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[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 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 object agnosias 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.
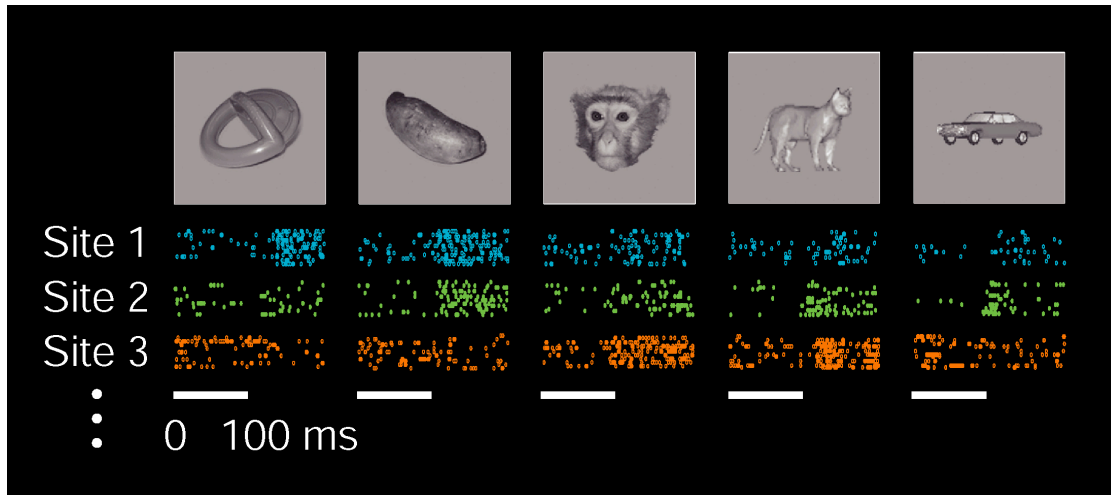
6.2.   Neuroanatomy of inferior temporal cortex

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

Figure 6.1*. 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., 2005a).*

89 receptive fields (DiCarlo and Maunsell, 2004) to descriptions of neurons with
90 receptive fields that span several tens of degrees (Rolls, 1991; Tanaka, 1993).
91 Most receptive fields in ITC include the foveal region.
92
93 **6.4.   Feature preferences in inferior temporal cortex**
94
95   Investigators have often found strong responses in ITC neurons elicited
96 by all sorts of different stimuli. For example, several investigators have shown
97 that ITC neurons can be driven by the presentation of faces, hands and body
98 parts (Gross et al., 1969; Perrett et al., 1982; Rolls, 1984; Desimone, 1991;
99 Young and Yamane, 1992). Other investigators have used parametric shape
100 descriptors of abstract shapes (Schwartz et al., 1983; Miyashita and Chang,
101 1988; Richmond et al., 1990). Logothetis and colleagues trained monkeys to
102 recognize paperclips forming different 3D shapes and subsequently found
103 neurons that were selective for paperclip 3D configurations (Logothetis and
104 Pauls, 1995).
105
106   While this wide range of responses may appear puzzling at first, it is
107 perhaps not too surprising given a simple model where ITC neurons are tuned to
108 "complex shapes". My interpretation of the wide number of stimuli that can drive
109 ITC neurons is that these units are sensitive to complex shapes which can be
110 found in all sorts of 2D patterns including fractal patterns, faces and paperclips.
111 This wide range of responses also emphasizes that we still do not understand
112 the key principles and tuning properties of ITC neurons.
113

