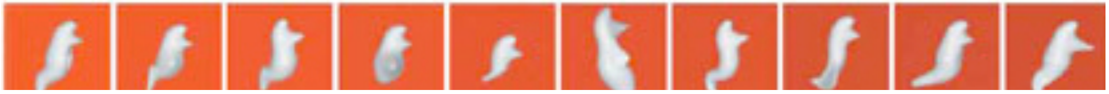
BEWARE: These are preliminary notes. In the future, they will become part of a textbook on Visual Object Recognition.

Chapter VIII. From inferior temporal cortex to cognition

In a basic attempt to simplify, categorize and organize our understanding of brain processes, we tend to use labels such as "vision", "audition", "memory", "planning" or "decision making". In some cases, these distinctions are warranted. In the case of the retina, we can safely assume that the responses are governed by the incoming visual stimuli and are largely independent of task demands (e.g. whether the subject is looking for his car keys), expertise (e.g. whether the person has seen a particular object before), cognitive state (e.g. whether the person is paying attention to one or another location within the visual field) and other variables. Since there is no feedback from the brain to the retina, we can assume with a certain degree of safety that cognitive demands such as the task at hand will not significantly impact the responses of retinal ganglion cells (except for eye movements of course). The distinctions become fuzzy as we ascend the visual hierarchy. Therefore, given the position of inferior temporal cortex (ITC) at the top of the ventral visual hierarchy, it should come as no surprise that the neurophysiological responses in ITC are strongly modulated by task demands and other non-visual constraints. In this chapter, we examine some of the ways through which the spatial surround, the temporal surround, learning and task constraints influence the neurophysiological responses in ITC.

8.1. What is in the brain of a neuron?

Figure 8.1. *Preferred stimuli for an ITC neuron. The neuron may respond to a complex combination of features that does not directly resemble known objects but forms a basis set for recognizing complex shapes. From (Yamane et al., 2008).*



One day, we may be able to inject a tracer in a neuron in ITC, follow all its inputs and thus construct a detailed quantitative model of what types of stimuli drive its responses[1]. Until that day, let us speculate about a possible simple scenario. As described in the previous chapter, ITC neurons receive "bottom-up" inputs from earlier ventral stream visual areas such as area V4. Additionally, ITC neurons receive "top-down" inputs from several areas including pre-frontal

---

[1] This statement assumes an understanding of the stimulus preferences of the input neurons.

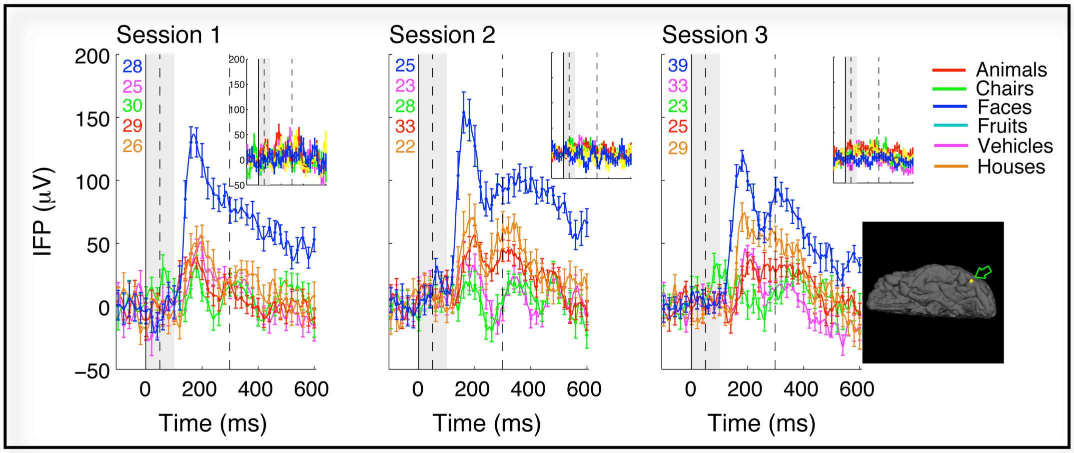
cortex. The simplest possible description is that the responses of neurons in ITC represent a combination of those bottom-up inputs with cognitive influences instantiated via top-down connections. The stimulus-driven part of the response could represent the (non-linear, weighted) sum of its inputs in the same way that we can think of orientation selectivity in primary visual cortex as arising from a suitable combination of LGN-type receptive fields. According to this simple hypothesis, neurons in ITC may respond to complex shape combinations (e.g. (Connor et al., 2007)). Several anthropomorphic approaches use two-dimensional renderings of objects such as cars or shoes. While these objects can certainly drive neurons in ITC, they do not necessarily imply that ITC neurons care about cars, shoes or any such shape. The specific features preferred by ITC neurons may be differentially represented in those complex objects leading investigators to describe a “car” preferring neuron. Care should be taken in the interpretation of such statements. A neuron in ITC could prefer complex combinations of stimulus features such as the ones illustrated in Figure 8.1 (Yamane et al., 2008).

