Chapter VI. Part 1. First steps into inferior temporal cortex

Inferior temporal cortex (ITC) is the highest echelon within the visual stream concerned with processing visual shape information. 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 history of how inferior temporal cortex became accepted and described as a visual area is a rather interesting one (Gross, 1994).

6.1. 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 (Dean, 1976; Weiskrantz and Saunders, 1984; Afraz et al., 2015) and several human objectagnosias are correlated with damage in the inferior temporal cortex (Damasio, 1990; Humphreys and Riddoch, 1993; Forde and Humphreys, 1999) (Chapter 4). Another piece of evidence for function could come from non-invasive functional imaging studies. While non-invasive studies have limited spatiotemporal resolution and a low signal to noise ratio, they can still provide tentative hints about the coarse mapping of stimuli to some indirect metric of brain activation. For example, upon presenting images of human faces and indirectly comparing the patterns of blood flow against those obtained when the same subject looks at pictures of houses, investigators typically report increased activity in a region of ITC called the fusiform gyrus (e.g. (Kanwisher et al., 1997)).

Localizing approximate anatomical regions relevant for visual processing is only the beginning of the story. 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 inputs, 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 a spatial resolution of single neurons and with millisecond temporal resolution.

---

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.
6.2. Neuroanatomy of inferior temporal cortex

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 a vast expanse of cortex that is usually subdivided into a posterior area (PIT), a central area (CIT) and an anterior area (AIT) (Felleman and Van Essen, 1991; Logothetis and Sheinberg, 1996; Tanaka, 1996). Biologists are fond of confusing people by using different names for the same things, a phenomenon that can be partly explained by independent investigators working on related topics in parallel and using different nomenclature to describe their findings. For example, inferior temporal cortex is also referred to as areas TEO and TE in the literature.

Like most other parts of cortex, the connectivity patterns of ITC are wide and complex (Markov et al., 2014). When we describe computational models of vision (Chapter 8), it is quite clear that most models represent a major simplification of the actual connectivity diagram. ITC receives feed-forward topographically organized inputs from areas V2, V3 and V4 along the ventral visual cortex. It also receives (fewer) inputs from areas V3A and MT along the ventral visual cortex, 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.

6.3. Receptive field sizes in ITC

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 shows 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. As we move to more anterior locations along the ITC, there is weaker and weaker retinotopical organization. Yet, this does not mean a lack of topography. On the contrary, nearby neurons share similar properties: for example, two nearby neurons are much more likely to respond in a similar fashion to a set of stimuli than neurons that are farther apart (Tanaka, 1996). The receptive fields in more anterior parts of ITC are often large but there is a wide range of estimations in the literature ranging from some neurons with ~2 degrees
receptive fields (DiCarlo and Maunsell, 2004) to descriptions of neurons with receptive fields that span several tens of degrees (Rolls, 1991; Tanaka, 1993). Most receptive fields in ITC include the foveal region.

6.4. 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 (Gross et al., 1969; Perrett et al., 1982; Rolls, 1984; Desimone, 1991; Young and Yamane, 1992). Other investigators have used parametric shape descriptors of abstract shapes (Schwartz et al., 1983; Miyashita and Chang, 1988; Richmond et al., 1990). 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 “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.

Figure 6.1. Example responses from 3 neurons in inferior temporal cortex (labelled “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., 2005a).
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; Tanaka, 1993; Kobatake and Tanaka, 1994). They argue for the presence of “columns” and higher-order structures like “hypercolumns” in the organization of shape preferences in ITC.

More recent work suggests that we may need to rethink the neural code for features in ITC (and perhaps earlier visual areas as well). Following up on the ideas developed by Yamane et al to let the neuron itself reveal what it likes rather than impose a strong bias in the stimulus selection, Xiao and colleagues developed a computational algorithm that is capable of generating images guided by neuronal firing rates. They use a genetic algorithm using the neuron’s firing rate as the fitness function. In a given generation, the investigators probe the responses to a set of images. Images that trigger high firing rates are kept, and the rest are modified and recombined by the generative algorithm. In Chapter 8, we will introduce deep hierarchical models of vision that start with pixels and yield a high-level feature representation. The generative algorithm deployed by Xiao and colleagues is essentially an inverted version of those computational models, starting with high level features and ending up with the generation of the pixels in an image.

By running this generative computational algorithm while recording the activity of a neuron in ITC, they discovered images that elicited higher firing rates than any natural image that had been used before to test the responses of the neurons. These images contain naturalistic combinations of textures and broad strokes, which have been described by investigators as impressionist (e.g. Monet) renderings of abstract art like a Kandinsky. The fundamental novel concept here is that neurons may be optimally activated by combinations of complex features that cannot be easily described in words. In contrast to the language-based anthropomorphic descriptions of neuronal feature preferences in ITC (“this neuron likes faces”, “this neuron likes chairs”, “this neuron likes convex curved shapes”), the new line of work suggests that neurons might be optimally activated by complex shapes that defy a definition. A rich basis set of neurons tuned to such complex features is capable of allowing the organism to discriminate real-world objects, but the basis set does not have to be based on real-world objects.
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 "noise" 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., 2005b). 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.

---

\(^2\) 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.
6.5. **Tolerance to object transformations**

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 (Ito et al., 1995; Logothetis and Pauls, 1995; Hung et al., 2005c). Even if the absolute firing rates are affected by the stimulus size, the rank order preferences among different objects can be maintained 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 (Ito et al., 1995; Logothetis and Pauls, 1995; Hung et al., 2005c). 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.

6.6. **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.

6.7. References


