs to be explained in terms of extrastriate visual cortex. It would be of interest to develop more quantitative and systematic approaches to examine feature preferences in extrastriate visual cortex (this also applies to other sensory modalities). Eventually, we should be able to describe a neuron's preferences in quantitative terms, starting from pixels. What types of shapes would a neuron respond to? This quantitative formulation should allow us to make predictions and extrapolations to novel shapes. It is not sufficient to show stimulus A and A" and then interpolate to predict the responses to A'. If we could really characterize the responses of the neuron, we should be able to predict the responses to a different shape B. Similarly, as emphasized multiple times, feature preferences are intricately linked to tolerance to object transformations. Therefore, we should be able to predict the neuronal response to different types of transformations of the objects. Much more work is needed to understand the computations and transformations along ventral visual cortex. How do we go from oriented bars to complex shapes such as faces? A big step would be to take a single neuron in, say, ITC, be able to examine the properties and responses of its afferent V4 units to characterize the transformations from V4 to ITC. This formulation presupposes that a large fraction of the ITC response is governed by its V4 inputs. However, we should keep in mind the complex connectivity in cortex and the fact that the ITC unit receives multiple other inputs as well (recurrent connections, bypass inputs from earlier visual areas, backprojections from the medial temporal lobe and pre-frontal cortex, connections from the dorsal visual pathway, etc). There is clearly plenty of virgin territory for the courageous investigators who dare explore the vast land of extrastriate ventral visual cortex and the computations involved in processing shapes.

7.6.   References

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. Journal of Cognitive Neuroscience 3, 1-8.

DiCarlo, J.J., and Maunsell, J.H.R. (2004). Anterior Inferotemporal Neurons of Monkeys Engaged in Object Recognition Can be Highly Sensitive to Object Retinal Position. Journal of Neurophysiology 89, 3264-3278.

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral cortex 1, 1-47.

Fujita, I., Tanaka, K., Ito, M., and Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. Nature 360, 343-346.

Gawne, T.J., and Richmond, B.J. (1993). How independent are the messages carried by adjacent inferior temporal cortical neurons? The Journal of neuroscience : the official journal of the Society for Neuroscience 13, 2758-2771.

Gross, C., Bender, D., and Rocha-Miranda, C. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. Science 166, 1303-1306.

Gross, C.G. (1994). How inferior temporal cortex became a visual area. Cerebral cortex 5, 455-469.

Hung, C., Kreiman, G., Quian-Quiroga, R., Kraskov, A., Poggio, T., and DiCarlo, J. (2005a). Using 'read-out' of object identity to understand object coding in the macaque inferior temporal cortex. In Cosyne (Salt Lake City).

Hung, C.P., Kreiman, G., Poggio, T., and DiCarlo, J.J. (2005b). Fast Read-out of Object Identity from Macaque Inferior Temporal Cortex. Science 310, 863-866.

Ito, M., Tamura, H., Fujita, I., and Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. J Neurophysiol 73, 218-226.

Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. Journal of Neuroscience 17, 4302-4311.

Kobatake, E., and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. J Neurophysiol 71, 856-867.

Kreiman, G., Koch, C., and Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. Nature neuroscience 3, 946-953.

Logothetis, N.K., and Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. Cerebral cortex 3, 270-288.

Logothetis, N.K., and Sheinberg, D.L. (1996). Visual object recognition. Annual Review of Neuroscience 19, 577-621.

Miyashita, Y., and Chang, H.S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. Nature 331, 68-71.

Perrett, D., Rolls, E., and Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. Experimental Brain Research 47, 329-342.

Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature 435, 1102-1107.

Richmond, B.J., Optican, L.M., and Spitzer, H. (1990). Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. I. Stimulus-response relations. Journal of Neurophysiology 64, 351-369.

Rolls, E. (1991). Neural organization of higher visual functions. Current opinion in neurobiology 1, 274-278.

Rolls, E.T. (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. Human Neurobiology 3, 209-222.

Sary, G., Vogels, R., and Orban, G.A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. Science 260, 995-997.

Schwartz, E., Desimone, R., Albright, T., and Gross, C. (1983). Shape-recognition and inferior temporal neurons. PNAS 80, 5776-5778.

Tanaka, K. (1993). Neuronal mechanism of object recognition. Science 262, 685-688.

Tanaka, K. (1996). Inferotemporal cortex and object vision. Annual Review of Neuroscience 19, 109-139.

Yamane, Y., Carlson, E.T., Bowman, K.C., Wang, Z., and Connor, C.E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. Nature neuroscience 11, 1352-1360.

Young, M.P., and Yamane, S. (1992). Sparse population coding of faces in the inferior temporal cortex. Science 256, 1327-1331.