8.1. Neuronal responses in ITC dynamically evolve over time

There is a progression in the response latencies of neurons along the ventral visual stream (**Chapter 6**). The response latencies in ITC are, on average, longer than the ones in earlier visual areas, with latency values from ~80 to ~120 ms (e.g. (Hung et al., 2005; Richmond et al., 1990; Tovee et al., 1993)). Neurons in ITC can also rapidly respond to very brief stimulus flashes, as short as 14 ms (Keysers et al., 2001). Furthermore, the activity of a small population of ITC neurons in the initial ~100 ms after stimulus onset provides sufficient information to decode the identity (or category) of the stimulus (Hung et al., 2005; Tovee et al., 1993; Treves et al., 1999). Slightly longer but similar latencies are observed in the human inferior temporal cortex (Liu et al., 2009). These latencies are consistent with the overall behavioral estimates of the amount of time required for object recognition (**Chapter 4**). These short latencies impose a strong constraint for the development of a biophysically plausible theory of visual object recognition.

Some studies have argued that the latency of ITC neurons depends on stimulus characteristics. For example, Sugase et al showed macaque monkeys images that were divided into three different categories (monkey faces, human faces, and simple geometric shapes) and presented several different exemplars within each category. They argued that whereas the early responses (before 100 ms) in ITC were able to discriminate among the three categories, they did not provide sufficient information to distinguish individual exemplars (Sugase et al., 1999). The finer information involved in identification was available more than 50 ms later. In another study, Meyers and colleagues used a machine learning approach to show that the population activity in ITC evolves over time and that different subpopulations of neurons may encode object and category information at different time points (Meyers et al., 2008).

Figure 8.2. *Stability of visual responses in the human ventral visual stream. Responses of an electrode in the human brain across multiple sessions (columns; the time between sessions was 46 and 24 hours, respectively). Each curve shows the average intracranial field potential (IFP) response to all images from a given category (red=animals, green=chairs, blue=faces, cyan=fruits, purple=vehicles, orange=houses). The error bars indicate ±SEM and the numbers of trials for each category contributing to the mean waveform are indicated in each subplot. The gray rectangle denotes the image presentation time. The vertical dashed lines denote the time period used for the analyses in the text (50 to 300 ms after stimulus onset). Each of these electrodes showed a statistically significant selective response (p<0.01, one-way ANOVA; Methods).*



8.1. Stability versus learning

What about longer time scales? If an electrode in ITC responds to certain shapes one day, will it still respond to the same shapes on a different day? The responses in ITC need to be sufficiently stable over long time spans if they are to subserve visual recognition. Indeed, recordings over multiple sessions spanning hours to days show that responses show temporal stability both in the human brain (Bansal et al., 2012) as well as in the macaque brain (Bondar et al., 2009).

At the same time, the responses of ITC neurons are not fixed. Rather, learning has a significant effect on the neuronal preferences in ITC. It is tempting to conjecture that, at least in part, learning to recognize novel shapes and objects depends in the long term on adjustments to the firing properties of ITC neurons. A series of elegant studies that speak to this property of ITC neurons was carried out by Miyashita (Higuchi and Miyashita, 1996; Miyashita, 1988; Miyashita and Chang, 1988). He presented sequences of fractal patterns to a monkey while recording the activity of ITC neurons. Day after day, the temporal sequence was the same and he observed that some of the neurons started to develop tuning for objects depending on the presentation order. The tuning of ITC neurons can also be sharpened by experience (Freedman et al., 2005). Furthermore, it is also possible to observe rapid learning effects whereby the neuronal preferences can be modified in an unsupervised fashion during a single recording session (Li and Dicarlo, 2008).

8.4 The effect of spatial clutter and attentional modulation

Most of the examples that we have provided so far involve the presentation of isolated stimuli on a uniform background. In the real world, stimuli rarely show up on a uniform background and the visual system has to deal with the problem of clutter and identifying objects embedded amidst other objects and complex backgrounds. The difficulty of this problem is emphasized by games such as “Where is Waldo?” where you need to find an object surrounded by many similar stimuli. Many animal species astutely capitalize on the difficulty of recognizing objects in clutter by using camouflage.