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

120

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

128

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

143

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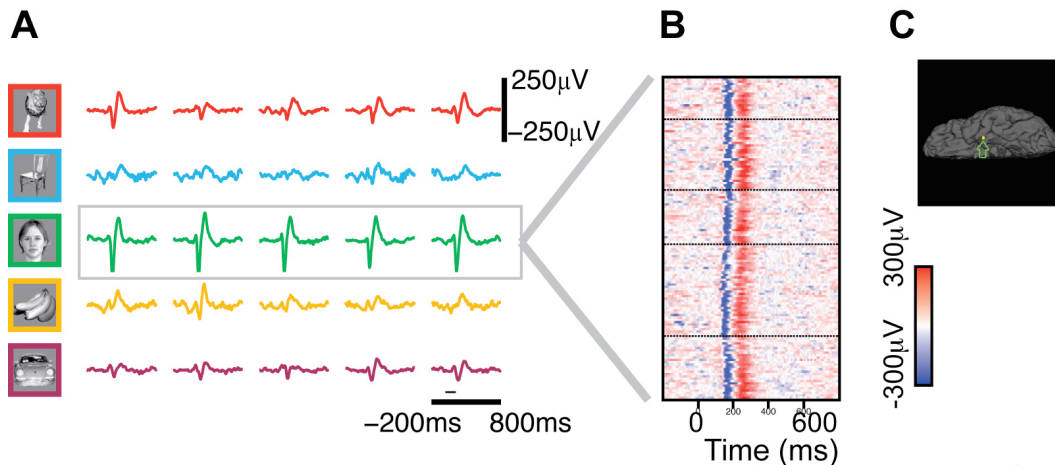158   real-world objects.

159

*Figure 6.2. 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.*

160   While each neuron shows a preference for some shapes over others, the
161   amount of information conveyed by individual neurons about overall shape is
162   limited (Rolls, 1991). Additionally, there seems to be a significant amount of
163   "noise"[2] in the neuronal responses in any given trial. Can the animal use the
164   neuronal representation of a population of ITC neurons to discriminate among
165   objects in single trials? Hung *et al* addressed this question by recording
166   (sequentially) from hundreds of neurons and using statistical classifiers to
167   decode the activity of a pseudo-population[3] of neurons in individual trials (Hung
168   et al., 2005b). They found that a relatively small group of ITC neurons (~200)
169   could support object identification and categorization quite accurately (up to
170   ~90% and ~70% for categorization and identification respectively) with a very
171   short latency after stimulus onset (~100 ms after stimulus onset). Furthermore,
172   the pseudo-population response could extrapolate across changes in object
173   scale and position. Thus, even when each neuron conveys only noisy information
174   about shape differences, populations of neurons can be quite powerful in
175   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.

176
177 **6.5.   Tolerance to object transformations**
178
179   As emphasized in **Lecture 1**, a key property of visual recognition is the
180 capacity to recognize objects in spite of the transformations of the images at the
181 pixel level. Several studies have shown that ITC neurons show a significant
182 degree of tolerance to object transformations.
183
184   ITC neurons can show similar responses in spite of large changes in the
185 size of the stimuli (Ito et al., 1995; Logothetis and Pauls, 1995; Hung et al.,
186 2005c). Even if the absolute firing rates are affected by the stimulus size, the
187 rank order preferences among different objects can be mainained in spite of
188 stimulus size changes (Ito et al., 1995). ITC neurons also show more tolerance to
189 object position changes than units in earlier parts of ventral visual cortex (Ito et
190 al., 1995; Logothetis and Pauls, 1995; Hung et al., 2005c). ITC neurons also
191 show a certain degree of tolerance to depth rotation (Logothetis and Sheinberg,
192 1996). They even show tolerance to the particular cue used to define the shape
193 (such as luminance, motion or texture) (Sary et al., 1993).
194
195   An extreme example of tolerance to object transformations was provided
196 by recordings performed in human epileptic patients. These are subjects that
197 show pharmacologically-resistant forms of epilepsy. They are implanted with
198 electrodes in order to map the location of seizures and to examine cortical
199 function for potential surgical treatment of epilepsy. This approach provides a
200 rare opportunity to examine neurophysiological activity in the human brain at high
201 spatial and temporal resolution. Recording from the hippocampus, entorhinal
202 cortex, amygdala and parahippocampal gyrus, investigators have found neurons
203 that show responses to multiple objects within a semantically-defined object
204 category (Kreiman et al., 2000). They have also shown that some neurons show
205 a remarkable degree of selectivity to individual persons or landmarks. For
206 example, one neuron showed a selective response to images where the ex-
207 president Bill Clinton was present. Remarkably, the images that elicited a
208 response in this neuron were quite distinct in terms of their pixel content ranging
209 from a black/white drawing to color photographs with different poses and views
210 (Quian Quiroga et al., 2005). As discussed above for the ITC neurons, we still do
211 not have any understanding of the circuits and mechanisms that give rise to this
212 type of selectivity or tolerance to object transformations.
213
214 **6.6.   The path forward**
215
216   Terra incognita (extrastriate ventral visual cortex), has certainly been
217 explored at the neurophysiological level. The studies discussed here constitute a
218 non-exhaustive list of examples of the type of responses that one might see in
219 areas such as V2, V4 and ITC. While the field has acquired a certain number of
220 such examples, there is an urgent need to put together these empirical
221 observations into a coherent theory of visual recognition. In our Lecture 6, we will

222 discuss some of the efforts in this direction and the current status in building
223 computational models to test theories of visual recognition.
224
225   As a final note, I conclude here with a list of questions and important
226 challenges in the field to try to better describe what we do *not* know and what
227 needs to be explained in terms of extrastriate visual cortex. It would be of interest
228 to develop more quantitative and systematic approaches to examine feature
229 preferences in extrastriate visual cortex (this also applies to other sensory
230 modalities). Eventually, we should be able to describe a neuron's preferences in
231 quantitative terms, starting from pixels. What types of shapes would a neuron
232 respond to? This quantitative formulation should allow us to make predictions
233 and extrapolations to novel shapes. It is not sufficient to show stimulus A and A"
234 and then interpolate to predict the responses to A'. If we could really characterize
235 the responses of the neuron, we should be able to predict the responses to a
236 different shape B. Similarly, as emphasized multiple times, feature preferences
237 are intricately linked to tolerance to object transformations. Therefore, we should
238 be able to predict the neuronal response to different types of transformations of
239 the objects. Much more work is needed to understand the computations and
240 transformations along ventral visual cortex. How do we go from oriented bars to
241 complex shapes such as faces? A big step would be to take a single neuron in,
242 say, ITC, be able to examine the properties and responses of its afferent V4 units
243 to characterize the transformations from V4 to ITC. This formulation presupposes
244 that a large fraction of the ITC response is governed by its V4 inputs. However,
245 we should keep in mind the complex connectivity in cortex and the fact that the
246 ITC unit receives multiple other inputs as well (recurrent connections, bypass
247 inputs from earlier visual areas, backprojections from the medial temporal lobe
248 and pre-frontal cortex, connections from the dorsal visual pathway, etc). There is
249 clearly plenty of virgin territory for the courageous investigators who dare explore
250 the vast land of extrastriate ventral visual cortex and the computations involved in
251 processing shapes.
252
253
254 **6.7.   References**
255

256 Afraz A, Boyden ES, DiCarlo JJ (2015) Optogenetic and pharmacological suppression
257   of spatial clusters of face neurons reveal their causal role in face gender
258   discrimination. Proceedings of the National Academy of Sciences of the
259   United States of America 112:6730-6735.
260 Damasio A (1990) Category-related recognition defects as a clue to the neural
261   substrates of knowledge. TRENDS IN NEUROSCIENCES 13:95-98.
262 Dean P (1976) Effects of inferotemporal lesions on the behavior of monkeys.
263   Psychological Bulletin 83:41-71.
264 Desimone R (1991) Face-selective cells in the temporal cortex of monkeys. Journal
265   of Cognitive Neuroscience 3:1-8.