Neurons in ITC (as well as neurons in earlier parts of ventral visual cortex) are significantly affected by the presence of other stimuli. The responses in area V4 (Connor et al., 1997; Ghose and Maunsell, 2008) and ITC (Chelazzi et al., 1998; De Baene et al., 2007; Li et al., 2009; Miller et al., 1993; Missal et al., 1999; Rolls and Tovee, 1995; Sato, 1989; Sheinberg and Logothetis, 2001; Zoccolan et al., 2005; Zoccolan et al., 2007) to the neuron’s preferred stimuli are suppressed by the addition of a second object within the receptive field. The degree of suppression varies substantially across recording areas and experimental conditions: some studies report strong suppression (e.g. (Rolls and Tovee, 1995)) or even that neurons may compute the average of the responses to the individual objects (De Baene et al., 2007; Zoccolan et al., 2005) while others report almost no suppression (e.g. (Gawne and Martin, 2002; Quian Quiroga et al., 2005)).

Given the (sometimes massive) response reduction observed in single neurons, how can we recognize objects in natural scenes at all? There are least two non-exclusive answers. First, small amounts of clutter can be overcome by the combination of multiple neurons (Agam et al., 2010; Li et al., 2009). Ultimately, even large neuronal populations will be impaired by heavy clutter. In those cases, the visual system uses *attention* to filter out parts of the visual input (Reynolds and Chelazzi, 2004). In the simplest instantiation, one can focus attention on a given part of the visual field, enhancing our discriminative power within the spotlight at the expense of discrimination outside. Several studies have demonstrated that spatial attention strongly enhances the responses of neurons (Fries et al., 2001; Luck et al., 1997; Moran and Desimone, 1985; Reynolds et al., 1999; Treue and Maunsell, 1999). The strength of spatial attention affects are particularly strong in ITC and seem to follow the reverse order of the visual hierarchy (Hochstein and Ahissar, 2002).

8.1. Does it matter to you? Then it matters to ITC as well

In addition to attentional modulation (mentioned above), neuronal activity in ITC is influenced by any aspect of cognition that you may think of. ITC neurons are interested in what you are doing with the visual information, what you are paying attention to, whether you want to retain the information, whether you have perceived the stimulus.

For example, imagine a neuron that responds vigorously to object A and not to object B and consider the following simple task. In some trials, the monkey has to search for and saccade to target A whereas in other trials the monkey has to search for and saccade to target B. The neuron will strongly enhance its activity when A is the target compared to those trials when B is the target even when the visual stimulus is identical in both conditions (Chelazzi et al., 1998). Furthermore, following up on the same example, in each trial, the monkey is instructed that the target is A or B, and there is a delay before the image with the two objects shows us. During the delay, even though there is no visual stimulus, the neuron fires more strongly when A is the target. In other words, the neuronal response correlates with the short-term memory required to solve the task. It is tempting to think of these neuronal responses as a correlate of visual imagery (Miyashita, 1993).

An extreme example of response modulation in ITC is provided by the phenomenon of binocular rivalry. If you present a stimulus A to the right eye and a stimulus B to your left eye, most of the time you do *not* perceive a mixture of A and B. Instead, your perception alternates in a seemingly random fashion between brief periods where you see A and brief periods where you see B (Alais and Blake, 2005; Blake and Logothetis, 2002). Given that the stimulus is constant (except for very small eye movements), what brain processes correlate with your alternating percepts? In a remarkable study, Sheinberg and Logothetis showed that almost all neurons in ITC strongly respond when the monkey is perceiving the preferred stimulus and they are essentially shut down while the monkey perceives the non-preferred stimulus (Sheinberg and Logothetis, 1997). In other words, a neuron selective to object A will fire if and only if the monkey perceives A, even though A is still present on the screen when the monkey perceives B.

8.1. References

Agam, Y., Liu, H., Pappanastassiou, A., Buia, C., Golby, A.J., Madsen, J.R., and Kreiman, G. (2010). Robust selectivity to two-object images in human visual cortex. Current Biology 20, 872-879.

Alais, D., and Blake, R. (2005). Binocular rivalry (Cambridge, Mass.: MIT Press).

Bansal, A., Singer, J., Anderson, W., Golby, A., Madsen, J., and Kreiman, G. (2012). Temporal stability of visually selective responses in intracranial field potentials recorded from human occipital and temporal lobes. Journal of Neurophysiology 108, 3073-3086.

Blake, R., and Logothetis, N. (2002). Visual competition. Nature Reviews Neuroscience 3, 13-21.

Bondar, I., Leopold, D., Richmond, B., Victor, J., and Logothesis, N. (2009). Long-term stability of visual pattern selective responses of monkey temporal lobe neurons PloS one 9.

Chelazzi, L., Duncan, J., Miller, E.K., and Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. Journal of Neurophysiology 80, 2918-2940.

Connor, C.E., Brincat, S.L., and Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. Current opinion in neurobiology 17, 140-147.

Connor, C.E., Preddie, D.C., Gallant, J.L., and Van Essen, D.C. (1997). Spatial attention effects in macaque area V4. The Journal of neuroscience : the official journal of the Society for Neuroscience 17, 3201-3214.

De Baene, W., Premereur, E., and Vogels, R. (2007). Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. J Neurophysiol 97, 2900-2916.

Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2005). Experience-Dependent Sharpening of Visual Shape Selectivity in Inferior Temporal Cortex. Cerebral cortex.

Fries, P., Reynolds, J., Rorie, A., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. Science 23, 1560-1563.

Gawne, T.J., and Martin, J.M. (2002). Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. J Neurophysiol 88, 1128-1135.

Ghose, G.M., and Maunsell, J.H. (2008). Spatial summation can explain the attentional modulation of neuronal responses to multiple stimuli in area V4. The Journal of neuroscience : the official journal of the Society for Neuroscience 28, 5115-5126.

Higuchi, S., and Miyashita, Y. (1996). Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. PNAS 93, 739-743.

Hochstein, S., and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. Neuron 36, 791-804.

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

Keysers, C., Xiao, D.K., Foldiak, P., and Perret, D.I. (2001). The speed of sight. Journal of Cognitive Neuroscience 13, 90-101.

Li, N., Cox, D.D., Zoccolan, D., and DiCarlo, J.J. (2009). What response properties do individual neurons need to underlie position and clutter "invariant" object recognition? J Neurophysiol 102, 360-376.

Li, N., and Dicarlo, J.J. (2008). Unsupervised Natural Experience Rapidly Alters Invariant Object Representation in Visual Cortex. Science 321, 1502-1507.

Liu, H., Agam, Y., Madsen, J.R., and Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. Neuron 62, 281-290.

Luck, S.J., Chelazzi, L., Hillyard, S.A., and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J Neurophysiol 77, 24-42.

Meyers, E., Freedman, D., Kreiman, G., Miller, E., and Poggio, T. (2008). Dynamic Population Coding of Category Information in ITC and PFC. Journal of Neurophysiology 100, 1407-1419.

Miller, E.K., Gochin, P.M., and Gross, C.G. (1993). Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. Brain Res 616, 25-29.

Missal, M., Vogels, R., Li, C.Y., and Orban, G.A. (1999). Shape interactions in macaque inferior temporal neurons. J Neurophysiol 82, 131-142.

Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. Nature 335, 817-820.

Miyashita, Y. (1993). Inferior Temporal Cortex: Where visual perception meets memory. Annual Review of Neuroscience 16, 245-263.

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

Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. Science 229, 782-784.

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.

Reynolds, J.H., and Chelazzi, L. (2004). Attentional modulation of visual processing. Annu Rev Neurosci 27, 611-647.

Reynolds, J.H., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. The Journal of neuroscience : the official journal of the Society for Neuroscience 19, 1736-1753.

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.T., and Tovee, M.J. (1995). The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. Experimental brain research Experimentelle Hirnforschung Experimentation cerebrale 103, 409-420.

Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. Experimental brain research Experimentelle Hirnforschung Experimentation cerebrale 77, 23-30.

Sheinberg, D.L., and Logothetis, N.K. (1997). The role of temporal areas in perceptual organization. Proceedings of the National Academy of Sciences, USA 94, 3408-3413.

Sheinberg, D.L., and Logothetis, N.K. (2001). Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. The Journal of neuroscience : the official journal of the Society for Neuroscience 21, 1340-1350.

Sugase, Y., Yamane, S., Ueno, S., and Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. Nature 400, 869-873.
Tovee, M.J., Rolls, E.T., Treves, A., and Bellis, R.P. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. J Neurophysiol 70, 640-654.
Treue, S., and Maunsell, J. (1999). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. Journal of Neuroscience 19, 7591-7602.
Treves, A., Panzeri, S., Rolls, E., Booth, M., and Wakeman, E. (1999). Firing Rate Distribution and Efficiency of Information Transmission of Inferior Temporal Cortex Neurons to Natural Visual Stimuli. Neural Computational 11, 30.
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.
Zoccolan, D., Cox, D.D., and DiCarlo, J.J. (2005). Multiple object response normalization in monkey inferotemporal cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 25, 8150-8164.
Zoccolan, D., Kouh, M., Poggio, T., and DiCarlo, J.J. (2007). Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 27, 12292-12307.