266 DiCarlo JJ, Maunsell JHR (2004) Anterior Inferotemporal Neurons of Monkeys
267   Engaged in Object Recognition Can be Highly Sensitive to Object Retinal
268   Position. Journal of Neurophysiology 89:3264-3278.
269 Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the
270   primate cerebral cortex. Cerebral cortex 1:1-47.
271 Forde E, Humphreys G (1999) Category-specific recognition impairments: a review
272   of important case studies and influential theories. Aphasiology 13:169-193.
273 Fujita I, Tanaka K, Ito M, Cheng K (1992) Columns for visual features of objects in
274   monkey inferotemporal cortex. Nature 360:343-346.
275 Gawne TJ, Richmond BJ (1993) How independent are the messages carried by
276   adjacent inferior temporal cortical neurons? Journal of Neuroscience
277   13:2758-2771.
278 Gross C, Bender D, Rocha-Miranda C (1969) Visual receptive fields of neurons in
279   inferotemporal cortex of the monkey. Science 166:1303-1306.
280 Gross CG (1994) How inferior temporal cortex became a visual area. Cerebral cortex
281   5:455-469.
282 Humphreys G, Riddoch M (1993) Object agnosias. BAILLIERES CLINICAL
283   NEUROLOGY 2:339-359.
284 Hung C, Kreiman G, Poggio T, Dicarlo J (2005a) Ultra-fast object recognition from
285   few spikes. In. Boston: MIT.
286 Hung C, Kreiman G, Quian-Quiroga R, Kraskov A, Poggio T, DiCarlo J (2005b) Using
287   'read-out' of object identity to understand object coding in the macaque
288   inferior temporal cortex. In: Cosyne. Salt Lake City.
289 Hung CP, Kreiman G, Poggio T, DiCarlo JJ (2005c) Fast Read-out of Object Identity
290   from Macaque Inferior Temporal Cortex. Science 310:863-866.
291 Ito M, Tamura H, Fujita I, Tanaka K (1995) Size and position invariance of neuronal
292   responses in monkey inferotemporal cortex. J Neurophysiol 73:218-226.
293 Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in
294   human extrastriate cortex specialized for face perception. Journal of
295   Neuroscience 17:4302-4311.
296 Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in
297   the ventral visual pathway of the macaque cerebral cortex. J Neurophysiol
298   71:856-867.
299 Kreiman G, Koch C, Fried I (2000) Category-specific visual responses of single
300   neurons in the human medial temporal lobe. Nature Neuroscience 3:946-
301   953.
302 Logothetis NK, Pauls J (1995) Psychophysical and physiological evidence for viewer-
303   centered object representations in the primate. Cerebral cortex 3:270-288.
304 Logothetis NK, Sheinberg DL (1996) Visual object recognition. Annual Review of
305   Neuroscience 19:577-621.
306 Markov NT et al. (2014) A Weighted and Directed Interareal Connectivity Matrix for
307   Macaque Cerebral Cortex. Cerebral cortex 24:17-36.
308 Miyashita Y, Chang HS (1988) Neuronal correlate of pictorial short-term memory in
309   the primate temporal cortex. Nature 331:68-71.
310 Perrett D, Rolls E, Caan W (1982) Visual neurones responsive to faces in the monkey
311   temporal cortex. Experimental Brain Research 47:329-342.

312   Quian Quiroga R, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual
313   representation by single neurons in the human brain. Nature 435:1102-1107.
314   Richmond BJ, Optican LM, Spitzer H (1990) Temporal encoding of two-dimensional
315   patterns by single units in primate primary visual cortex. I. Stimulus-
316   response relations. Journal of Neurophysiology 64:351-369.
317   Rolls E (1991) Neural organization of higher visual functions. Current opinion in
318   neurobiology 1:274-278.
319   Rolls ET (1984) Neurons in the cortex of the temporal lobe and in the amygdala of
320   the monkey with responses selective for faces. Human Neurobiology 3:209-
321   222.
322   Sary G, Vogels R, Orban GA (1993) Cue-invariant shape selectivity of macaque
323   inferior temporal neurons. Science 260:995-997.
324   Schwartz E, Desimone R, Albright T, Gross C (1983) Shape-recognition and inferior
325   temporal neurons. PNAS 80:5776-5778.
326   Tanaka K (1993) Neuronal mechanism of object recognition. Science 262:685-688.
327   Tanaka K (1996) Inferotemporal cortex and object vision. Annual Review of
328   Neuroscience 19:109-139.
329   Weiskrantz L, Saunders R (1984) Impairments of visual object transforms in
330   monkeys. Brain : a journal of neurology 107:1033-1072.
331   Yamane Y, Carlson ET, Bowman KC, Wang Z, Connor CE (2008) A neural code for
332   three-dimensional object shape in macaque inferotemporal cortex. Nature
333   neuroscience 11:1352-1360.
334   Young MP, Yamane S (1992) Sparse population coding of faces in the inferior
335   temporal cortex. Science 256:1327-1331.
336